Project Deliverable Report

DELIVERABLE 4.2

PREDICTIVE SIMULATION MODEL

WORK PACKAGE NUMBER: WP4

WORK PACKAGE TITLE: MODELLING & DIGITAL TWINS

TYPE: REPORT

AUTHORS: CRISTINA CORNARO, MARCELLO PETITTA, GIANLUIGI BOVESECCHI, ALI SOHANI, FRANCESCO BISO, FEDERICO ANDREOZZI, LUCA ROSATI, EMILIANO SERI

REGACE Action Information	
Action full title	Responsive Greenhouse Agrivoltaics System with CO2 Enrichment for Higher Yields
Action acronym	REGACE
Grant agreement number	101096056
Project coordinator	Prof. Ibrahim Yehia
Project start date and duration	1 February 2023, 36 months
Project website	https://regaceproject.com/

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

Document Information

Deliverable Information	
Work package number	**4**
Work package title	**Modelling & Digital Twins**
Deliverable number	**4.2**
Deliverable title	**PREDICTIVE SIMULATION MODEL**
Description	**Develop a predictive simulation model of the greenhouse system that includes microclimate modelling, PV system modelling and provides an estimation for crop growth.**
Lead beneficiary	**Tor Vergata University - UNITOV**
Lead Author(s)	**Tor Vergata University - UNITOV**
Contributor(s)	**Cristina Cornaro, Marcello Petitta, Gianluigi Bovesecchi, Ali Sohani, Francesco Biso, Federico Andreozzi, Luca Rosati, Emiliano Seri - UNITOV, Ibrahim Yehia - TRDC, Papaioannou Chrysoula, Catherine Baxevanou, Dimitrios Fidaros - UTH, Uwe Schmidt, Thorsten Rocksch - HU**
Revision number	**1.0**
Revision Date	**27 January 2026**
Status (Final (F), Draft (D), Revised Draft (RV))	**F - Final**
Dissemination level (Public (PU), Restricted to other program participants (PP), Restricted to group speci-fied by consortium (RE), Confidential for consortium members only (CO))	**PU**

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Document History			
Revision	**Date**	**Modification**	**Author**
0.1	2025/12/11	Outline	Cristina Cornaro
0.2	2025/12/20	Initial draft	Cristina Cornaro
0.3	2026/01/16	Inputs from all contributors	Cristina Cornaro, Marcello Petitta, Gianluigi Bovesecchi, Ali Sohani, Francesco Biso, Federico Andreozzi, Luca Rosati, Emiliano Seri - UNITOV, Ibrahim Yehia - TRDC, Papaioannou Chrysoula, Catherine Baxevanou, Dimitrios Fidaros - UTH, Uwe Schmidt, Thorsten Rocksch - HU
1.0	2026/01/27	Final version for submission	Cristina Cornaro

Approvals				
	Name	**Organisation**	**Date**	**Signature (initials)**
Coordinator	Prof. Ibrahim Yehia	Alzahrawy Society	2026/01/27	IY
WP Leaders	Cristina Cornaro	Tor Vergata University	2026/01/27	CC

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

Table of Contents

LIST OF FIGURES ... 6
LIST OF TABLES ... 12
EXECUTIVE SUMMARY ... 14
INTRODUCTION ... 16
CHAPTER 1. DATA DESCRIPTION AND MANAGEMENT ... 17
INTRODUCTION ... 17
1.1 UTH GREENHOUSE AT “FARM OF VELESTINO” (GREECE) ... 17
1.2 FSC GREENHOUSE AT “FATTORIA SOLIDALE DEL CIRCEO” (ITALY) ... 21
CONCLUSIONS ... 29
CHAPTER 2. DYNAMIC MODELLING OF THE GREENHOUSE MICROCLIMATE ... 31
INTRODUCTION ... 31
2.1 IDA-ICE MODELLING OF THE GREENHOUSE (FSC) ... 32
2.2 VALIDATION RESULTS ... 39
CONCLUSIONS ... 46
REFERENCES ... 48
CHAPTER 3. CFD MODELLING OF THE GREENHOUSE MICROCLIMATE ... 49
INTRODUCTION ... 49
3.1 AVAILABLE DATA ... 49
3.2 OVERVIEW OF THE CFD MODELLING APPROACH ... 50
3.3 SIMULATION DOMAINS AND NUMERICAL SETUP ... 54
3.4 SIMULATED SCENARIOS ... 57
3.5 MODEL VALIDATION ... 58
3.6 KEY RESULTS AND DISCUSSION ... 60
3.7. IMPACT OF PV PANELS ON RADIATION AVAILABILITY ... 61
3.8 INTERACTION BETWEEN PVS, CROPS AND MICROCLIMATE ... 63
CONCLUSIONS ... 64
REFERENCES ... 67
CHAPTER 4. MODELLING OF THE PV MODULES PERFORMANCE ... 69
INTRODUCTION ... 69
4.1 SOLAR IRRADIANCE INSIDE THE GREENHOUSE ... 70
4.2 EVALUATION OF THE PV SYSTEM PERFORMANCE ... 75
4.3 PV PRODUCTION MODELLING ... 79
CONCLUSION ... 87
REFERENCES ... 90
CHAPTER 5. MODELING CROP GROWTH AND YIELD FORMATION ... 91
INTRODUCTION ... 91
5.1. OVERVIEW OF THE AQUACROP MODEL ... 91

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

5.2. AQUACROP SIMULATION AT FSC 96

5.3 A STATISTICAL APPROACH: MULTI-STATE MODELING OF GREENHOUSE CUCUMBER YIELD DYNAMICS UNDER MICROCLIMATE EFFECTS 111

5.4 RESULTS 114

CONCLUSION 115

REFERENCES 117

CHAPTER 6. PHOTOVOLTAIC GREENHOUSE DIGITAL TWIN 120

INTRODUCTION 120

6.1 SYSTEM ARCHITECTURE AND METHODOLOGY 121

6.2 DATA DESCRIPTION AND MANAGEMENT 133

6.3 ENERGY OPTIMIZATION MODEL 134

6.4 THERMODYNAMIC AND ENVIRONMENTAL MODEL 141

6.5 GNN RESULTS AND METHODS 147

6.6 CROP GROWTH MODELING COMPONENT 155

6.7 MULTI-OUTPUT MODELLING OF GREENHOUSE MICROCLIMATE VARIABLES 159

CONCLUSIONS 169

REPOSITORY DESCRIPTION (D4.3) 173

OVERALL CONCLUSIONS AND KEY FINDING 177

ANNEX 179

ANNEX A – DETAILED MATHEMATICAL FORMULATIONS AND UDF DESCRIPTIONS (CHAPTER 3) 179

ANNEX B – COMPUTATIONAL GRIDS AND NUMERICAL PARAMETERS (CHAPTER 3) 188

ANNEX C – MATERIAL THERMOPHYSICAL AND OPTICAL PROPERTIES (CHAPTER 3) 191

ANNEX D – SUPPLEMENTARY FIGURES AND EXTENDED RESULTS (CHAPTER 3) 195

ANNEX E – LIST OF VARIABLES, FEATURES, AND TARGET NAMES OF THE MODELS (CHAPTER 6) 222

ANNEX F – LIST OF VARIBLE USED FOR ALL THE OTHER MODELS (CHAPTER 6) 223

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

List of Figures

Chapter 1

Fig. 1.1 UTH Greenhouse location 18
Fig. 1.2 CAD representation of the UTH greenhouse 18
Fig. 1.3 Sercom sensor's location 19
Fig. 1.4 REGACE sensors system 20
Fig. 1.5 Trisolar sensors system 20
Fig. 1.6 FSC Greenhouse location 21
Fig. 1.7 3D-CAD representation of the FSC greenhouse 22
Fig. 1.8 Greenhouse internal partition 22
Fig. 1.9 Control room 23
Fig. 1.10 Measurement points at FSC (a), sketch of a 3-level measuring point (b), view of the installation (c) 23
Fig. 1.11 Weather station 24
Fig. 1.12 Contour plot of CO_2 and Illumination inside the greenhouse 24
Fig. 1.13 Sketch of the irrigation system 25
Fig. 1.14 Irrigation GUI (left) and real-time monitoring of the irrigation (right) 26
Fig. 1.15 Photovoltaic system sensors 27
Fig. 1.16 AccuPAR LP-80 ceptometer 28
Fig. 1.17 LAI sampling in row crops 28

Chapter 2

Fig. 2.1 View of the greenhouse located at FSC. The shading structures located to the east, south, and west are highlighted 33
Fig. 2.2 GH 3D model developed in IDA ICE 5.1 showing the subdivision into the two main zones and their respective subzones 33
Fig. 2.3 Time trend and carpet plot of air temperature measured by the experimental weather station located at FSC during the simulation period (21 May 2025 25 August 2025) 35
Fig. 2.4 Time trend and carpet plot of relative humidity measured by the experimental weather station located at FSC during the simulation period (21 May 2025 25 August 2025) 35
Fig. 2.5 Time trend of aerodynamic conductance during the simulation period (21 May 2025 25 August 2025) 38
Fig. 2.6 Time trend of canopy conductance during the simulation period (21 May 2025 25 August 2025) 38
Fig. 2.7 Time trend of the Leaf Area Index (LAI) associated with the tomato crop during the simulation period (21 May 2025 25 August 2025) 38
Fig. 2.8 Time trend of the Leaf Area Index (LAI) associated with the eggplant crop during the simulation period (21 May 2025 25 August 2025) 38
Fig. 2.9 Time trend and carpet plot of measured and simulated air temperature for subzone 1.3 of the FSC greenhouse during the simulation period 40
Fig. 2.10 Time trend and carpet plot of measured and simulated air temperature for subzone 2.3 of the FSC greenhouse during the simulation period 41
Fig. 2.11 Scatter plots of simulated versus measured air temperature for subzones 1.2 and 1.3 of the FSC greenhouse 42
Fig. 2.12 Scatter plots of simulated versus measured air temperature for subzones 2.2 and 2.3 of the FSC greenhouse 42
Fig. 2.13 Time trend and carpet plot of measured and simulated relative humidity for Zone 1 of the FSC greenhouse during the simulation period 44
Fig. 2.14 Time trend and carpet plot of measured and simulated relative humidity for Zone 2 of the FSC greenhouse during the simulation period 44
Fig. 2.15 Correlation plots of simulated versus measured relative humidity for subzones 1.2 and 1.3 of the FSC greenhouse 45
Fig. 2.16 Correlation plots of simulated versus measured relative humidity for subzones 2.2 and 2.3 of the FSC greenhouse 45

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

Chapter 3

Fig. 3.1 – Detail of PV panel computational grid. 53

Chapter 4

Fig. 4.1. The profiles of global horizontal irradiance outside and inside the greenhouse (GHI_out and GHI_in) for (a) a clear-sky day and (b) an overcast day for FSC. 71
Fig. 4.2. Scatter plot of GHI_in versus GHI_out at the FSC site, with the fitted linear regression line. 72
Fig. 4.3. Comparing the monthly profiles of the received energy on the horizontal surface inside and outside the greenhouse of the FSC site: (a) Raw values; (b) The specific values, which obtains by dividing the total monthly values by the equivalent days with available irradiance data. 73
Fig. 4.4. Comparison of the observed and PVlib estimated normalized power for the FSC site using the four representative effective irradiance definitions: (a) West GHI_eff; (b) East GHI_eff; (c) Average GHI_eff; (d) Minimum GHI_eff. 74
Fig. 4.5. Comparing the monthly values of PR for FSC, UTH, and ESTER Lab. 76
Fig. 4.6. The contour plots of the clearness index, as the ratio between GTI and clear-sky GTI, for different cases: (a) FSC; (b) UTH; (c) ESTER Lab. 77
Fig. 4.7. The conducted experiments at ESTER Lab to confirm the influence of shadowing on PV power production: (a) Single vertical shadow; (b) Single horizontal shadow; (c) Single diagonal shadow; (d) Double shadowing. 78
Fig. 4.8. The profiles for the normalized power and effective irradiance for the shadow experiments at ESTER Lab: (a) Single vertical shadow; (b) Single horizontal shadow; (c) Single diagonal shadow; (d) Double shadowing. 79
Fig. 4.9. Checking the power estimation accuracy of DBPG2 model compared to PVlib using the test data for: (a, b) FSC; (c, d) UTH; (e, f) ESTER Lab. 83
Fig. 4.10. Comparison of normalized power output profiles throughout the day between measured observations and simulated results obtained from PVlib and DBPG2 models: (a, b) FSC; (c, d) UTH; (e, f) ESTER sites on selected dates. 84
Fig. 4.11. Comparing the monthly values of the observed PR with PVlib and the DBPG2 model estimation for each case. 85

Chapter 5

Fig. 5.1 Chart of AquaCrop indicating the main components of the soil–plant–atmosphere continuum and the parameters driving phenology, canopy cover, transpiration, biomass production, and final yield. 92
Fig. 5.2 Calculation scheme of AquaCrop with indication of the four steps and the processes affected by water and temperature stress. 93
Fig. 5.3 FAO AquaCrop software main menu. 95
Fig. 5.4 Soil map of the Lazio Region. 96
Fig. 5.5 Geological map of the Lazio Region. 96
Fig. 5.6 Reference evapotranspiration for the two plots cultivated with tomato. 99
Fig. 5.7 Reference evapotranspiration for the two plots cultivated with eggplant. 99
Fig. 5.8 Growing Degree Days (GDD) for the two plots cultivated with tomato. 101
Fig. 5.9 Growing Degree Days (GDD) for the two plots cultivated with eggplant. 101
Fig. 5.10 Measured and interpolated Leaf Area Index (LAI) for each of the two tomato plots. 102
Fig. 5.11 Measured and interpolated Leaf Area Index (LAI) for each of the two eggplant plots. 102
Fig. 5.12 Canopy Cover (CC) for the tomato reference plot. 103
Fig. 5.13 Canopy Cover (CC) for the tomato PV plot. 103
Fig. 5.14 Canopy Cover (CC) for the eggplant reference plot. 103
Fig. 5.15 Canopy Cover (CC) for the eggplant PV plot. 103
Fig. 5.16 Cumulative fresh yield for the tomato reference plot: observed vs simulated. 107
Fig. 5.17 Correlation between observed and simulated cumulative fresh yield for the tomato reference plot. 107
Fig. 5.18 Cumulative fresh yield for the tomato PV plot: observed vs simulated. 108
Fig. 5.19 Correlation between observed and simulated cumulative fresh yield for the tomato PV plot. 108

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Fig. 5.20 Cumulative fresh yield for the eggplant reference plot: observed vs simulated. 110
Fig. 5.21 Correlation between observed and simulated cumulative fresh yield for the eggplant reference plot. 110
Fig. 5.22 Cumulative fresh yield for the eggplant PV plot: observed vs simulated. 110
Fig. 5.23 Correlation between observed and simulated cumulative fresh yield for the eggplant PV plot. 111
Fig. 5.24 Conceptual three-state space for daily cucumber yield (low, medium, high) with bidirectional transitions. 112
Fig. 5.25 Correlation patterns among detrended within-greenhouse anomalies. 113
Fig. 5.26 Example of daily yield trajectories showing plateaus and thresholds used to discretize yield states. 114
Fig. 5.27 Pooled model: 30-day transition probabilities between yield states under average covariate conditions. 115

Chapter 6

Fig. 6.1 Scheme of the methodology architecture 129
Fig. 6.2 Scatter plot for the E6 XGBoost model for PV Power prediction 140
Fig. 6.3 Time series plot for the E6 XGBoost model for PV Power prediction for three days in September/October 2025 141
Fig. 6.4 Engineered features for the low-cost thermodynamic model 143
Fig. 6.1: Scatter plot for the LC1 XGBoost model for greenhouse internal temperature prediction. 146
Fig. 6.6 Conceptual scheme of the GNN 148
Fig. 6.7 Scatter plot for the F GNN model for greenhouse internal temperature prediction 154
Fig. 6.8 Scatter plot for the H GNN model for greenhouse internal temperature prediction 154
Fig. 6.9 Design of the greenhouse with the poles and the associated numbers 160
Fig. 6.10: Scatter plots for the 8 temperature sensors in the greenhouse poles. 161
Fig. 6.11: Scatter plots for the 8 relative humidity sensors in the greenhouses poles. 161
Fig. 6.12: Scatter plots for the 8 CO_2 sensors in the greenhouse poles. 162
Fig. 6.13: Scatter plots for the 8 PAR sensors in the greenhouse poles. 162
Fig. 6.14: Scatter plots for the 4 Power production predictions. 163
Fig. 6. 1 Scatter plots for temperature, relative humidity, CO_2 and PAR for pole 3 level adjustable. 166
Fig. 6.2 Scatter plots for temperature, relative humidity, CO_2, PAR for pole 3 level adjustable and Power. 168

Appendix A

Fig. A.1 Transfer equation operation 181

Appendix B

Fig. B.1 2D grid of 6 chambers cross section. 188
Fig. B.2 2D grid of 2 chambers with the PVs horizontal. 188
Fig. B.3 2D grid of 2 chambers with PVs inclined to the east. 189
Fig. B.4 2D grid of 2 chambers with PVs inclined to the west. 189
Fig. B.5 2D grid of 2 chambers with plants. 190
Fig. B.6 2D grid of 2 chambers with plants and horizontal PV panels. 190
Fig. B.7 3D grid of one chamber with plants and horizontal PV panels. 190

Appendix D

Fig. D.1a Comparison of measured and simulation predicted PAR in the location of sensors I4t, 09/06/24. 195
Fig. D.2 Comparison of measured and simulation predicted PAR in the location of sensors I2, 09/06/24. 195
Fig. D.3 Comparison of measured and simulation predicted PAR in the location of sensors I4s, 09/06/24. 196
Fig. D.4 Comparison of measured and simulation predicted PAR in the location of sensors I3, 09/06/24. 196
Fig. D.5 Comparison of measured and simulation predicted PAR in the location of sensors PAR8, 17/12/24. 196
Fig. D.6 Comparison of measured and simulation predicted PAR in the location of sensors 197

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

PAR11, 17/12/24.
Fig. D.7 Comparison of measured and simulation predicted PAR in the location of sensors P35, 17/12/24. 197
Fig. D.8 Comparison of measured and simulation predicted PAR in the location of sensors P5, 17/12/24. 197
Fig. D.9 Comparison of measured and simulation predicted PAR+NIR in the location of sensors I2, 18/04/25. 198
Fig. D.10 Comparison of measured and simulation predicted PAR+NIR in the location of sensors I4s, 18/04/25. 198
Fig. D.11 Comparison of measured and simulation predicted PAR+NIR in the location of sensors I3, 18/04/25. 198
Fig. D.12 Comparison of measured and simulation predicted PAR in the location of sensors P35, 18/04/25. 199
Fig. D.13 Comparison of measured and simulation predicted PAR in the location of sensors P39, 18/04/25. 199
Fig. D.14 Comparison of measured and simulation predicted PAR in the location of sensors PAR10, 18/04/25. 199
Fig. D.15 Comparison of measured and simulation predicted PAR in the location of sensors P5, 18/04/25. 200
Fig. D.16 Comparison of measured and simulation predicted PAR in the location of sensors P9, 18/04/25. 200
Fig. D.17 Comparison of measured and simulation predicted PAR in the location of sensors PAR11, 18/04/25. 200
Fig. D.18 Comparison of measured and simulation predicted PAR+NIR in the location of sensors I2, 14/05/24. 201
Fig. D.19 Comparison of measured and simulation predicted PAR+NIR in the location of sensors I4s, 14/05/24. 201
Fig. D.20 Comparison of measured and simulation predicted PAR+NIR in the location of sensors I4t, 14/05/24. 201
Fig. D.21 Comparison of measured and simulation predicted PAR+NIR in the location of sensors I4t, 14/05/24. 202
Fig. D.22 PAR isocontours for 09/06/24 without PV, without crops and without shading consideration at 10.00 [h]. 202
Fig. D.23 PAR isocontours for 09/06/24 without PV, without crops and without shading consideration at 12.00 [h]. 202
Fig. D.24 PAR isocontours for 09/06/24 without PV, without crops and without shading consideration at 16.00 [h]. 202
Fig. D.25 PAR isocontours for 09/06/24 without PV, without crops and without shading consideration at 18.00 [h]. 203
Fig. D.26 PAR isocontours for 17/12/24 without PV, without crops and without shading consideration at 10.00 [h]. 203
Fig. D.27 PAR isocontours for 17/12/24 without PV, without crops and without shading consideration at 12.00 [h]. 203
Fig. D.28 PAR isocontours for 17/12/24 without PV, without crops and without shading consideration at 14.00 [h]. 203
Fig. D.29 PAR isocontours for 17/12/24 without PV, without crops and without shading consideration at 16.00 [h]. 204
Fig. D.30 PAR isocontours for 09/06/24 without PV, without crops accounting shading at 10.00 [h]. 204
Fig. D.31 PAR isocontours for 09/06/24 without PV, without crops accounting shading at 12.00 [h]. 204
Fig. D.32 PAR isocontours for 09/06/24 without PV, without crops accounting shading at 16.00 [h]. 204
Fig. D.33 PAR isocontours for 09/06/24 without PV, without crops accounting shading at 18.00 [h]. 205
Fig. D.34 PAR isocontours for 17/12/24 without PV, without crops accounting shading at 10.00 [h]. 205
Fig. D.35 PAR isocontours for 17/12/24 without PV, without crops accounting shading at 12.00 [h]. 205

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Fig. D.36 PAR isocontours for 17/12/24 without PV, without crops accounting shading at 14.00 [h]. 205
Fig. D.37 PAR isocontours for 17/12/24 without PV, without crops accounting shading at 16.00 [h]. 206
Fig. D.38 Differentiation at predicted PAR radiation due to shading consideration for the 09/06/24. 206
Fig. D.39 Differentiation at predicted PAR radiation due to shading consideration for the 17/12/24. 206
Fig. D.10 PAR isocontours for 18/04/25 with PV at horizontal position, without crops accounting shading at 10.00 [h]. 207
Fig. D.41 PAR isocontours for 18/04/25 with PV at horizontal position, without crops accounting shading at 12.00 [h]. 207
Fig. D.42 PAR isocontours for 18/04/25 with PV at horizontal position, without crops accounting shading at 16.00 [h]. 207
Fig. D.43 PAR isocontours for 18/04/25 with PV at horizontal position, without crops accounting shading at 18.00 [h]. 207
Fig. D.44 PAR isocontours for 18/04/25 with PV inclined, without crops accounting shading at 10.00 [h]. 208
Fig. D.45 PAR isocontours for 18/04/25 with PV inclined, without crops accounting shading at 12.00 [h]. 208
Fig. D.46 PAR isocontours for 18/04/25 with PV inclined, without crops accounting shading at 16.00 [h]. 208
Fig. D.47 PAR isocontours for 18/04/25 with PV inclined, without crops accounting shading at 18.00 [h]. 208
Fig. D.48 PAR isocontours for 17/12/24 with PV at horizontal position, without crops accounting shading at 10.00 [h]. 209
Fig. D.49 PAR isocontours for 17/12/24 with PV at horizontal position, without crops accounting shading at 12.00 [h]. 209
Fig. D.50 PAR isocontours for 17/12/24 with PV at horizontal position, without crops accounting shading at 16.00 [h]. 209
Fig. D.51 PAR isocontours for 17/12/24 with inclined PV, without crops accounting shading at 10.00 [h]. 209
Fig. D.52 PAR isocontours for 17/12/24 with inclined PV, without crops accounting shading at 12.00 [h]. 210
Fig. D.53 PAR isocontours for 17/12/24 with inclined PV, without crops accounting shading at 14.00 [h]. 210
Fig. D.54 PAR radiation variation with horizontal PV at time zone 12h 18/04/25. 210
Fig. D.55 PAR radiation variation with inclined PV at time zone 12h 18/04/25. 211
Fig. D.56 PAR radiation variation with horizontal PV at time zone 12h 17/12/25. 211
Fig. D.57 PAR radiation variation with horizontal PV at time zone 12h 17/12/25. 211
Fig. D.58 PAR radiation variation at height 2 m across chamber at 10h of 18/04/25. 212
Fig. D.59 PAR radiation variation at height 2 m across chamber at 10h of 17/12/25. 212
Fig. D.60 PAR isocontours for 14/05/24 with cucumber plants, without PV panels at 10.00 [h]. 212
Fig. D.61 PAR isocontours for 14/05/24 with cucumber plants, without PV panels at 12.00 [h]. 213
Fig. D.62 PAR isocontours for 14/05/24 with cucumber plants, without PV panels at 16.00 [h]. 213
Fig. D.63 PAR isocontours for 14/05/24 with cucumber plants, without PV panels at 18.00 [h]. 213
Fig. D.64 Temperature isocontours for 14/05/24 with cucumber plants, without PV panels at 10.00 [h]. 214
Fig. D.65 Temperature isocontours for 14/05/24 with cucumber plants, without PV panels at 12.00 [h]. 214
Fig. D.66 Temperature isocontours for 14/05/24 with cucumber plants, without PV panels at 16.00 [h]. 215
Fig. D.67 Temperature isocontours for 14/05/24 with cucumber plants, without PV panels at 18.00 [h]. 215
Fig. D.68 Velocity magnitude isocontours for 14/05/24 with cucumber plants, without PV panels at 12.00 [h]. 215
Fig. D.69 Streamlines for 14/05/24 with cucumber plants, without PV panels at 12.00 [h]. 216
Fig. D.70 PAR isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 10.00 [h]. 216

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Fig. D.71 PAR isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 12.00 [h]. 216
Fig. D.72 PAR isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 14.00 [h]. 217
Fig. D.73 PAR isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 16.00 [h]. 217
Fig. D.74 Temperature isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 10.00 [h]. 217
Fig. D.75 Temperature isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 12.00 [h]. 218
Fig. D.76 Temperature isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 14.00 [h]. 218
Fig. D.77 Temperature isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 16.00 [h]. 218
Fig. D.78 Velocity magnitude isocontours and streamlines for 11/10/24 with cucumber plants, and PV panels in horizontal position at 10.00 [h]. 219
Fig. D.79 Velocity magnitude isocontours and streamlines for 11/10/24 with cucumber plants, and PV panels in horizontal position at 12.00 [h]. 219
Fig. D.80 Velocity magnitude isocontours and streamlines for 11/10/24 with cucumber plants, and PV panels in horizontal position at 14.00 [h]. 219
Fig. D.81 Velocity magnitude isocontours and streamlines for 11/10/24 with cucumber plants, and PV panels in horizontal position at 16.00 [h]. 220
Fig. D.82 Velocity magnitude isocontours and streamlines for 18/04/25 11h with inclined PV and without plants. 220

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

List of Tables

Chapter 2

Table 2.1 Temperature correlation parameters calculated for each analyzed subzone 41
Table 2.2 Relative humidity correlation parameters calculated for each analyzed subzone 43

Chapter 3

Table 3.1 Timetable of UTH greenhouse operation 50
Table 3.2. Coefficients of regression analysis 52
Table 3.3 CFD simulations validation days 57
Table 3.4 External climatic data 28/04/24 57
Table 3.5 External climatic data 17/12/24 58

Chapter 4

Table 4.1. The performance evaluation of the PV system in FSC site. 75
Table 4.2. The performance evaluation of the PV system in UTH site. 76
Table 4.3. The performance evaluation of the PV system in ESTER Lab. 76
Table 4.4. Comparing the observed PR for the whole period with PVlib and the DBPG2 model estimation for each case. 87

Chapter 5

Table 5.1 Default soil characteristics based on AquaCrop Handbook I. 94
Table 5.2 Soil pedological analyses for the FSC site (gravimetric water contents). 97
Table 5.3 Soil pedological analyses for the FSC site (volumetric water contents). 97
Table 5.4 Input parameters used in simulating tomato response with AquaCrop. 105
Table 5.5 Input parameters used in simulating eggplant response with AquaCrop. 105
Table 5.6 Performance metrics of the fresh yield simulation for the tomato reference plot. 107
Table 5.7 Performance metrics of the fresh yield simulation for tomato plots (reference vs PV). 109
Table 5.8 Performance metrics of the fresh yield simulation for the eggplant reference plot. 109
Table 5.9 Performance metrics of the fresh yield simulation for eggplant plots (reference vs PV). 111
Table 5.10 Greenhouse code and treatment description. 113

Chapter 6

Table 6.1 Progressive experiments for greenhouse temperature prediction 139
Table 6.2 Performance of thermodynamic and environmental models 146
Table 6.3 Ablation study of GNN-based digital twin components 153
Table 6.4 Crop growth modeling experimental results 157
Table 6.5 Performance metrics for all pole pairs and target variables 164
Table 6.6 Performance metrics for the XGBoost "Only Microclimate" model 166
Table 6.7 Summary of limitations and lessons learned 167
Table 6.8 Summary of key achievements by component 170
Table 6.9 Summary of limitations and lessons learned 171

Appendix A

Table A.1 Characteristics of 12 arc constitute the six chambers roof arcs 180
Table A.2 181

Appendix C

Table C.1 Materials thermophysical properties [1-8] 191
Table C.2 PV structure I and thermophysical properties (Baxevanou 2020) 191
Table C.3 PV structure II and thermophysical properties 192
Table C.4 Thermophysical properties of plants parts (Cengel Jiayu Zhang 2025, Mousavizadeh 2010) 192
Table C.5 Materials optical properties [9-15] 193
Table C.6 PV components extinction coefficients 193
Table C.7 PV components refractive index 193

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Table C.8 Simulation Effective thermophysical properties 194
Table C.9 Simulation effective material optical properties 194

Appendix D

Table D.1 PV solar cell composition 195

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Executive Summary

The activities carried out in our project have focused on developing advanced modelling techniques for a novel photovoltaic (PV) greenhouse system. This effort spans multiple tasks, each addressing a specific aspect of the system, from microclimate modelling to the integration of various models into a comprehensive Digital Twin. As also outlined in the work package description of the proposal, Crop growth modelling has been implemented using Aquacrop and validated with data coming from the pilot testing activity for FSC site while an innovative statistical approach has been applied to UTH site. Below is a summary of the main activities and achievements at the end of the project:

Dynamic Modelling of the Greenhouse Microclimate and Energy Demand

- Developed a dynamic model using advanced software for Dynamic Building Simulation (DBS) to simulate greenhouse microclimates.
- Integrated physical processes such as evapotranspiration and plant growth during time through look up tables into the model.
- Calibrated and validated the model with data for UTH and FSC sites enhancing understanding of thermo-hygrometric behaviour and energy demand.

CFD Modelling of the Greenhouse Microclimate

- Structured CFD modelling at two levels: radiation transport and internal microclimate.
- Developed 2D models for radiation transport through greenhouse covers and validated them.
- Created 3D models for internal microclimate, incorporating heat exchange, plant interaction, and CO2 distribution.
- Conducted extensive parametric studies.

Modelling of the PV Modules Performance

- Reviewed available PV performance models for bifacial technologies and developed a semi-empirical performance model (DBPG2).
- Used common practice model (PVlib) to first evaluation of PV performance to compare with data available for FSC and UTH sites
- Calibrated the model using PV module parameters and validated it with data for FSC and UTH sites.
- Compare the performance of DBPG2 model with common practice tools (PVlib).

Digital Twins

- Integrated outputs from previous tasks into a comprehensive empirical model based on machine learning (ML).
- Utilized Recurrent Neural Networks (RNNs) and XGBoost techniques for capturing temporal dependencies in microclimate data.
- Explored the potential of Graph Neural Networks (GNNs) to handle complex interactions within the PV greenhouse system.
- Developed models that predict the general behavior of the PV/greenhouse with appropriate uncertainty levels.
- Criticism and recommendations discussed for future developments in real world applications

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

Main Results

- **Dynamic Modeling**: Successfully developed and validated a dynamic model for greenhouse microclimates, incorporating key physical processes and calibrated with real-world data.
- **CFD Modeling**:
  - Completed 2D and 3D CFD models for radiation transport and internal microclimate.
  - Conducted detailed parametric studies to improve model accuracy
- **PV Modules Performance**:
  - Developed a semi-empirical performance model for bifacial PV modules.
  - Achieved significant accuracy in predicting PV performance by accounting for multiple environmental factors.
- **Digital Twins**:
  - Integrated various model outputs into a unified ML-based empirical model.
  - Demonstrated strong predictive capabilities using RNNs and XGBoost techniques for greenhouse temperature and humidity (simpler models)
  - Successfully completed a Spatio Temporal GNN (STGNN) validated at UTH and FSC sites testing different approaches of network construction

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Introduction

This deliverable presents the progress and achievements in developing advanced modelling frameworks for photovoltaic (PV) greenhouse systems, with the overall objective of improving understanding, prediction, and optimization of their environmental, energy, and productive performance. The work builds on the methodological vision outlined in the project proposal, whereby complementary deterministic and data-driven approaches are combined to capture the intrinsic complexity of PV-integrated greenhouses, where microclimate dynamics, energy flows, crop processes, and resource efficiency are tightly interconnected.

Owing to the high heterogeneity of data at the different pilot site representing real world operation the modelling framework is implemented and validated on the two pilot sites provided with sensors and monitoring systems able to produce reliable datasets, these are the FSC and UTH pilot greenhouses.

Several complementary deterministic models have been developed, building on established approaches in literature and adapted to the specific characteristics of PV greenhouse systems. These models are calibrated and validated using experimental datasets and continuous monitoring data from operational pilots, allowing a deeper understanding of system behaviour in terms of microclimate regulation, energy production and consumption, water use, and crop development.

A dynamic model of greenhouse microclimate and energy demand has been implemented using advanced building performance simulation tools (IDA ICE) and validated against real operational data, enabling accurate representation of thermo-hygrometric behaviour and supporting the assessment of control strategies and design options. In line with the original methodological concept, the crop processes are firstly addressed through the established model AquaCrop, testing and validating the crop growth on a specific crop cycle at the FSC pilot greenhouse. High-resolution simulations based on Computational Fluid Dynamics, which provide deeper insight into radiation transport, airflow patterns, and heat exchange processes in complex greenhouse geometries are investigated, tested and validated at the UTH pilot site to support the interpretation and refinement of the overall modelling framework.

The photovoltaic subsystem is represented through a dedicated performance model for the bifacial PV modules integrated into the greenhouse structures. Experimental characterization performed in laboratory conditions, together with long-term monitoring data collected at pilot sites, supports the development of a novel semi-empirical model suited for the specific features of bifacial technologies. This enables reliable prediction of PV energy yield under realistic operating conditions.

Environmental, agronomic, and energy data collected at the pilots, together with the insight gained from the deterministic model outputs, contributed to the development of the REGACE Digital Twin.

This is supported by data-driven modelling approaches based on machine learning, including recurrent neural networks (RNN) for time-dependent processes, XGboost network and graph neural networks (GNN) for capturing complex interactions between system components. The resulting framework is designed to provide predictive capability across the full chain from solar radiation and environmental control to energy production and crop performance.

Overall, the activities presented in this deliverable establish a solid and coherent methodological foundation for the development of reliable, scalable, and transferable models of PV greenhouse systems.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Chapter 1. Data Description and Management

Introduction

This chapter provides a comprehensive description of the data sources, monitoring infrastructure, and data management procedures adopted within the FSC and UTH pilot greenhouses, developed in the framework of the Project. The objective of this chapter is to document the experimental setup and the associated datasets in a transparent and reproducible manner, forming the basis for the subsequent analyses on microclimate dynamics, irrigation efficiency, photovoltaic (PV) system performance, and crop response in an agrivoltaic context.
The FSC and UTH greenhouses represent a complex experimental system in which environmental, hydraulic, photovoltaic, and agronomic components are tightly integrated. To properly assess the interactions between these subsystems, a high-resolution, multi-layer monitoring strategy was implemented. This includes continuous measurements of greenhouse microclimate and external meteorological conditions, detailed monitoring of soil moisture and irrigation processes, real-time acquisition of PV system performance metrics, and periodic collection of crop growth and yield data.
Given the heterogeneity of the data in terms of temporal resolution, physical units, and measurement technologies, particular attention was devoted to data acquisition protocols, synchronization, quality control, and preprocessing. This chapter therefore not only describes the physical characteristics of the experimental site and sensor deployment but also outlines the data handling workflows adopted to ensure consistency, reliability, and suitability for modelling and performance assessment.
Overall, this chapter establishes the methodological foundation for the integrated analysis of energy, water, and crop productivity carried out in the following sections, with specific emphasis on the dual objectives of agronomic sustainability and renewable energy optimization in greenhouse agrivoltaic systems.

1.1 UTH greenhouse at “farm of Velestino” (GREECE)

1.1.1. Location

The Greek greenhouse is located in Thessaly (Greece) at the farm of Velestino (UTH), in central Greece (39° 23' 41.14" N; 22° 45' 29.16" E), see Fig. 1.1.
The experimental greenhouse is consisted of 6 chambers of 25 m x 9.6 m with gothic roof with gutter height 5 m and top height 7.35 m, oriented NE-SW, as shown in Fig. 1.2
The 1st chamber is the engine room. For the REGACE project chambers 3, 4 and 5 were used. Chamber 3 is the conventional room. In the chamber 4 and 5 PV panels were installed. In each chamber were installed 96 bifacial monocrystalline PV panels of 75 W (0.53 m x 1.1 m) with a total installed power of 7.2 kW in each chamber. In chamber 5 also CO_2 enrichment system was installed too. PV panels were installed above the height of shading curtain mounted on a system that can vary their inclination according to the sun position and the crop needs. Inside the greenhouse there are also benches for hydroponic cultivation (6 rows of 19 m in each chamber).

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

Fig. 1.1. UTH Greenhouse location

During the REGACE experiments cucumber cultivation was grown. The greenhouses are heated with a hot water pipe system while an evaporative pad with axial fans tunnel operation is used for cooling. A shading curtain is used for the control of radiation and temperature along with cover openings; however, their operation is not considered in present simulations. In the Fig. 1.2 the plan of the considered greenhouse is presented.

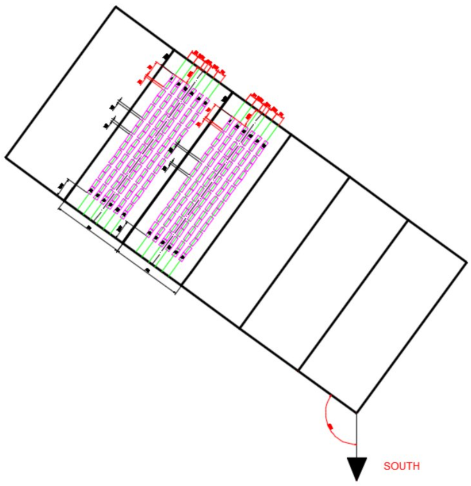


Fig. 1.2. CAD representation of the UTH Greenhouse

1.1.2 - UTH microclimate and PV sensors equipment and location

Three sets of sensors were used for measurement of internal microclimate:

- Sercom system already installed in all the chambers for their automated operation,
- New REGACE system operated after 08/03/24 in the chambers 3 and 4,
- Trisolar sensors operated after 11/09/2024 operated in chambers 4 and 5.

The 'Sercom' system is consisted of 4 stations each one measuring PAR radiation (PAR8, PAR9, PAR10 and PAR11), relative humidity (RT08, TR09, RT10 and RT11), CO_2 concertation (C-8, C09, C10 and C11), and temperature of each chamber in positions shown in Fig. 1.3. The 'Sercom' system is complemented by an external meteorological station which

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

measures total and diffuse radiation, wind direction and speed, CO_2 concentration, temperature and relative humidity.

The new REGACE system focus on measurements on chamber 4 with 5 stations (st2, st3, st4, st5 and st6) there and only one station in chamber 3 (st 7). Each station measures radiation (400 – 1100 nm) (I2, I3, I4s, I4t, d4, I5, I6, I7), temperature and relative humidity (RT2, RT3, RT4, RT5, RT6, RT7) and anemometers that measure the wind direction and the wind speed (A2, A3, A4, A5, A6 and A7). Specially in the station 4 there exist three pyranometers two total radiation pyranometers in different positions (the I4s under the shading curtain and the i4t above the shading curtain) and one pyranometer that measures diffusive radiation. The REGACE system sensors are presented in Fig. 1.4. One more station (st1) is installed outside the greenhouses which measures total and diffuse radiation, wind direction and speed, CO2 concentration, and temperature and relative humidity.

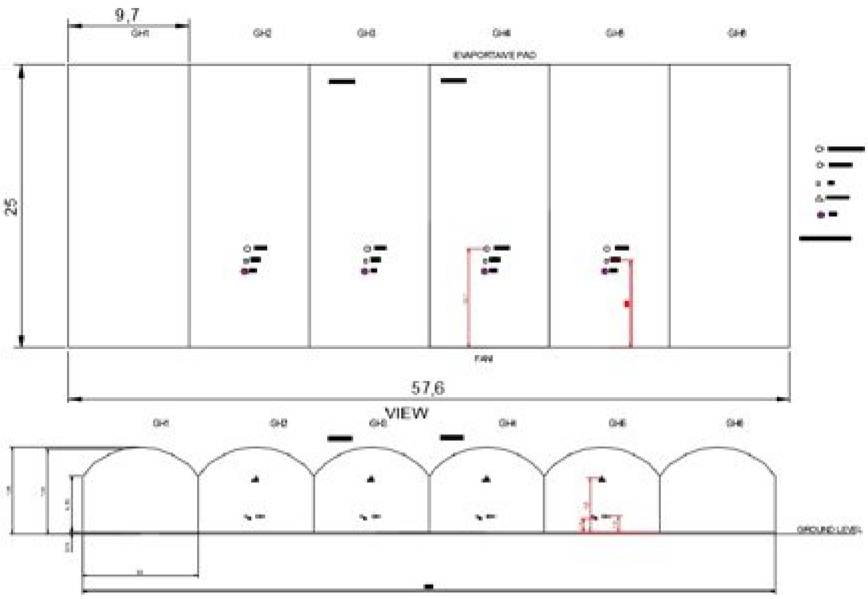
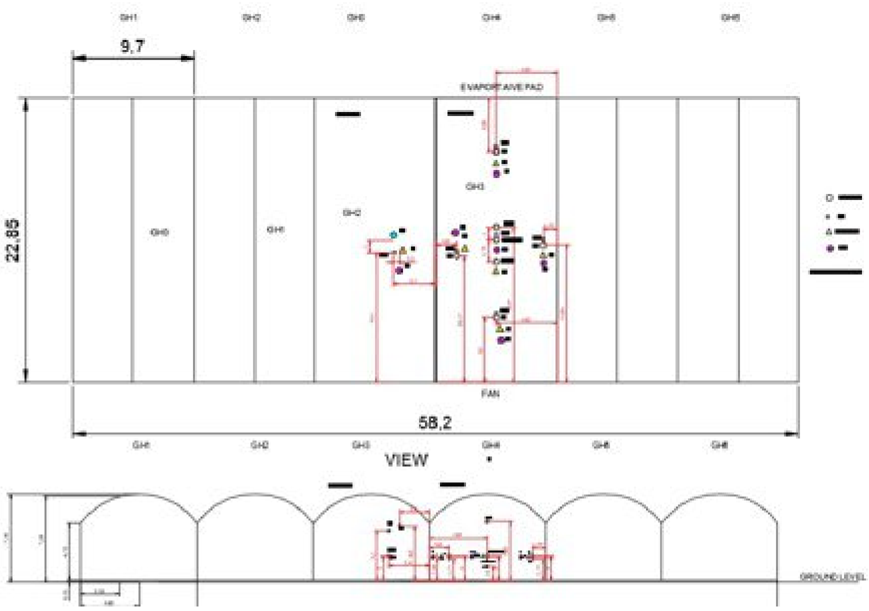


Fig. 1.3. Sercom sensor's location



Funded by the European Union
by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Fig. 1.4. REGACE sensors system

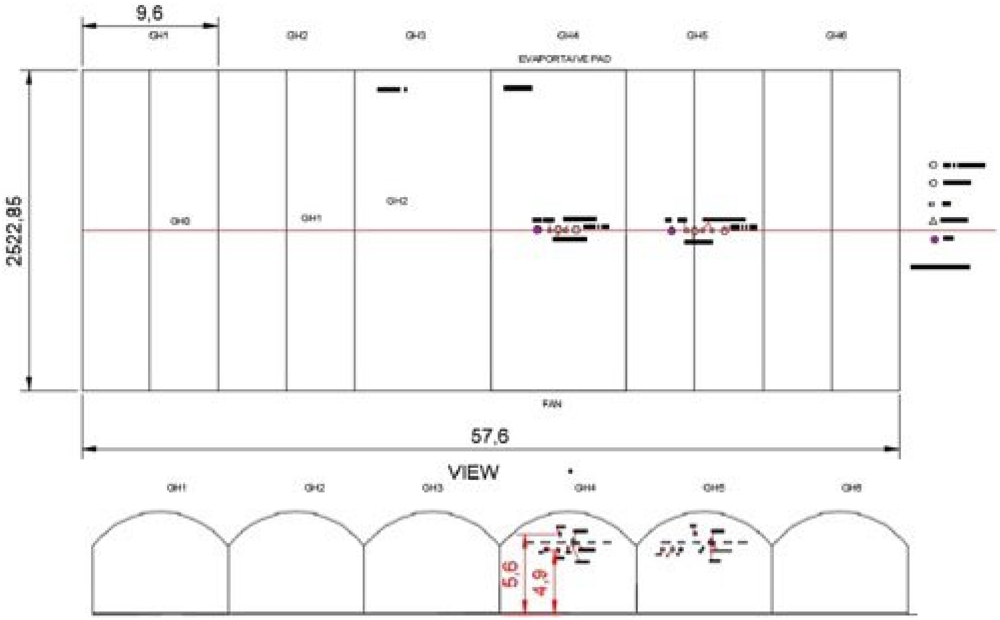


Fig. 1.5. Trisolar sensors system

The ‘Trisolar’ system focuses on measuring the PAR radiation, temperature and CO_2 concentration in the chambers with the PV installation (chamber 4 and chamber 5). In chamber 4 they measure relative humidity at one position (HR t-36), CO_2 concentration at one point (CO2-t36), panel and sun angle, air temperature at one point (t36) and on a PV surface (t46-surf) and radiation above the PVs (P325) under the PVs (P39) on the upper PV surface (P32-sur) and on the lower PV surface (P34-sur). In the chamber 5 they measure CO2 in two points (CO2-t7, CO2-t14), relative humidity in one point (RH-t7), air temperature at one point (t6) and PV surface temperature at two points (t15-surf and t-16surf), PAR radiation at three points above and under PVs (P2, P5 and P9) and on the upper PV surface (P3-sur) and on the lower PV surface (P4-sur). The position of ‘Trisolar’ sensors is given in Fig. 1.5.
Each greenhouse compartment is equipped with:

- a continuous roof vent opened whenever the greenhouse air temperature exceeded 21°C during the day or 18°C during the night, or whenever the greenhouse air relative humidity exceeded 95% during the night or 87% during the day (max vent opening for dehumidification of 10%);
- a pad (15 cm thick, 9.6 m width and 2.0 m height) and fan (capacity of 35 000 m3 h^{-1}, 1.1 kW) system operating whenever the air temperature exceeded 25ºC,
- a pipe rail heating system (12 pipes of 5.1 cm diameter) operating to maintain a greenhouse air temperature of 18°C during the day and 14°C during the night (max pipe temperature set at 60°C),
- a thermal/shading screen with light transmission of about 50% and energy saving of 60% used during the night reducing mainly thermal radiation heat transfer, whenever the outside air temperature was 2°C lower than the heating temperature set point or during the day whenever the outside solar radiation exceeded 750 W m^{-2}.

Each greenhouse compartment is equipped with six hydroponic gutters of 20 m each located 0.5 m above the ground in 1.6 m between each other, holding 19 perlite slabs of 35 L each

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

(average substrate water content of 25-30%) the substrates used for rooting/cultivation of the crop. A drip irrigation system was used for the crop fertigation needs with 5 drippers per slab. External environmental data were obtained through on-site measurements (referred to year 2020). These data include parameters such as temperature, humidity, radiation, wind speed and direction, and rainfall.

1.2 FSC greenhouse at "Fattoria Solidale del Circeo" (ITALY)

1.2.1 Location

The Italian greenhouse is located in Pontinia (Latina, Italy) at the Fattoria Solidale del Circeo (FSC), in central-southern Italy (41° 23' 44.85" N; 13° 8' 57.68" E), see Fig. 1.6.

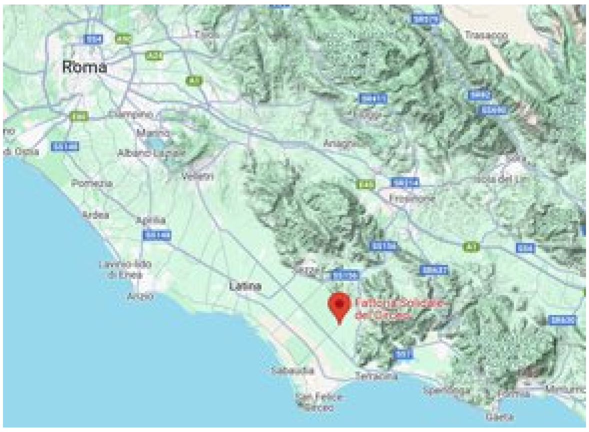


Fig. 1.6. FSC Greenhouse location

The greenhouse is 18x10 m and a maximum height of approximately 5.2 m (Fig. 1.7). It is divided into two equal zones (EAST and WEST) based on its longitudinal orientation.

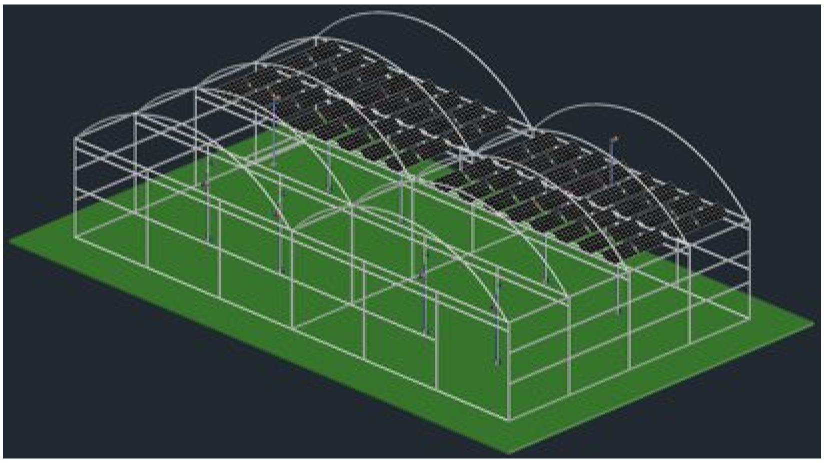


Fig. 1.7. – 3D-CAD representation of the FSC greenhouse

Each zone covers 90 m² and is further subdivided into two sectors: one shaded by the PV panels and one unshaded, which serves as a reference. This results in four distinct sectors: EAST_PV (PV E), EAST_no_PV (REF E), WEST_PV (PV W), and WEST_no_PV (REF W), see Fig. 1.8.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

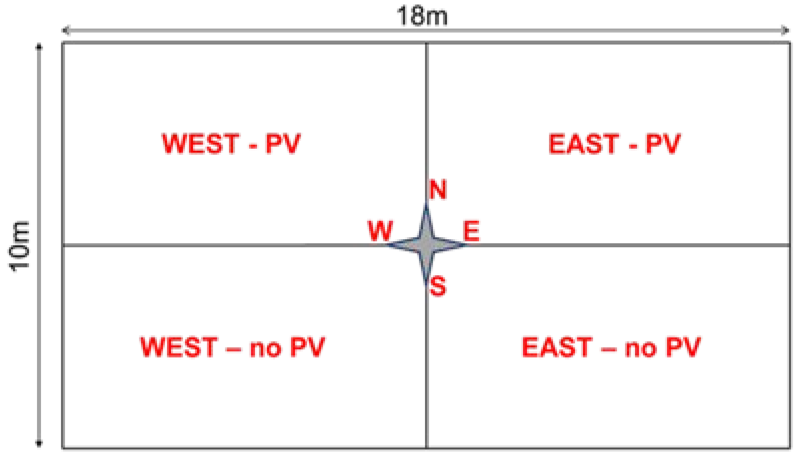


Fig. 1.8. Greenhouse internal partition

Two different crops are cultivated simultaneously, one in the EAST zone and one in the WEST zone. To assess the impact of the PV system on plant growth and productivity, the unshaded sectors serve as control references for evaluating the influence of photovoltaic coverage on the adjacent PV-shaded sectors.
The monitoring infrastructure is organized in 3 independent part:

- Internal greenhouse microclimate and external meteorological data;
- Soil moisture, irrigation control, and hydraulic parameters;
- PV performance metrics and actuator feedback for tracking.

The sensors installed in the greenhouse are provided by Renke and comprehend temperature, relative humidity, CO_2 concentration, photosynthetically active radiation, wind speed, luxmeter, pyranometer, and soil properties.
These sensors are connected to two different laptops via RS485 Modbus connection, acquired using custom program developed in MATLAB and Python (Fig. 1.9), the data are back upped daily at midnight and copied in a shared folder.

Fig. 1.9. Control room

1.2.2 - FSC microclimate sensors equipment and location

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

For the microclimate monitoring eight primary sensor nodes were deployed, two per sector, each stratified vertically across three layers (Fig. 1.10): ground level, canopy (adjustable) level, and below-PV.
Two supplementary nodes (positions 9 and 10) are located above the PV panels to quantify ambient parameters above PV panels

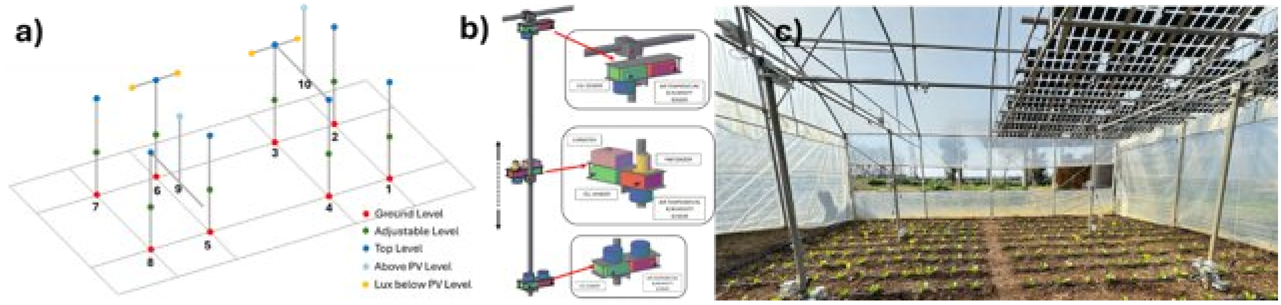


Fig. 1.10. Measurement points at FSC (a), sketch of a 3-level measuring point (b), view of the installation (c)

The sensors installed for the greenhouse microclimate and environment monitoring are:

- Ground level: CO_2, Temperature and Relative Humidity;
- Adjustable level: CO_2, Temperature, Relative Humidity, PAR and Luxmeter;
- Below PV level: CO_2, Temperature and Relative Humidity;
- Above PV level: CO_2, Temperature, Relative Humidity, Luxmeter and Pyranometer (GHI and RHI);
- Environmental parameters (Fig. 1.11): CO_2, Temperature, Relative Humidity, Luxmeter, Anemometer and Pyranometer (GHI);

Fig. 1.11. - Weather station

Data acquisition is continuous and processed using custom program developed in MATLAB every 5 minutes stored in 7 different files (one for each sector, two for the above PV level and one for the environmental parameter).

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

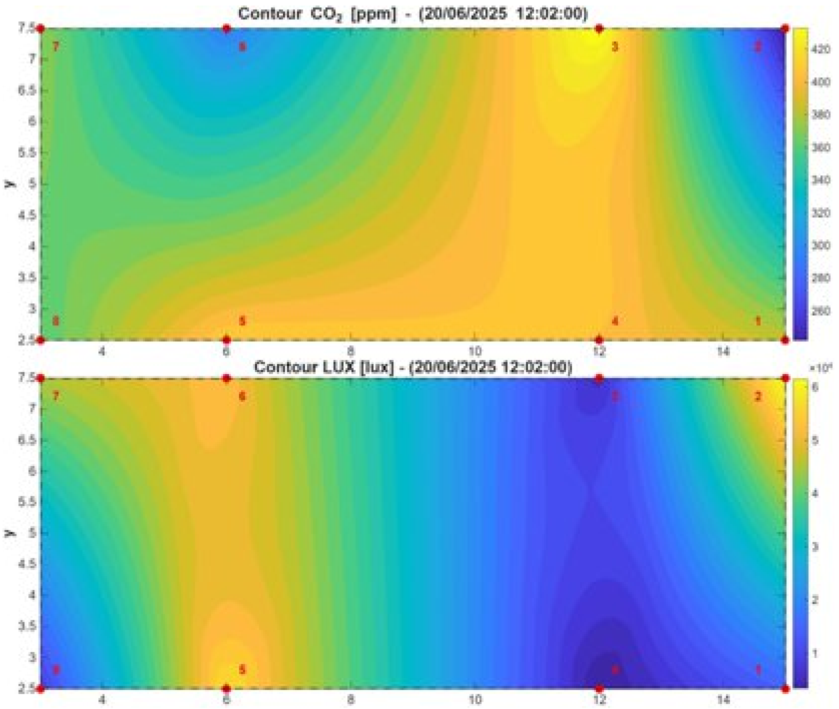


Fig. 1.12. Contour plot of CO2 and Illumination inside the greenhouse

This configuration enables continuous, multi-level monitoring of key environmental variables across the plant growth cycle. By maintaining the adjustable level just above the crop canopy, the system ensures precise quantification of microclimatic parameters, particularly those related to light availability (LUX, PAR), which directly influence photosynthetic activity and biomass accumulation.

An example of the microclimate monitoring is shown in Fig. 1.12 where the CO2 and LUX contour plot is reported.

1.2.3 - Irrigation data and sensors

In the greenhouse a drip irrigation system was implemented to ensure optimized water distribution with minimal waste. The system architecture is subdivided into independent irrigation sectors (Fig. 1.13), each instrumented with the following components:

- soil temperature and relative humidity sensors (4 units total): Strategically positioned in the root zone to enable accurate evaluation of soil moisture conditions and mitigate the risk of hydric stress (red dots in Fig. 1.13);
- a flow and temperature sensor measures the volumetric flow rate and temperature of irrigation water supplied to each sector (blue dots in Fig. 1.13);
- electro-actuated solenoid valve regulates the hydraulic line for each sector, enabling localized control of irrigation events.
- pressure sensor: monitors inlet pressure to assess system integrity and detect anomalies (black dot in Fig. 1.13);
- combined flow and temperature sensor provides baseline flow and temperature measurements for the entire system to support diagnostic and performance assessments (black dot in Fig. 1.13).

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

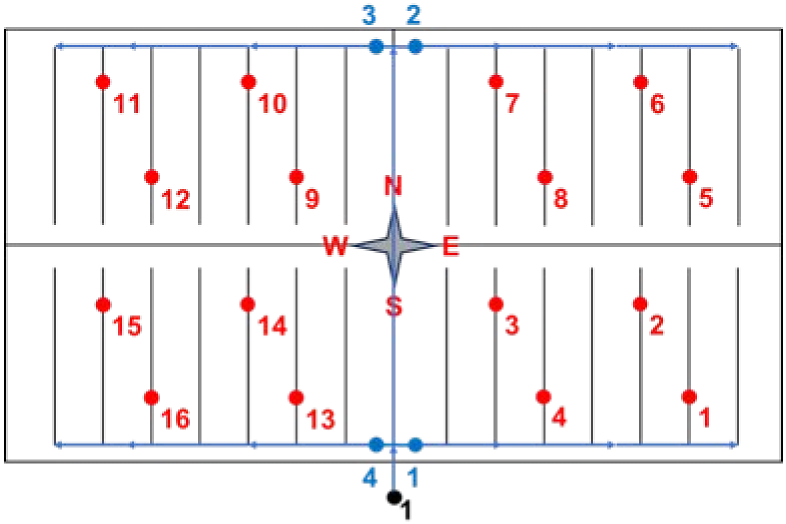


Fig. 1.13. Sketch of the irrigation system.

The irrigation strategy is governed by soil relative humidity data and it is fully automatic. When humidity readings drop below a user-defined lower threshold, the system activates the relevant solenoid valve until the humidity surpasses the upper threshold, at which point the valve is closed. This control logic can operate in either parallel or sequential mode (Fig. 1.14).
A manual timed irrigation mode is also available, allowing the operator to configure and initiate sector-specific irrigation sequences with customizable durations.

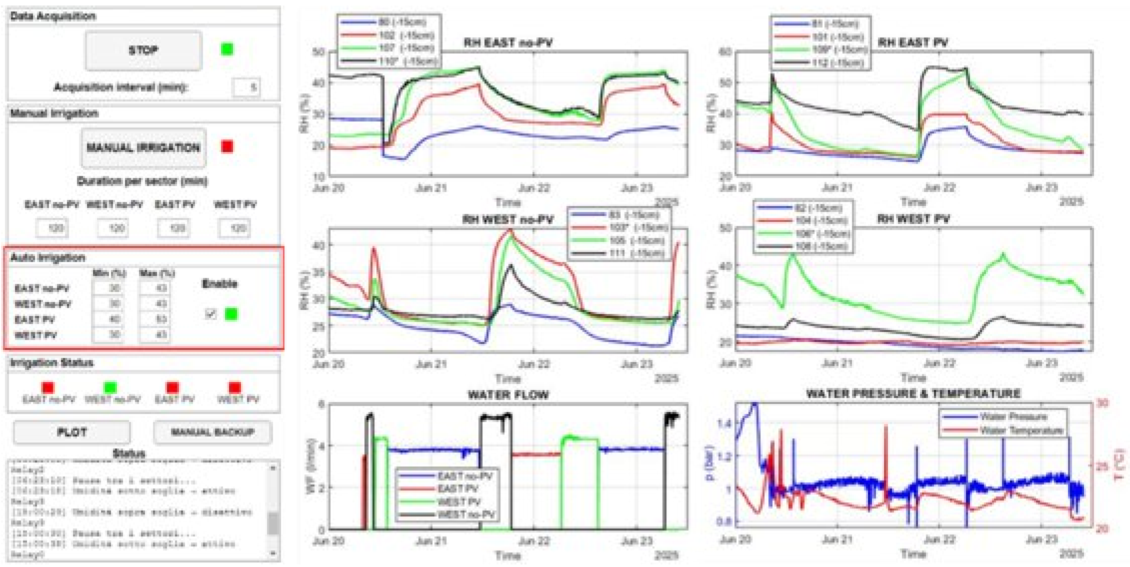


Fig. 1.14. Irrigation GUI (on the left) and real time monitoring of the irrigation (right).

1.2.4 Photovoltaic system data and sensors

Inside the greenhouse 64 custom bifacial PV panels are installed 75W rated, 32 in the east side section and 32 in the west one. In each section the panels are series connected and the two series are parallel connected at the inverter.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

The photovoltaic panel tracking system is single-axis (east–west), optimized both for electrical energy production and plant growth, promoting the transmission of solar radiation during the early hours of the day.
A suite of sensors has been installed symmetrically on both the east and west sides of the greenhouse, to monitor the performance of the photovoltaic system (Fig. 1.15). The suite includes the following components:

- two pyranometers mounted in-plane with PV modules to measure the Global Tilted Irradiance (GTI) and the Reflected Tilted Irradiance (RTI);
- two back-of-module (T_{BOM}) temperature sensors fixed directly on the rear surface of the PV panels to monitor thermal conditions affecting module efficiency and degradation;
- two illuminance sensors (luxmeters) positioned beneath the PV panel plane, one located in an area persistently free from panel-induced shading and the other intentionally placed in a region subject to shading. These allow differential measurement of light availability for the understory crop canopy.
- one combined temperature and relative humidity sensor and one CO_2 sensor, both installed beneath the panel structure to assess sub-panel microclimatic conditions;
- one inclinometer.

While not directly involved in electrical performance monitoring, the luxmeters and CO_2 sensor play a critical role in optimizing panel orientation. Based on real-time feedback from these sensors, the control algorithm can trigger the reorientation of the panels to a near-vertical position in scenarios where light availability or CO_2 concentration beneath the panels falls below crop-sustaining thresholds. This ensures dynamic optimization of both energy harvesting and agronomic performance in the agrivoltaic system

Fig. 1.15. Photovoltaic system sensors.

These data are integrated with the production data acquired by the SMA Inverter (Fig. 1.9), that collect the power production minutely or every 15 minutes according to the user. Ther inverter moreover provides information about the monthly and annual yield of the power plant, but no information about the current and the voltage of the string. The PV system performance inside the greenhouse is evaluated using a number of widely-used indicators, namely: (1) The instantaneous AC power is acquired from the SMA inverter, (2) The final yield (Yf), which is computed as the delivered electrical energy normalized by the installed nominal power, and (3) The performance ratio, which is the ratio between the final yield and the reference yield. The Reference Yield (Yr) is the received irradiance energy normalized by the bifacial standard test condition.

1.2.5 Crop and vegetation data

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

Two main experimental cultivation campaigns were conducted at the Tor Vergata pilot site (FSC) within the framework of the Project. Cycle I took place during the winter–spring season of 2025 and focused on the cultivation of lettuce and zucchini. Cycle II was subsequently carried out during the spring–summer season, during which tomato and eggplant were grown. For each crop, cultivation was implemented in two comparable experimental plots (PV and non-PV; see Fig. 1.8) to enable a direct comparison and to evaluate the effects of the photovoltaic installation on crop performance.
During Cycle I, the monitoring infrastructure was not yet fully procured, installed, or operational. In contrast, during Cycle II, all monitoring systems were fully deployed; therefore, the discussion below primarily refers to Cycle II.
Crop monitoring was primarily based on two key indicators. The first was fresh yield, which represents the target output of the cultivation process. The second was the leaf area index (LAI), a fundamental parameter for assessing plant growth and canopy development.
Fresh yield was measured for each row within each plot using a standard weighing scale and was recorded whenever fruits reached harvest maturity throughout the entire production cycle. In contrast, LAI is inherently more complex to quantify; therefore, a dedicated instrument (AccuPAR LP-80 ceptometer) was employed (Fig. 1.16), with multiple measurements repeatedly carried out over the course of the cultivation cycle.

Fig. 1.16. AccuPAR LP-80 ceptometer.

The instrument estimates the leaf area index (LAI) by measuring photosynthetically active radiation (PAR) above and below canopy. Based on the inter-row and intra-row spacing, the instrument was used to sample LAI at two representative locations within each of the four plots, with measurements conducted up to twice per week using the most appropriate sampling strategy (Fig. 1.17).

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

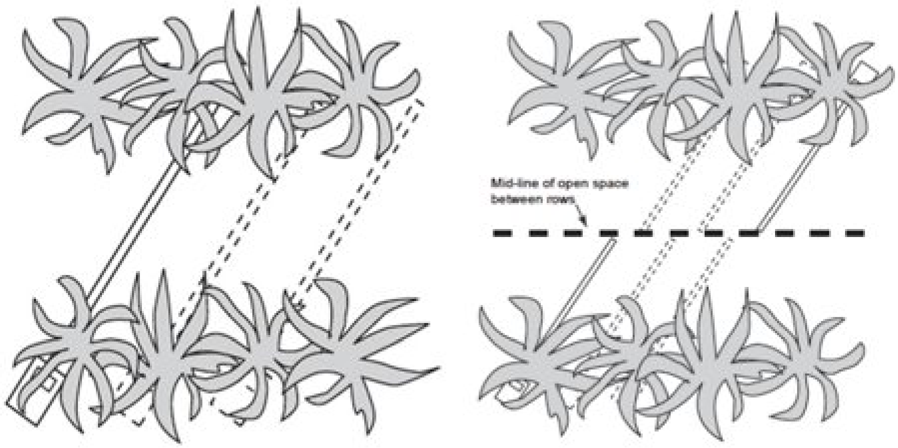


Fig. 1.17. LAI sampling in row crops.

By integrating the microclimatic and irrigation data described in Sections 1.2 and 1.3 with the data discussed here, it was possible to estimate the input variables required by crop growth models, drawing on relationships well established in the literature.

1.2.6.1 Data Pre-processing and Quality Control

1.2.6.1.1 Greenhouse Environmental data

The raw data stored in CSV format by the data acquisition system underwent a thorough preprocessing procedure. The first step consisted of verifying that all observations fell within the operational ranges of the sensors. Subsequently, the dataset was screened for erroneous measurements, identified as values lying well outside the expected ranges of the monitored physical variables. When potentially anomalous data points were detected, the adopted procedure involved visual inspection of a representative portion of the corresponding time series in order to determine whether the observation should be removed and replaced with a missing value flag. Once the dataset had been cleaned, a resampling step was applied to harmonize the temporal resolution of the data to five-minute intervals.
Finally, for the purposes of crop growth analysis and specifically for Cycle II, the preprocessed dataset was first resampled to an hourly time step. Hourly values were computed as the means of the available observations provided that at least 75% of the expected five-minute measurements were valid; otherwise, the corresponding hourly value was labelled as missing. The resulting hourly data set was then subjected to an imputation procedure in which missing values were replaced, on an hour-by-hour basis, with the mean of the values recorded at the same hour during the three preceding and three following days. Imputation was performed only when at least four of the six neighboring values used for reconstruction were available.
Since the FAO AquaCrop model operates with a minimum temporal resolution of one day, an additional resampling step to a daily time scale was carried out in order to derive the complete set of input variables required for the modelling activities.

1.2.6.1.2 Irrigation Data

Similarly to the climatic data described in the previous section, the raw irrigation-system data acquired by the data acquisition platform were subjected to a dedicated preprocessing workflow. The flow-rate time series was resampled to a uniform temporal resolution of five minutes. Data quality checks were then performed to ensure consistency with the nominal discharge rates of the irrigation system; subsequently, missing observations were imputed by assigning zero-flow values.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

1.2.6.1.3 Photovoltaic System Data

The photovoltaic dataset was subjected to a dedicated preprocessing and quality-control workflow prior to the performance assessment. First, a night/low-signal filter was applied to flag measurements collected under negligible irradiance conditions. In particular, the data points were filtered when global tilted irradiance was below 20 W m^{-2}. An additional statistical filter based on the temperature-corrected performance ratio was applied to remove inconsistent operating points (outliers); observations outside the accepted PR range were flagged and eliminated. This ensured that subsequent PR aggregation and modelling steps relied on physically plausible PV operating conditions while maintaining a continuous timestamp grid. The accepted PR range was found for each case separately by plotting frequency of the temperature-corrected performance ratio.

Conclusions

This chapter presents a detailed overview of the experimental sites, monitoring infrastructure, and data management framework implemented at the FSC and UTH pilot greenhouse. The adopted approach integrates environmental, irrigation, photovoltaic, and crop-related datasets into a coherent and high-resolution monitoring system, specifically designed to capture the complex interactions characterizing agrivoltaic greenhouse environments.

The multi-level sensor deployment enables a fine-scale characterization of greenhouse microclimatic conditions, both spatially and vertically, while the automated irrigation system provides precise control and monitoring of water use based on soil moisture feedback. In parallel, the photovoltaic monitoring setup allows for continuous assessment of energy production and panel operating conditions, as well as indirect evaluation of the effects of PV shading on the crop microclimate. Crop and vegetation data complement these measurements by providing key indicators of plant growth and productivity, enabling a direct link between environmental drivers and agronomic outcomes.

A robust data preprocessing and quality control workflow was applied to all datasets to address issues related to sensor errors, missing values, and temporal misalignment. The harmonization of data at multiple temporal scales ensures compatibility with crop growth models and performance analyses, while preserving the information content of the original high-frequency measurements.

In summary, the data infrastructure described in this chapter constitutes a solid and reliable foundation for the subsequent analytical and modelling activities. By ensuring data consistency, traceability, and integration across domains, it supports a comprehensive evaluation of agrivoltaic greenhouse performance in terms of energy production, water use efficiency, and crop productivity.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Chapter 2. Dynamic modelling of the greenhouse microclimate

Introduction

This chapter presents the development and validation of a numerical simulation model aimed at reproducing the microclimatic conditions of the photovoltaic greenhouse located at the *Fattoria Solidale del Circeo* (FSC) in Pontinia, Latina, Italy. The study focuses on the analysis of air temperature and relative humidity dynamics, which represent key variables for both crop growth and greenhouse operational performance.

The main objective is to develop a simulation framework capable of accurately representing the internal thermal and hygrometric environment under real operating conditions, thereby providing a reliable basis for performance analysis and future optimization studies.

The work builds upon previous simulation activities carried out for a similar greenhouse located in Volos, Greece, extending and adapting the modeling approach to a different climatic context, structural configuration, and crop typology (Cornaro et al. 2025).

While the general modeling strategy remains consistent with the reference case, significant refinements were required to account for site-specific characteristics, local climatic conditions, and different cultivation practices.

Attention was therefore devoted to improving the representation of the greenhouse microclimate through detailed data acquisition, model calibration, and validation against experimental measurements.

To ensure a high degree of correspondence between simulated and measured conditions, the modeling activity was structured into four main phases, each addressing a specific aspect of the simulation workflow:

- The first phase consisted of an on-site survey carried out at the Fattoria Solidale del Circeo, aimed at collecting detailed information on the greenhouse geometry, construction materials, photovoltaic system layout, and operational characteristics. This phase was essential for defining the boundary conditions and physical parameters required to accurately describe the real system within the simulation environment.
- The second phase involved the development of a detailed geometric and thermal model of the greenhouse using the simulation software IDA ICE 5.1. Care was taken to ensure a faithful representation of the real structure, including the spatial subdivision of the greenhouse into multiple zones to capture horizontal and vertical microclimatic gradients. Thermal properties of the envelope, internal heat gains, and ventilation strategies were implemented in accordance with the collected technical data.
- The third phase focused on the generation and validation of a site-specific climatic file to be used as input for the simulation model. This file was developed using data collected from an experimental weather station installed at the Circeo site and subsequently validated through comparison with measurements from the regional ARSIAL–Cotarda meteorological station.
  This step was crucial to ensure the reliability of external boundary conditions, particularly with respect to solar radiation, air temperature, and humidity, which strongly influence greenhouse microclimate dynamics.
- The fourth phase addressed the refinement of the crop–microclimate interaction modeling, with particular emphasis on the modification and calibration of the evapotranspiration process.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Given the significant role of vegetation in regulating heat and moisture exchanges within the greenhouse, this phase included the adjustment of key plant-related parameters such as Leaf Area Index (LAI)**,** canopy conductance**,** and aerodynamic conductance**.**
These refinements were introduced to improve the representation of crop-driven latent heat fluxes and their impact on internal temperature and relative humidity levels.
The simulation period considered in this study spans from 21 May 2025 to 25 August 2025**,** corresponding to the complete growing cycle of tomato and eggplant crops cultivated inside the greenhouse. This time window was selected to evaluate the model performance under typical summer conditions, characterized by high solar radiation, elevated external temperatures, and significant thermal and hygrometric variability.
The model results were subsequently validated through a systematic comparison with measured data, allowing for a quantitative assessment of the accuracy and reliability of the proposed simulation approach.

2.1 IDA-ICE modelling of the greenhouse (FSC)

2.1.1 Model construction

Based on the available planimetric and geometric data, the greenhouse was modeled using the CAD tools integrated within IDA ICE 5.1. Care was taken to ensure that the simulated geometry faithfully reproduced the real structure in terms of dimensions, shape, and orientation.
This level of geometric accuracy was considered essential to correctly capture solar radiation distribution, internal air stratification, and the overall microclimatic behavior of the greenhouse.
The greenhouse model was divided into two main areas, referred to as Main Zone 1 and Main Zone 2, corresponding respectively to the eastern and western sectors of the structure.
This subdivision was introduced to account for the asymmetric exposure to solar radiation and for potential differences in microclimatic conditions between the two sides of the greenhouse.
Each of these main zones was further subdivided into multiple vertical subzones to reproduce the vertical stratification of temperature and relative humidity observed in the experimental measurements. Specifically, each sector was divided into four subzones at different height levels, consistently with the four measurement levels of the internal greenhouse microclimate.
These measurements were obtained through the installation of ten monitoring poles, which form part of the experimental monitoring system described in the previous chapter.
The vertical zoning strategy was therefore designed to directly match the spatial resolution of the available experimental data, enabling a meaningful comparison between simulated and measured variables during the validation phase.
Focusing on Main Zone 1, corresponding to the eastern sector of the greenhouse, the following subzones were defined:

- Subzone 1.1 has a surface area of 90 m^2 and extends from ground level up to a height of 1 m.
- Subzone 1.2, with the same surface area, spans the vertical range between 1 m and 2 m above ground level.
- Subzone 1.3 also covers 90 m^2 and extends from 2 m to 3.2 m, corresponding to the transition zone between the vertical perimeter walls and the circular vaulted roof of the greenhouse.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

- Finally, the Roof subzone represents the upper volume enclosed by the circular vault, with a surface area of 90 m² and a maximum height of approximately 5.2 m.

An identical vertical and horizontal subdivision was applied to Main Zone 2, corresponding to the western sector of the greenhouse, ensuring consistency in the spatial discretization of the model and allowing a direct comparison between the two sectors.

Fig. 2.1 View of the greenhouse located at FSC. The shading structures located to the east, south, and west are highlighted

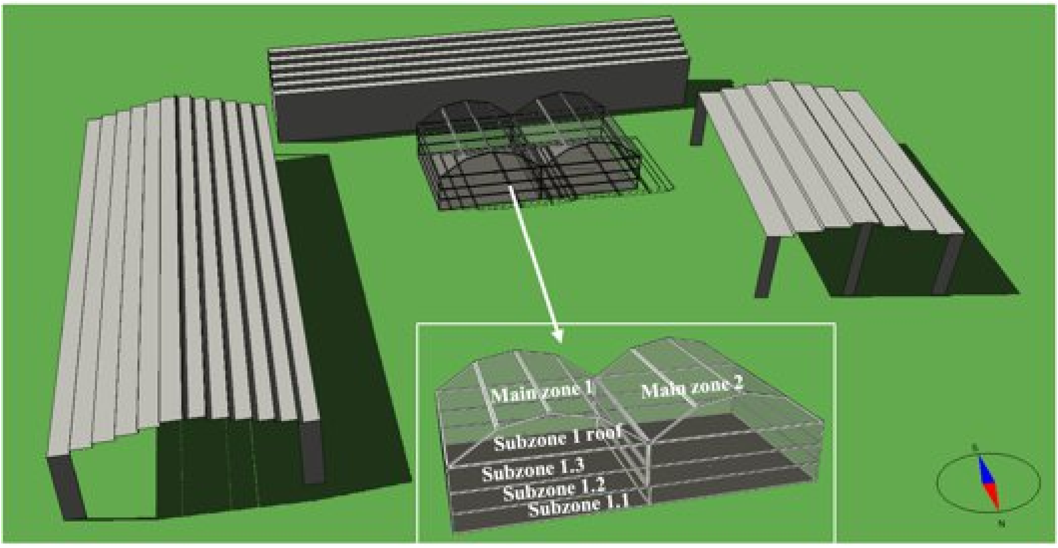


Fig. 2.2 GH 3D model done with IDA ICE 5.1 showing the subdivision in the two main zones and into their respective subzones.

To ensure accurate simulation results, the surrounding buildings located to the south, east, and west of the greenhouse were also included in the simulation model.
These structures play a significant role in shading the greenhouse, particularly during the winter season and during low solar altitude conditions and therefore have a non-negligible impact on the incoming solar radiation and the resulting internal microclimate.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Their inclusion was essential to realistically reproduce the boundary conditions affecting the greenhouse throughout the simulation period.

2.1.2 Climate file

Once the CAD model of the greenhouse was completed, it became necessary to generate a climate file corresponding to the locality of Pontinia, where the *Fattoria Solidale del Circeo* is located, in order to upload it into simulation software.
Meteorological data were obtained from the experimental weather station installed at the *Fattoria Solidale del Circeo*.
This data, compiled in an Excel file, corresponds to measurements of external variables recorded outside the greenhouse, including:

- Temperature (°C)
- Relative humidity (%)
- Global Horizontal Irradiance (W/m²)
- Wind speed in the prevailing direction (m/s)
- Prevailing wind direction (°)

The dataset was subsequently imported into the IDA ICE software through the creation of a climate file in the ".PRN" format.
The .PRN file requires a specific hourly sequence of variables for an entire year, amounting to 8,784 hours.
Since the ".PRN" format requires the decomposition of total global radiation into direct normal and diffuse components, it was necessary to develop a MATLAB script to perform this calculation.
Within this script, the tool *pvl_erbs*, available in the MATLAB photovoltaic library, was employed. The *pvl_erbs* model takes as inputs the incident global horizontal irradiance (denoted as GHI), the true solar zenith angle in decimal degrees (denoted as Z), and the day of the year (denoted as *doy*).
The model outputs the desired radiation components: Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI), both expressed in W/m², as well as the clearness index (K_t), defined as the ratio between global radiation and solar radiation on a horizontal plane.
Among the necessary inputs, only the global radiation was directly measured. Therefore, additional MATLAB code was implemented to calculate the true solar zenith angle and the day of the year. Moreover, because the meteorological station records data at five-minute intervals, a further MATLAB routine was developed to average the measurements, obtaining the hourly resolution required by the .PRN file.
Since the .PRN file used by IDA ICE must cover an entire solar year, and the greenhouse experiment began during the current year, an initial sample climate file was generated using a Test Reference Year (TRY) corresponding to the nearest location to Pontinia, namely Rome Ciampino Airport.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

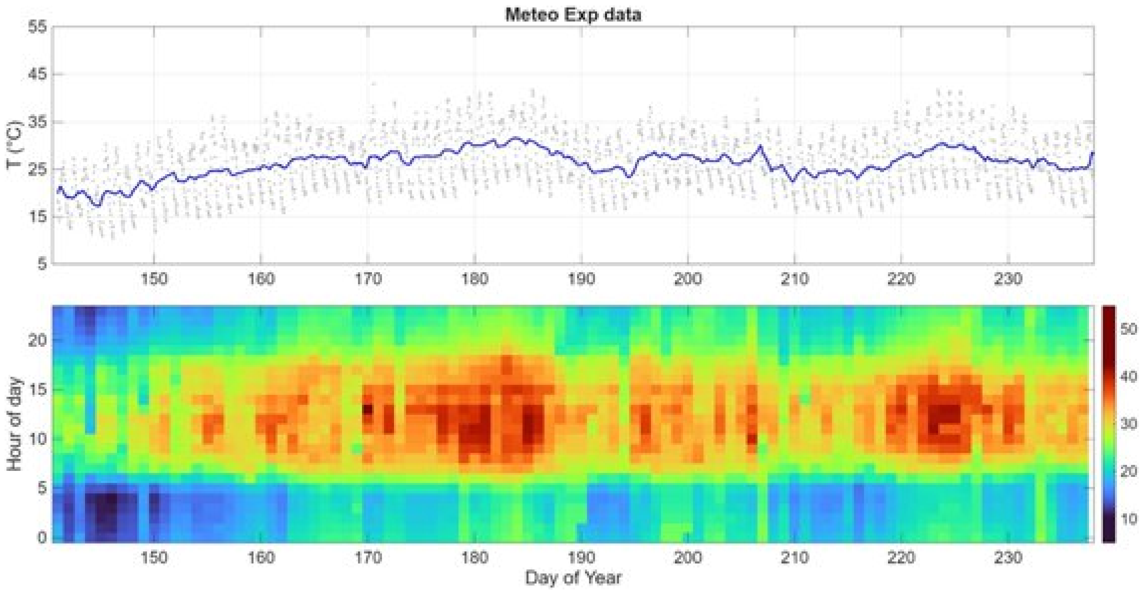


Fig. 2.3 Time trend and carpet plot of the T measured by the experimental weather station located at FSC. The x-axis shows the days of the year corresponding to the simulation period, which extends from 21 May 2025 to 25 August 2025

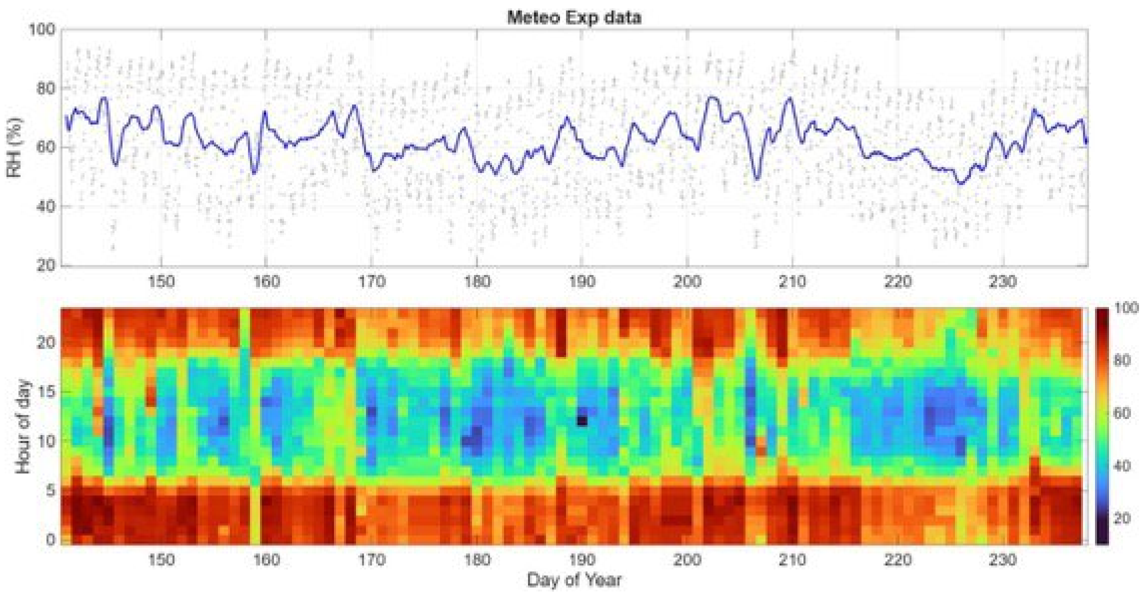


Fig. 2.4 Time trend and carpet plot of the RH measured by the experimental weather station located at the FSC. The x-axis shows the days of the year corresponding to the simulation period, which extends from 21 May 2025 to 25 August 2025

This file served as a placeholder, and its data were progressively replaced with the measurements obtained from the weather station until December 31, 2025.
The sequence of variables required in the .PRN file, derived from the converted Excel dataset, is as follows:

- Time (h), starting from 0 for 01.01.2025 at 00:00 up to 8,784 for 31.12.2025 at 24:00;
- Outdoor dry-bulb air temperature (°C);
- Relative Humidity (%);
- Prevailing wind direction (°);

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

- Measured wind speed in the prevailing direction (m/s);
- Direct normal component of global radiation (W/m²);
- Diffuse horizontal component of global radiation (W/m²).

The carpet plots and the time trends related to the external values of Temperature (T), Relative Humidity (RH) associated with the simulation period are reported in Figure 2.3 and 2.4.

2.1.3 Ventilation and infiltrations

In the IDA ICE model, infiltration and ventilation were defined to represent the air exchange conditions of the greenhouse.
A fixed infiltration rate of 1 Air Changes per Hour (ACH) was assigned at the building level and distributed to the zones proportionally to their volume.
At the same time, the leakage area in all zones was set to zero, assuming no uncontrolled air leakage through the envelope.
Natural ventilation is provided exclusively through operable windows: the two longitudinal perimeter façades are equipped with windows 1 m high and 18 m long, opened at 8%, while the two transverse façades include windows 1 m high and 10 m long, also opened at 8%.
As a result, air exchange within the greenhouse is governed by the combination of the imposed air change rate and the controlled opening of large façade windows, rather than by envelope airtightness defects.

2.1.4 Evapotranspiration and use of crop data to vary the evapotranspiration with look up tables

In order to take into account, the plants evapotranspiration in the greenhouses a tool within IDA-ICE 5.1 was developed. For doing that, a custom zone within DBS software that considers the crop evapotranspiration process was designed. This is modelled as proposed in (Katsoulas and Stanghellini 2019).
In particular, the Penman-Monteith equation is used to compute the evapotranspiration mass flow, $\dot{m}_{ET_0}$(in kg s-1) (Monteith 1965):

$$\dot{m}_{ET_0} = A_c \frac{\Delta R_n + \rho C_p D_i g_a}{\lambda \left[\Delta + \gamma \left(1 + \frac{g_a}{g_c} \right) \right]} \tag{2.1}$$

where R_n is the net radiation intercepted by the crop (Wm^{-2}), D_i is the vapor deficit of the air (kPa), g_a and g_c are the crop aerodynamic and canopy conductance (ms^{-1}), Δ is the slope of the saturation vapor pressure-temperature relationship ($kPa\ K^{-1}$), γ is the psychrometric constant ($kPa\ K^{-1}$), A_c is the crop area (m^2), and ρ, C_p,λ are the air density (kgm^{-3}), specific heat capacity ($J\ kg^{-1}K^{-1}$) and latent heat of vaporization ($J\ Kg^{-1}$).
On the other hand, Δ is computed as:

$$\Delta = \frac{4098 P_{v,sat}}{(T + 237.3)^2} \tag{2.2}$$

with T and $P_{v,\,sat}$ being the air temperature (°C) and the saturation vapor pressure (kPa).
Following (Katsoulas and Stanghellini 2019), the intercepted net radiation is modelled as:

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

$$R_n = a\left(1 - e^{-k_s LAI}\right)R_s \tag{2.3}$$

where R_s is the solar radiation per unit area (Wm^{-2}), a is an empirical constant (indicating the fraction of net radiation absorbed by the crops), k_s is the extinction coefficient for shortwave radiation, and LAI is the leaf area index.
In the zonal model, the coefficients a and k_s are set to 0.70 and 0.55. In addition, the solar radiation R_s is calculated as the shortwave radiation reaching the floor divided by its area (since the crops are assumed to be at the ground level). Finally, the latent power related to the evapotranspiration mass flow is calculated as in (Baglivo et al. 2020):

$$P_{ET} = \dot{m}_{ET_0}\left(l + C_{p,v}T\right) \tag{2.4}$$

($C_{p,v}$ is the specific heat capacity of the vapor, $J\,kg^{-1}K^{-1}$).
In this study the simulation was carried out for the most critical summer season with a variable simulation time step with maximal time step of 0.5 hours and an output time step of 1 hour.

2.1.4.1 Aerodynamic conductance (g_a) and canopy conductance (g_c)

Aerodynamic conductance (g_a) and canopy conductance (g_c) play a fundamental role in governing heat and mass exchange processes between vegetation and the surrounding air, directly influencing both sensible and latent heat fluxes within greenhouse environments.
Aerodynamic conductance describes the efficiency of turbulent transport of heat and water vapor between the crop canopy and the bulk air and is primarily affected by canopy structure, leaf roughness, and airflow conditions.
Canopy conductance, on the other hand, represents the physiological control exerted by plants on transpiration through the opening and closing of stomata and is strongly influenced by crop development and environmental conditions.
In the previous simulation model developed for the greenhouse located in Volos, Greece, both aerodynamic and canopy conductance were assumed to be constant in time.
While this simplified parameterization allowed for a first-order representation of crop–microclimate interactions, it did not account for the significant changes in canopy structure and plant physiological activity occurring throughout the cultivation cycle.
Based on insights derived from scientific literature, including experimental studies on greenhouse crops, it became evident that both aerodynamic and canopy conductance exhibit pronounced temporal variability as a function of crop growth stage, Leaf Area Index (LAI), and canopy architecture (Haofang et al. 2023).
For this reason, the assumption of constant conductance values was considered inadequate for accurately simulating the microclimate of the Circeo's greenhouse.
Compared to the Volos model, a methodological improvement was therefore introduced in the present work by implementing a time-dependent representation of aerodynamic and canopy conductance. This was achieved using source tables within the IDA ICE simulation environment, allowing these parameters to vary over time in accordance with crop growth and development.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

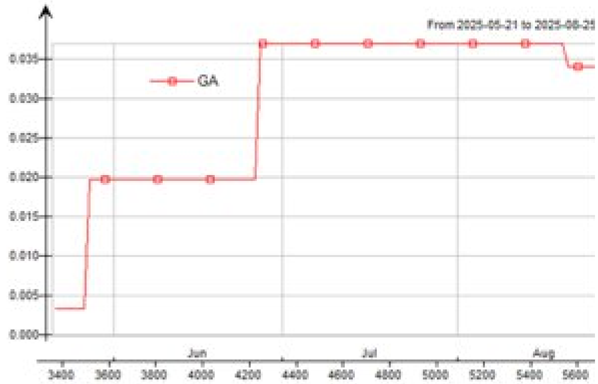


Fig. 2.5 Time trend of aerodynamic conductance. The x-axis represents the annual hours corresponding to the simulation period from 21 May 2025 to 25 August 2025

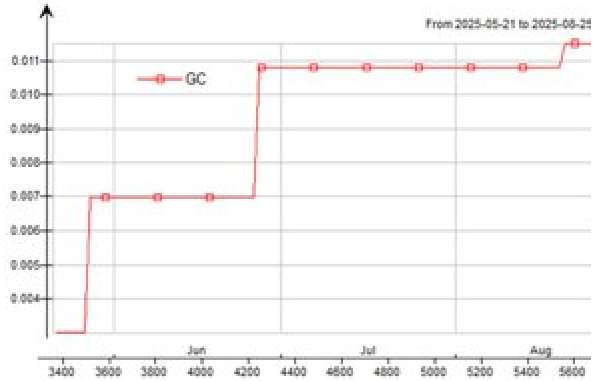


Fig. 2.6 Time trend of canopy conductance. The x-axis represents the annual hours corresponding to the simulation period from 21 May 2025 to 25 August 2025

The temporal profiles adopted in the source tables were derived from literature-based relationships and adapted to the specific crops cultivated in the greenhouse.

By allowing aerodynamic and canopy conductance to evolve over the growing season, the model is able to more realistically reproduce crop transpiration dynamics and their impact on internal air temperature and relative humidity.

This enhancement represents a significant improvement with respect to the previous Volos simulation model and contributes to the improved agreement observed between simulated and measured microclimatic variables in the validation phase.

2.1.4.2 Leaf Area Index (LAI)

In order to improve the representation of crop-driven latent heat fluxes within the greenhouse, a dedicated modeling approach was adopted to explicitly account for the temporal evolution of the Leaf Area Index (LAI). For each cultivated crop, specific source tables were created within the IDA ICE simulation environment and populated with experimentally measured LAI values, allowing the model to directly reflect the actual development of the crop canopy over time.

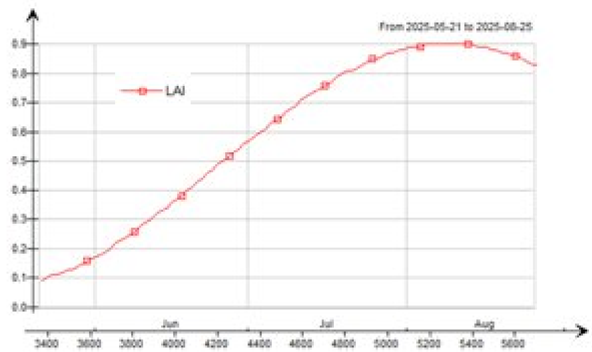


Fig. 2.7 Time trend of the LAI associated to tomato crops. The x-axis represents the annual hours corresponding to the simulation period from 21 May 2025 to 25 August 2025

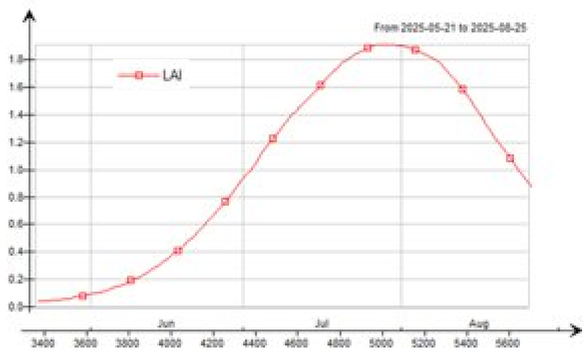


Fig. 2.8 Time trend of the LAI associated to the eggplant crop. The x-axis represents the annual hours corresponding to the simulation period from 21 May 2025 to 25 August 2025

In the present study, this approach was applied to the tomato and eggplant crops cultivated inside the greenhouse (Xuewen G.et al 2021).

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

The LAI values were measured on-site using the ACCUPAR LP-80 instrument and associated with the corresponding stages of crop growth.
Separate source tables were defined for tomato and eggplant, ensuring that crop-specific differences in canopy development were accurately represented.
These time-dependent LAI profiles were then integrated into the evapotranspiration calculation framework of the model. The importance of considering a variable LAI during crop growth is well established in the literature which identifies LAI as a key parameter controlling crop transpiration and evapotranspiration dynamics (Richard G. ALLEN et al. 1998).
According to it, the increase in LAI over the growing season leads to a progressive enhancement of the transpiring surface and, consequently, to higher latent heat fluxes. Neglecting this temporal evolution may therefore result in a systematic underestimation of evapotranspiration, particularly during the mature stages of crop development.
In the previous simulation model developed for the greenhouse in Volos, Greece, the LAI was assumed to be constant throughout the simulation period.
This simplification led to an underestimation of the evapotranspiration process and of its contribution to the greenhouse microclimatic balance, especially under conditions of high canopy density.
In the present model of the *Fattoria Solidale del Circeo* greenhouse, evapotranspiration is instead calculated as a function of the time-varying LAI, in combination with the dynamic canopy and aerodynamic conductance described in the previous sections.
This integrated formulation allows latent heat fluxes to evolve consistently with crop growth, resulting in a more physically realistic representation of plant–air interactions.

2.2 Validation results

This section presents the validation of the greenhouse simulation model developed using IDA ICE, based on a comparison between measured and simulated air temperature and relative humidity data.
The analysis was conducted relating to the main two greenhouse zones (Main Zone 1 and Main Zone 2) for the following subzones:

- subzones 1.2 and 1.3 associated to the eastern sector;
- subzones 2.2 and 2.3 associated to the western sector.

Model validation was performed using:

- Time-series plots, to evaluate the capability of the model to reproduce the temporal evolution of the monitored variables.
- Scatter plots, to assess the correlation between simulated and measured values, quantified through the coefficient of determination (R^2).

The combined analysis of trend comparison and statistical correlation provides a comprehensive evaluation of the model's reliability in reproducing the greenhouse microclimatic conditions.

2.2.1 Temperature trend validation

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

The validation of indoor air temperature was carried out by comparing simulated and measured temperature data for subzones 1.2, 1.3, 2.2, and 2.3 through both time-series analysis and scatter plot evaluation.
The time-series comparison demonstrates that the simulation model is able to accurately reproduce the temporal evolution of air temperature in all analyzed subzones.
The simulated temperature profiles closely follow the experimental measurements, successfully capturing daily thermal cycles as well as longer-term variations observed throughout the monitoring period. In particular, the model correctly represents the timing and magnitude of both temperature peaks and minimum in subzones 1.2, 1.3, 2.2, and 2.3, indicating an adequate representation of the dominant thermal processes occurring inside the greenhouse.
The scatter plot analysis further confirms the strong agreement between simulated and measured temperatures across all subzones. A clear linear correlation is observed in each case, with coefficients of determination (R^2) ranging between approximately 0.88 and 0.89.
These high values indicate that the simulation model is capable of explaining most of the variability observed in the experimental temperature data.
Furthermore, the regression lines exhibit slopes close to unity, suggesting the absence of significant systematic bias in temperature prediction.

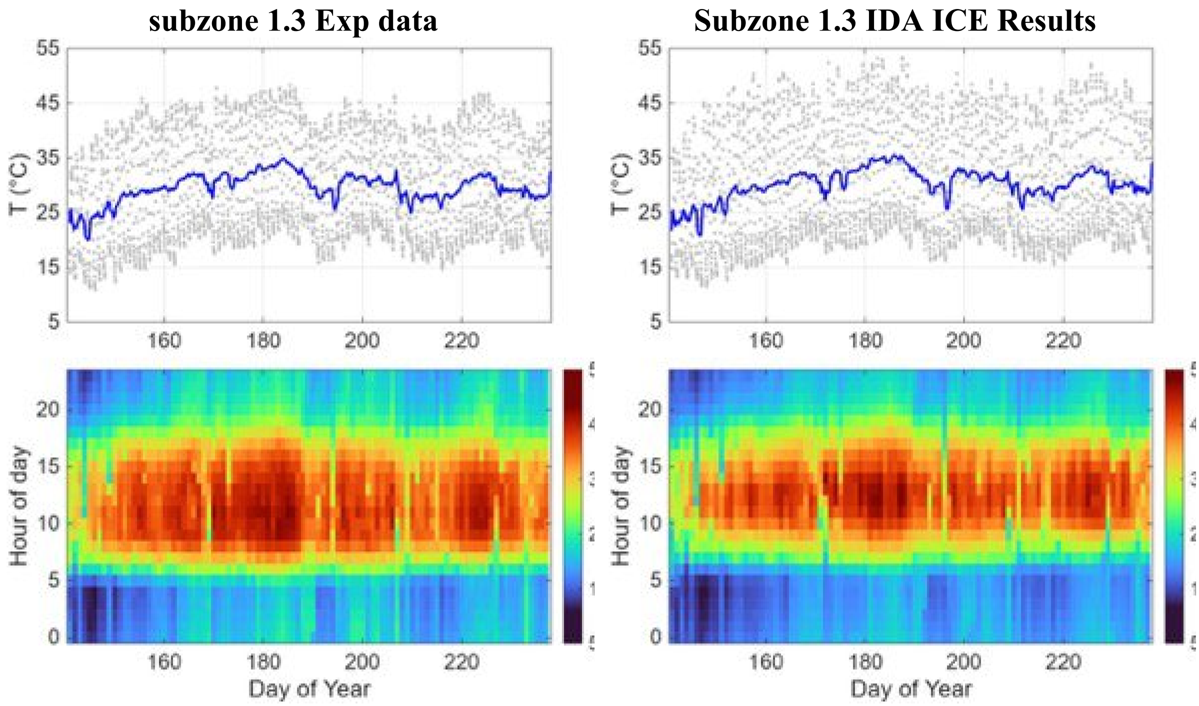


Fig. 2.9 Time trend and carpet plot of the temperature measured and simulated associated to subzone 1.3 of the greenhouse located at the FSC. The x-axis shows the days of the year corresponding to the simulation period, which extends from 21 May 2025 to 25 August 2025

Minor discrepancies between simulated and measured values are mainly observed during periods characterized by rapid temperature variations, such as high solar radiation conditions or nighttime cooling phases. These deviations can be attributed to uncertainties in boundary conditions, simplifications in the modeling of solar gains and shading effects, as well as potential measurement inaccuracies. Nevertheless, the magnitude of these differences remains limited and does not compromise the overall reliability of the model.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Table 2.1 summarizes the temperature correlation parameters calculated for each analyzed subzone.

Table 2.1 temperature correlation parameters calculated for each analyzed subzone

Temperature (T) Correlation Parameters							
Main Zone	**Subzone**	**ρ**	**R^2**	**RMSE**	**MAE**	**MBE**	**NMBE [%]**
1	2	0.9413	0.8860	3.222	2.373	-1.030	-3.53
1	3	0.9437	0.8905	3.600	2.530	0.732	2.48
2	2	0.9368	0.8776	3.143	2.341	0.006	0.02
2	3	0.9362	0.8764	3.786	2.668	0.813	2.78

The Pearson correlation coefficient (ρ) confirms a very strong linear relationship between simulated and measured temperatures in all cases, with values consistently higher than 0.93. The coefficients of determination (R^2) further support the robustness of the model performance. Error-based indicators, including the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), indicate moderate deviations that remain within acceptable limits for greenhouse thermal simulations. Finally, the Mean Bias Error (MBE) and the Normalized Mean Bias Error (NMBE) reveal only minor tendencies toward underestimation or overestimation, with values generally close to zero, confirming the absence of significant systematic errors in the temperature predictions.

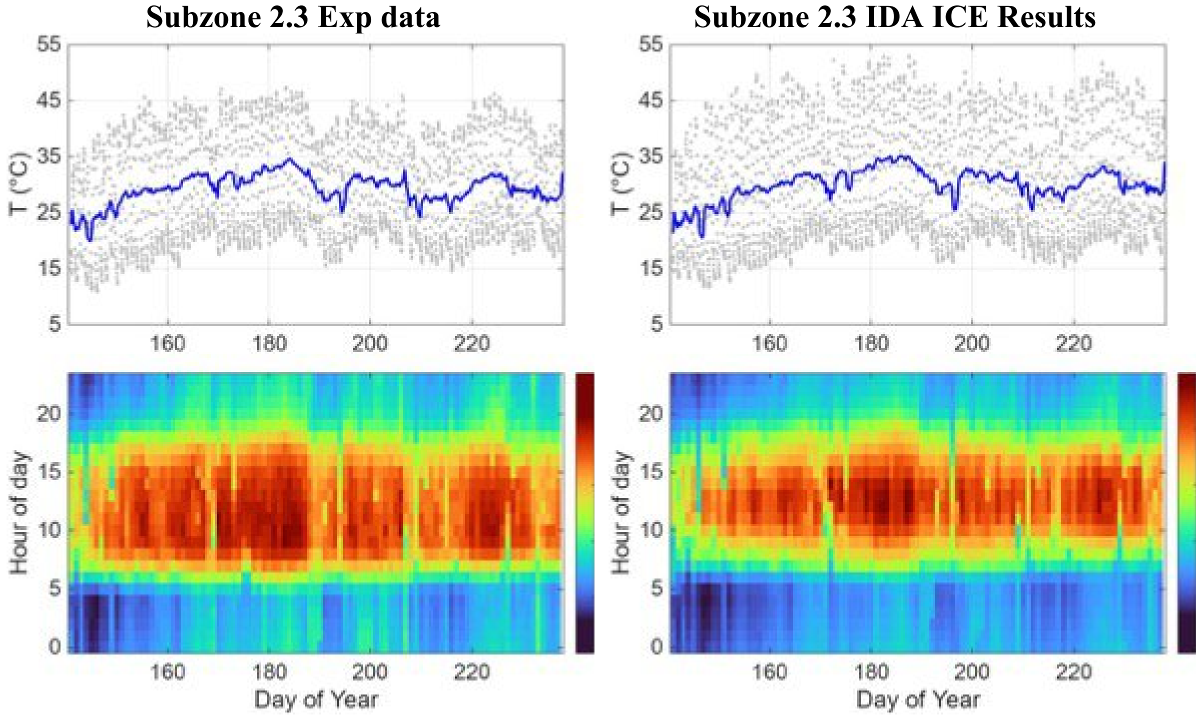


Fig. 2.10 Time trend and carpet plot of the temperature measured and simulated associated to the subzone 2.3 of the greenhouse located at the FSC. The x-axis shows the days of the year corresponding to the simulation period, which extends from 21 May 2025 to 25 August 2025

The carpet plots and time trends of both experimental and simulated temperatures, together with the associated relative humidity profiles during the simulation period, are presented in the figures 2.10 - 2.12.

The carpet plots and the time trends related to the experimental and simulated temperature of the greenhouse as well as the relative humidity associated to the simulation period are reported below.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

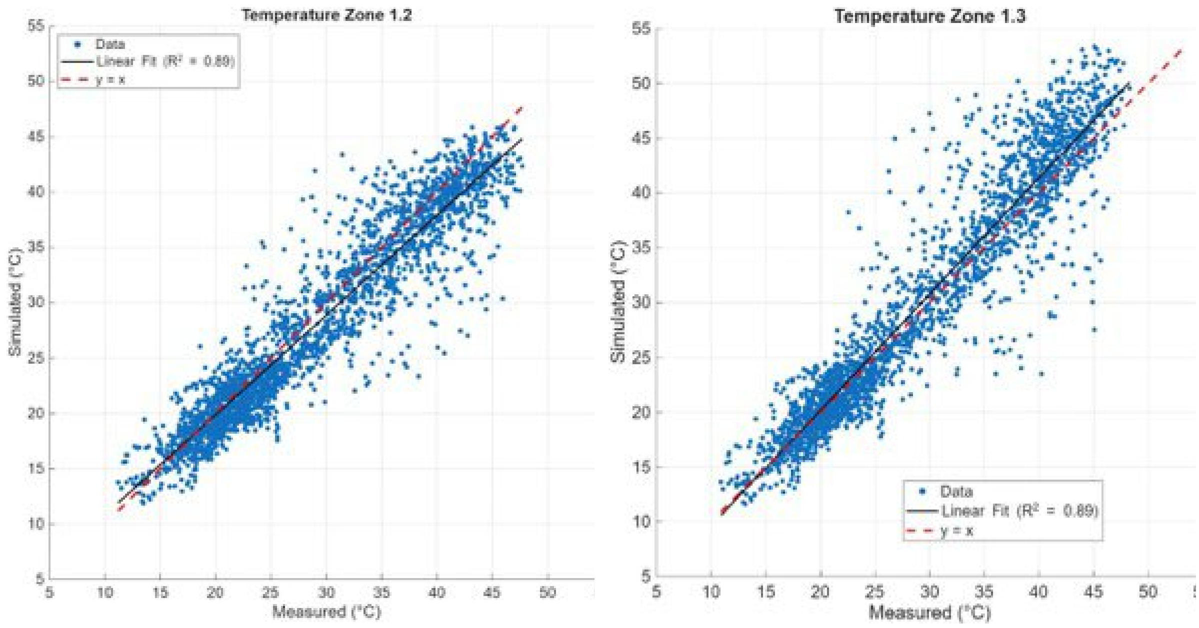


Fig. 2.11 Scatter plots of temperature associated to subzones 1.2 and 1.3 of the greenhouse located at FSC.

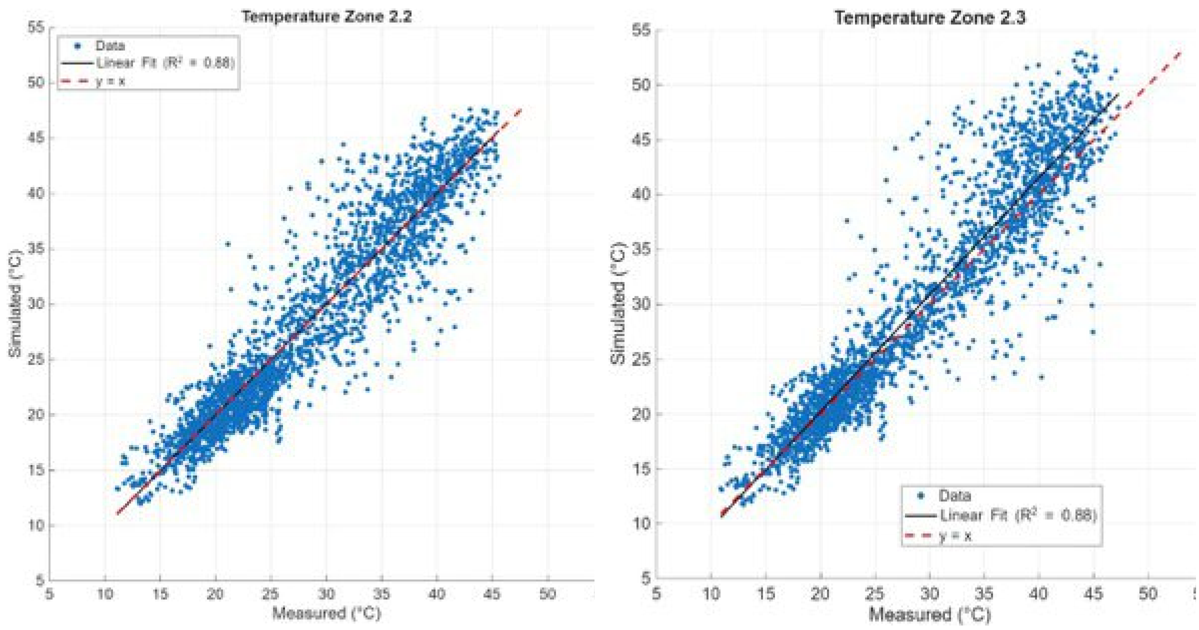


Fig. 2.12 Scatter plots of temperature associated to subzones 2.2 and 2.3 of the greenhouse located at FSC

2.2.2 Relative Humidity trend validation

The validation of relative humidity (RH) was conducted by comparing simulated and measured RH data for subzones 1.2, 1.3, 2.2, and 2.3 using both time-series analysis and scatter plot evaluation.

The comparison of the temporal trends indicates that the simulation model is able to reproduce the overall evolution of relative humidity within the greenhouse environment, capturing the main daily fluctuations as well as the broader seasonal patterns observed in the monitored data.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

The simulated RH profiles generally follow the experimental measurements, particularly with respect to the timing and magnitude of humidity variations during typical day–night cycles.
The model adequately represents the strong coupling between air temperature and relative humidity, successfully reproducing the decrease in RH during periods of increasing temperature and the corresponding increase during nighttime conditions.
This behavior suggests that the interactions among thermal conditions, ventilation processes, and internal moisture dynamics are reasonably captured by the simulation framework.
However, when compared to air temperature, the relative humidity results exhibit larger discrepancies between simulated and measured values, especially during periods characterized by elevated humidity levels or rapid environmental variations.
These differences are more pronounced during transitional phases and can be attributed to the intrinsic complexity of humidity-related phenomena, including plant evapotranspiration, moisture buffering effects of greenhouse materials, variability in air exchange rates, and spatial heterogeneity of internal moisture sources.
The scatter plot analysis reveals a good correlation between simulated and measured RH values across all analyzed subzones. The coefficients of determination (R^2), which range approximately between 0.79 and 0.84, indicate that a substantial portion of the variability observed in the experimental data is captured by the model. Although these values are lower than those obtained for air temperature validation, they are consistent with the higher level of uncertainty typically associated with humidity modeling in greenhouse environments.
The regression lines show a reasonable alignment with the ideal 1:1 relationship, suggesting the absence of strong systematic overestimation or underestimation over the investigated RH range.
Table 2.2 reports the statistical correlation parameters calculated for relative humidity in each analyzed subzone. The Pearson correlation coefficient (ρ) indicates a strong linear relationship between simulated and measured values, with coefficients generally close to or higher than 0.89. The coefficients of determination (R^2) further confirm the satisfactory predictive capability of the model.

Table 2.2 correlation parameters calculated for relative humidity in each analyzed subzone

Relative Humidity (RH) Correlation Parameters							
Main Zone	**Subzone**	**ρ**	**R^2**	**RMSE**	**MAE**	**MBE**	**NMBE [%]**
1	2	0.8973	0.8051	14.579	11.923	10.838	16.91
1	3	0.9161	0.8392	0.099	0.075	0.004	0.55
2	2	0.8862	0.7854	13.965	11.177	9.692	15.23
2	3	0.9174	0.8416	0.106	0.083	0.029	4.58

Zone 1.3 Exp data **Zone 1.3 IDA ICE Results**

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

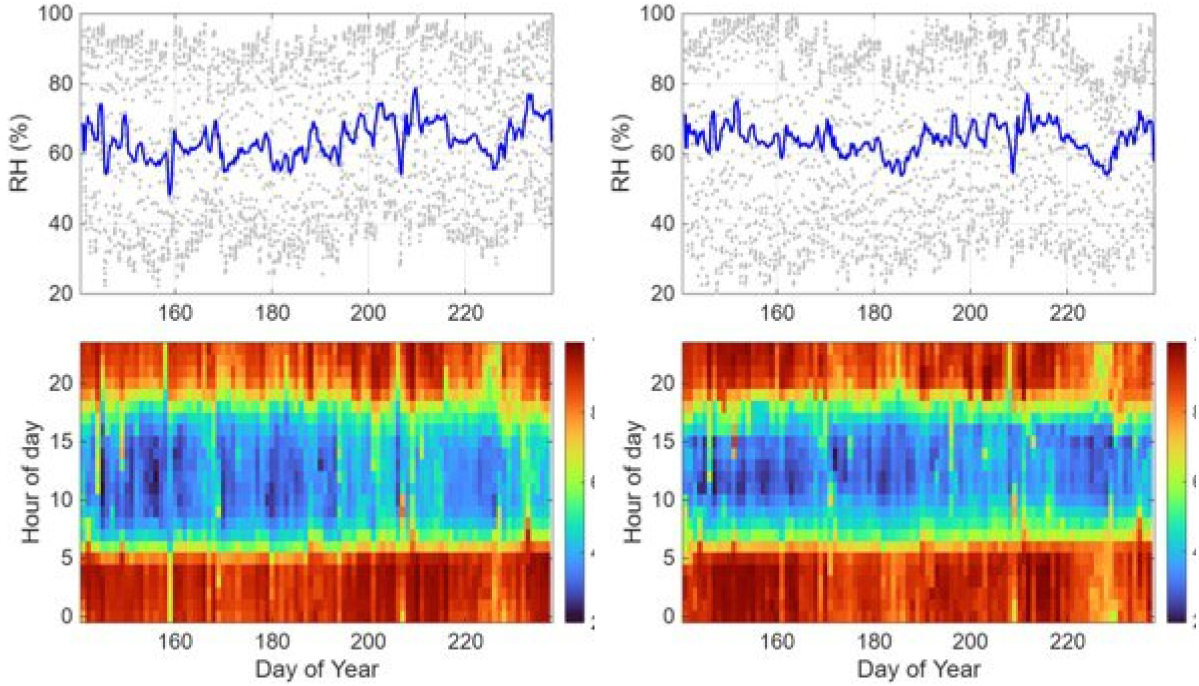


Fig. 2.13 Time trend and carpet plot of the relative humidity measured and simulated associated to the zone 1 of the greenhouse located at FSC. The x-axis shows the days of the year corresponding to the simulation period, which extends from 21 May 2025 to 25 August 2025

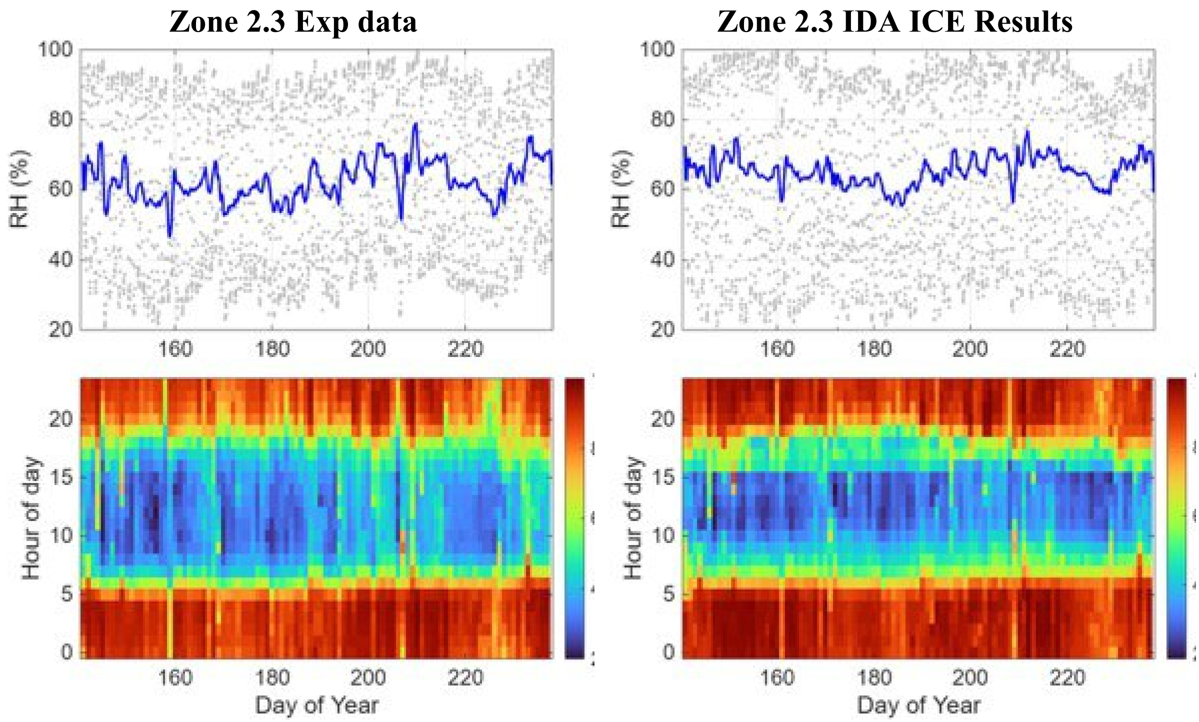


Fig. 2.14 Time trend and carpet plot of the relative humidity measured and simulated associated to the zone 2 of the greenhouse located at the "Fattoria Solidale del Circeo" farm. The x-axis shows the days of the year corresponding to the simulation period, which extends from 21 May 2025 to 25 August 2025.

The error indicators, including the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE), reveal moderate deviations, which are expected given the complexity of moisture transport and generation mechanisms. The Mean Bias Error (MBE) and the Normalized Mean Bias Error (NMBE) highlight a tendency toward RH overestimation in some subzones,

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

particularly under high-humidity conditions; however, the overall magnitude of the bias remains within acceptable limits for dynamic greenhouse simulations.
The carpet plots and time trends of both experimental and simulated relative humidity during the simulation period are presented in the following figures.

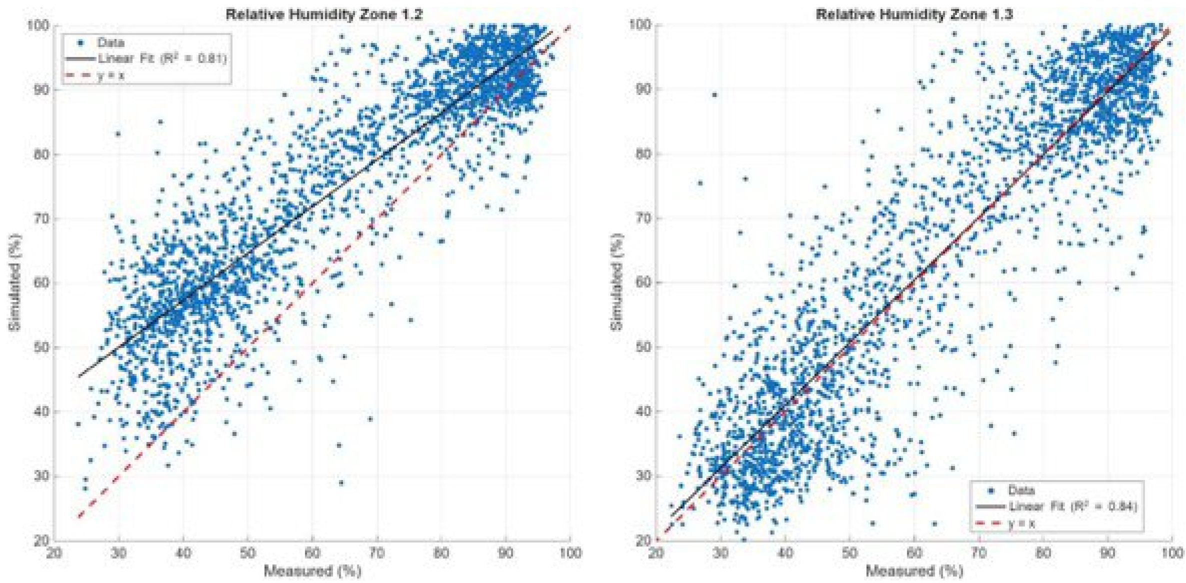


Fig. 2.15 Correlation plots of simulated relative humidity vs measured relative humidity of subzones 1.2 and 1.3 of the greenhouse located at FSC.

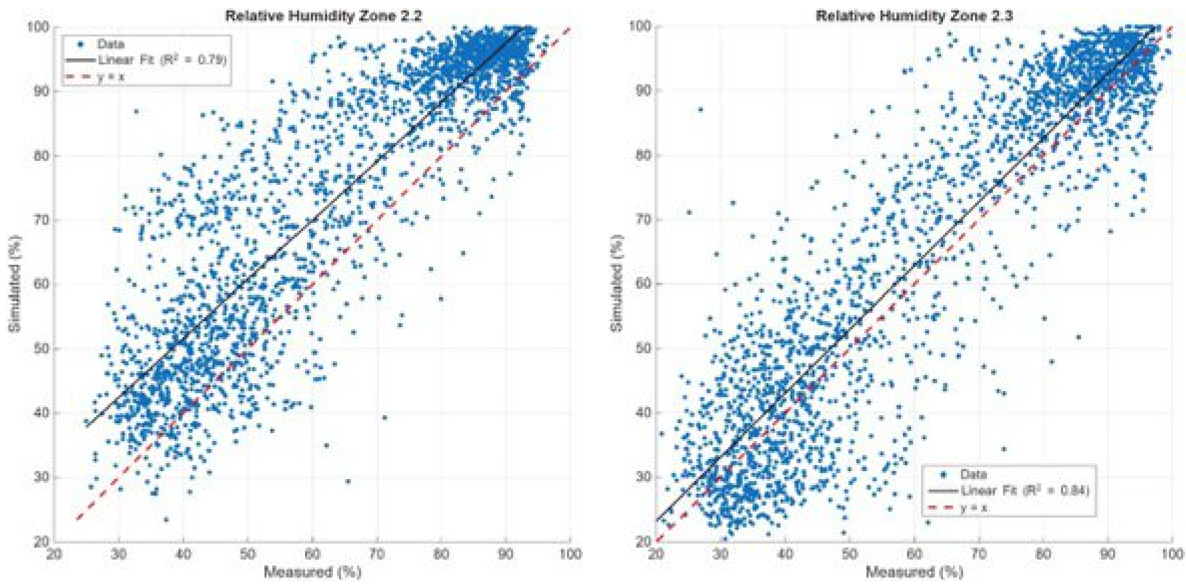


Fig. 2.16 Correlation plots of simulated relative humidity vs measured relative humidity of subzones 2.2 and 2.3 of the greenhouse located at FSC.

Conclusions

This chapter presented the development, implementation, and validation of a detailed dynamic simulation model aimed at reproducing the microclimatic conditions of a naturally ventilated photovoltaic greenhouse under real operating scenarios.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

The modeling activity was carried out using the IDA ICE simulation software, with a strong emphasis on the accurate representation of both external boundary conditions and internal crop–microclimate interactions, which are known to play a dominant role in greenhouse climate dynamics.
A refined vertical zoning strategy was adopted, dividing each greenhouse sector into multiple subzones along the height, to capture the thermal and hygrometric stratification observed in the experimental measurements.
A major contribution of this work lies in the development of a site-specific climatic file, specifically tailored to the location of the Fattoria Solidale del Circeo.
The climatic inputs used in this study were derived from high-resolution measurements collected by an on-site experimental weather station. These data were carefully processed, validated through comparison with a regional meteorological station, and converted into a format compatible with IDA ICE requirements.
Particular attention was devoted to the decomposition of global solar radiation into its direct and diffuse components, as well as to the temporal aggregation of the measurements. This procedure ensured that the external boundary conditions driving the simulations were highly representative of the actual environmental conditions experienced by the greenhouse during the monitored period, significantly enhancing the reliability of the simulation results.
Equally important is the methodological advancement introduced in the modeling of crop evapotranspiration, which represents one of the most challenging aspects of greenhouse microclimate simulation.
In this work, evapotranspiration was explicitly modeled using a Penman–Monteith-based formulation implemented within a custom simulation zone. Compared to previous modeling approaches, a substantial improvement was achieved by introducing time-dependent aerodynamic and canopy conductance, allowing these parameters to evolve throughout the cultivation cycle in accordance with crop growth and physiological activity. In addition, the Leaf Area Index (LAI) was modeled as a time-varying parameter based on experimentally measured data and implemented through dedicated source tables. This approach enabled the model to realistically capture the evolution of the transpiring surface and the associated latent heat fluxes, which strongly influence internal temperature and relative humidity levels.
The combined effect of a refined climatic input and an enhanced representation of evapotranspiration processes resulted in a physically consistent and robust description of greenhouse microclimate behavior.
The validation analysis confirmed a high level of agreement between simulated and measured air temperature values, with coefficients of determination approaching unity and regression slopes close to one for both analyzed zones. Relative humidity predictions, while inherently more complex due to the interplay of ventilation, moisture transport, and biological processes, also showed satisfactory agreement with experimental observations, with robust correlation levels across the analyzed period.
The remaining discrepancies observed during periods of rapid environmental variation can be reasonably attributed to uncertainties in sensor measurements, simplifications in the representation of airflow patterns, and the intrinsic variability of crop physiological responses. Nevertheless, the overall performance of the model demonstrates the effectiveness of the proposed methodological improvements.
In conclusion, this work highlights the critical importance of accurate climatic boundary conditions and advanced evapotranspiration modeling in the dynamic simulation of greenhouse environments. The validated modeling framework provides a solid and reliable basis for future studies aimed at evaluating alternative ventilation and climate control strategies, optimizing greenhouse operation, and investigating the interaction between crop development and

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

environmental conditions. More broadly, the approach presented in this chapter contributes to the advancement of simulation-based tools for sustainable greenhouse design and management, with potential benefits in terms of microclimate control, crop productivity, and energy efficiency.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

References

R.G. Allen, L.S. Pereira, D. Raes, M. Smith, Crop Evapotraspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56, FAO - Food and Agriculture Organization of the United Nations, 1998.

Cornaro, C., Petitta, M., Bovesecchi, G., Miraglia Fagiano, P., Voinea, C., Fornari, W., et al. (2025). Simulating the microclimate of a pilot greenhouse for testing innovative agri-voltaic system technology. Citation Proceedings. presented At 6th Building Simulation Applications Conference (BSA 2024), Bolzano.

Katsoulas, Nikolaos, and Cecilia Stanghellini. 2019. "Modelling Crop Transpiration in Greenhouses: Different Models for Different Applications." Agronomy 9 (7): 392. https://doi.org/10.3390/agronomy9070392.

Monteith, JL. 1965. "Evaporation and Environment. 19th Symposia of the Society for Experimental Biology. University Press, Cambridge, 205-234." Symp Soc Exp Biol.

Baglivo, Cristina, Domenico Mazzeo, Simone Panico, Sara Bonuso, Nicoletta Matera, Paolo Maria Congedo, and Giuseppe Oliveti. 2020. "Complete Greenhouse Dynamic Simulation Tool to Assess the Crop Thermal Well-Being and Energy Needs." Applied Thermal Engineering 179 (October):115698. https://doi.org/10.1016/j.applthermaleng.2020.115698.

Haofang Yan, Shuaishuai Deng, Chuan Zhang, Guoqing Wang, Shuang Zhao, Mi Li, Shaowei Liang, Jianhui Jiang, Yudong Zhou. Determination of energy partition of a cucumber grown Venlo-type greenhouse in southeast China. Agricultural Water Management 276 (2023) 108047. https://doi.org/10.1016/j.agwat.2022.108047

Yan, H., Acquah, S.J., Zhang, C., Wang, G., Huang, S., Zhang, H., Zhao, B., Wu, H., 2019. Energy partitioning of greenhouse cucumber based on the application of PenmanMonteith and Bulk Transfer models. Agric. Water Manag. 217, 201–211. https://doi.org/10.1016/j.agwat.2019.02.036.

Xuewen G., Rangjian Q., Baozhong Z., Shunsheng W., Jiankun G.,Shikai G., Zaiqiang Y. Energy budget for tomato plants grown in a greenhouse in northern China. Agricultural Water Management 255 (2021) 107039. https://doi.org/10.1016/j.agwat.2021.107039

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Chapter 3. CFD modelling of the greenhouse microclimate

Introduction

CFD is one of the tools used in REGACE to study the operation of PV systems inside a greenhouse driven by a responsive tracking system. The advantages it offers over other methods can be summarized as follows:

a) it can provide values of climatic parameters (flow field, temperature, radiation, humidity) over the entire domain under investigation, and not only at specific measurement points;
b) it avoids the simplifications that one-dimensional or lumped models are forced to adopt;
c) it can be validated with a relatively small number of measurements, without the need for large datasets, since it accurately simulates the physics of transport phenomena.

It can be considered a virtual experiment that saves both time and cost.
The role of CFD within the project is to:

- Qualitatively and quantitatively describe the development of flow, temperature, and radiation fields (in different wavelength bands) inside greenhouses equipped with PV systems, with and without the presence of crops, under different climatic conditions.
- Describe how the radiation available at plant level and on PV surfaces is affected by external climatic conditions, greenhouse cover materials and geometry, PV inclination, and the interaction between PV panels and plants.
- Support the design of PV panel spatial arrangements and provide input for the development of responsive tracking system software, based on the predicted effects of PV presence and operation on the internal microclimate.
- Carry out parametric studies that can be used to train the digital twin model operating in conjunction with the CFD model.

The objectives are to:

- Develop a 2D CFD model capable of predicting radiation propagation inside a greenhouse equipped with PV systems and validate the model with respect to radiation simulation.
- Predict radiation propagation and the development of temperature and flow fields in 2D geometries under various crops and climatic conditions.
- Develop a 3D CFD model capable of predicting radiation propagation and the development of temperature and flow fields in greenhouse compartments equipped with PV systems, crops, axial fans, and evaporative pads, enabling full 3D simulations.

3.1 Available Data

The simulated greenhouse is described in Section 1.1. The materials considered in the simulation inside the greenhouse are as follows:

- Cover: The roof cover is a polyethylene sheet with a thickness of 0.018 cm (180 μm).
- Side walls: The side walls are constructed of corrugated polycarbonate sheets with a thickness of 0.08 cm.
- Ground: The ground is divided into two areas: a 2.9 m wide strip at the front of each chamber, near the axial fans, used as a corridor and covered with concrete; and the remaining part of the chamber, where the ground consists of soil covered with geotextile.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

- Hydroponic bench: The hydroponic bench consists of a metal base on which the hydroponic substrate is placed. The substrate is composed of rock wool enclosed in a plastic casing (white/black plastic film).
- PVs: Monocrystalline bifacial PV panels with dimensions of 1.1 m × 0.53 m and a thickness of 5.18 mm. Each panel consists of 36 PV cells (0.105 m × 0.105 m), with the composition described in Table D.1 of Annex D.
- Plants: During the experiments, cucumbers were used as cultivated crops.
- Evaporative pads: At the rear side of each chamber, evaporative pads are installed with dimensions of 6.2 m × 2 m. Munters 60–30 pads with a thickness of 10 cm, made of cellulose paper with a porosity of 0.947, were used.
- Axial fans: In the front part of each chamber, an axial fan with an opening diameter of 1.4 m and overall dimensions of 1.4 m × 1.4 m is installed.
- Material properties: Detailed thermophysical and optical properties of the materials considered are provided in Annex C.

During the REGACE project, three experiments were carried out with the presence of PV systems. However, crops were already grown, and the internal climate was monitored before the installation of the PVs. Table 3.1 presents the timetable of greenhouse operation. The first, second, and third experiments were conducted with cucumber crops. Although the experiment continued beyond these phases, the subsequent results were not used for model validation.

Table 3.1 Timetable of UTH greenhouse operation

Period	From	To	GH2	GH3	GH4	GH5
Without PV	08/03/2024	13/05/2024	No crop	No crop	No crop	No crop
	14/05/2024	10/06/2024	Crop	Crop	Crop	Crop
	11/06/2024	11/09/2024	No crop	No crop	No crop	No crop
After PV installation 1st exp.	12/09/2024	12/12/2024	Crop	Crop	Crop + PV (no responding system)	Crop + PV (no responding system)
No crop	13/12/2024	15/01/2025	No crop	No crop	No crop + PV	No crop + PV
2nd exp.	16/01/2025	03/04/2025	CO_2	Crop	Crop + PV	Crop + PV + CO_2
No crop	04/04/2025	24/04/2025	No crop	No crop	No crop + PV	No crop + PV
3rd exp.	25/04/2025	13/06/2025	No crop	Crop	Crop + PV	Crop + PV

3.2 Overview of the CFD Modelling Approach

3.2.1 General modelling strategy and assumptions

For the development of the CFD model, numerous submodules were implemented and incorporated into the commercial CFD code FLUENT in the form of User-Defined Functions (UDFs). First, a model was developed to calculate normally incident solar radiation on each computational cell of the Gothic-type roof cover. Subsequently, another model was implemented to calculate the shading effects caused by the roofs of adjacent greenhouse chambers.

A key requirement was the ability to define appropriate radiation boundary conditions on the side walls of internal chambers, enabling the simulation of only one or two chambers instead of all six. This approach allowed for a finer spatial and volumetric discretization, which would

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

not have been feasible if all six chambers had been included. This was achieved through the implementation of a radiation transfer equation.
Another important aspect was the definition of suitable boundary conditions for the ground. To this end, an energy balance model was developed to calculate the ground surface temperature. For unsteady simulations, or for steady-state cases in which external climatic conditions are unknown, it was also necessary to develop a model for the estimation of external air temperature.
Since one of the main objectives of the simulations was the accurate prediction of the radiation field, the correct optical properties of all involved materials, including air, had to be considered. The optical properties of air depend on its water vapor content; therefore, a dedicated model was developed to calculate air optical properties as a function of air humidity.
Within the greenhouse, two main factors govern the development of the microclimate: the PV panels and plants. PV panels induce shading and interact with the air through reflection and thermal emission, while also modifying the flow and temperature fields. Plants, in turn, affect the flow and temperature fields, as well as relative humidity and CO_2 concentration, and reflect and emit radiation. Accordingly, two specific plant-related models were developed:

- an evapotranspiration model;
- a radiation interaction model.

3.2.2 Radiation modelling and shading treatment

External incident radiation is one of the fundamental boundary conditions of the CFD model. For model validation, measured external radiation data were used to calculate the radiation incident normal to the greenhouse cover and side walls, considering both beam and diffuse components in four wavelength bands: UV (0.1–0.4 μm), PAR (0.4–0.76 μm), NIR (0.76–1.1 μm), and IR (1.1–1000 μm). The use of four wavelength bands is necessary because plants primarily utilize PAR, while PV systems operate in both PAR and NIR ranges.
To account for external incident radiation, two dedicated UDFs were developed. The detailed mathematical formulation of these UDFs is presented in Annex A. When specific measurements of external radiation are not available, the radiation incident on a horizontal plane is calculated from standard climatic data. In this case, the relative model based on Duffie (2006), and analytically presented by Baxevanou (2010), was adopted, considering the monthly average total and diffuse solar energy incident on horizontal surfaces.
For the calculation of shading caused by adjacent greenhouse roofs, the following procedure is applied:

- For each time step, the plane defining the solar azimuth and perpendicular to the horizontal plane is determined.
- The ellipse resulting from the intersection of this plane with a cylinder representing the adjacent roof is then identified.
- For each point on the roof, the angle of the line passing through that point and tangent to the ellipse is calculated.
- If this angle is smaller than the solar elevation angle, the point is considered shaded (Flag = 0) at that time; otherwise, the point is considered unshaded (Flag = 1).

As will be discussed in the grid development section, in order to generate a sufficiently dense mesh to accurately resolve transport phenomena near solid surfaces, within solids, and in porous materials, the computational domain was limited to one or two greenhouse chambers. In the case of multiple greenhouses, which is the typical configuration in commercial

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

facilities—the main source of uncertainty lies in the boundary conditions applied at the side walls, particularly for radiation in the UV and NIR wavelength bands.
The procedure adopted to address this issue was as follows:

a. Multiple simulations of a six-chamber greenhouse were performed under different external climatic conditions, at various times of the day and for different days.
b. The radiation in the PAR and NIR bands incident on the side walls of chambers 4 (left side wall) and 5 (right side wall) was calculated.
c. A regression analysis was performed using the total radiation incident on a horizontal plane and the solar elevation angle as independent variables, and the PAR and NIR radiation on the side walls as dependent variables, in order to derive a transfer function. This function allows the estimation of side-wall radiation in general cases where only external radiation and solar elevation are known.
d. The resulting values were applied as total diffuse radiation boundary conditions on the side walls of the two-chamber computational geometry.

Further details of the methodology are provided in Annex A, while the results of the regression analysis are reported in Table 3.2.

Table 3.2 Coefficients of regression analysis

Coefficient	Left PAR	Left NIR	Right PAR	Right NIR
B0	1.927485209	-11.85194593	8.422348573	-8.233466166
B1	1.617224319	0.250962399	1.395276544	0.051657124
B2	0.439892051	1.210131394	0.463801558	1.285292506

3.2.3 Representation of PV panels

PV panels are modeled as realistically as possible. They are treated as semi-transparent solid media and positioned at their actual locations within the computational domain. Since their real thickness (5.18 mm) is too small to be adequately resolved by the computational grid, the panels are represented as flat plates with an effective thickness of 1 cm. The real and effective thermophysical properties of the PV panels (as required for the simulations) are reported in Annex C.
Each panel is discretized into 10 horizontal and 22 vertical computational cells, providing a sufficiently fine mesh for the calculation of temperature and radiation evolution within the solid. The panel surfaces are both thermally and optically coupled with the surrounding semi-transparent walls, and the energy and radiation equations are solved within the solid material.
Each PV panel consists of two main components: an opaque layer containing the PV cells and a transparent structural layer. From an optical standpoint, the entire panel is treated as a homogeneous mixture of these two components, as detailed in Annex C.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

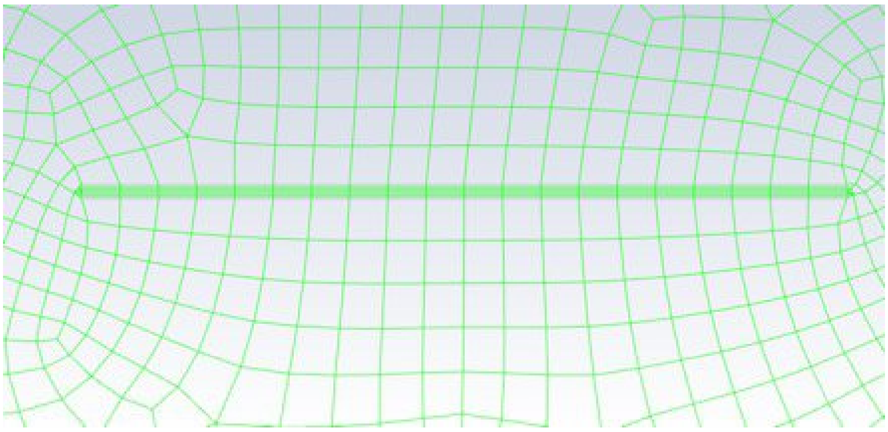


Fig. 3.1 Detail of PV panel computational grid.

3.2.4 Representation of crops and evapotranspiration

Crops interact with the temperature, radiation, and flow fields in several ways. From a fluid dynamics perspective, crops are treated as porous media composed of air and solid elements (leaves, fruits, and stems). As porous media, crops induce pressure losses by introducing a sink term in the momentum equations and modify both the heat transfer coefficient in the energy equation and the effective viscosity in the turbulence equations (Fidaros, 2010; Baxevanou, 2018; Baxevanou, 2020). Details regarding the effects of crops, modeled as porous media, on the momentum, turbulence, and air heat transfer equations are provided in Annex A.
With respect to radiation, plants absorb PAR radiation and exchange energy with the surrounding environment and greenhouse surfaces through reflection and emission of IR radiation, according to their temperature. In CFD simulations, plants are modeled as porous media consisting of a fluid fraction (the greenhouse air, defined by the porosity γ) and a solid fraction (leaves, fruits, and stems). Since only the fluid phase of the porous medium is explicitly accounted for in the calculations, radiation exchange associated with the plants is incorporated by defining an "effective air" with equivalent optical properties. These properties are chosen such that, when applied to the porous volume, they reproduce the same optical behavior as the real crop canopy. This modeling approach is adopted here, and further details on material properties and effective optical properties are provided in Annex A (Fidaros, 2010; Baxevanou, 2018).
Finally, plants absorb sensible heat and release water vapor through evapotranspiration. This process depends on plant growth parameters, expressed in terms of plant height and leaf area density (LAD), as well as on absorbed PAR radiation and ambient air temperature and humidity. Accordingly, two additional UDFs were developed: one to calculate the sink term in the energy equation and one to calculate the source term in the air–water vapor scalar transport equation. Both UDFs are described in detail in Annex A.

3.2.5 Treatment of airflow, heat and mass transfer

Another important issue in greenhouse simulations concerns the definition of ground boundary conditions. In some studies, the ground is assumed to be adiabatic, which is unrealistic, or modeled as a surface with a constant heat flux, while others assume an isothermal surface, which is more appropriate. In this latter case, however, the ground temperature must be properly defined. To address this, an energy balance model was developed in the form of a UDF to calculate ground surface temperature as a function of the day of the year, incident

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

radiation, ground properties, and the air temperature inside the greenhouse. The energy balance model and the corresponding UDF are described in detail in Annex A.
For the general case in which the external air temperature at a given time is unknown, and simulations must rely on available climatic data, the external air temperature is computed using an additional UDF. The calculation procedure for estimating external air temperature as a function of the day of the year and the hour of the day is presented in Annex A. Two separate UDFs were implemented, one for daytime conditions and one for nighttime conditions.
Air is a mixture of dry air and water vapor, and its optical properties depend on the water vapor content. For this reason, a UDF was developed to calculate the optical properties of air as a function of temperature and humidity. The calculation methodology and the corresponding UDF are presented in Annex A.
Finally, accurate modeling of the axial fans and evaporative pads is required (Fidaros, 2018). In many relevant studies, axial fans and evaporative pads are not explicitly included in the computational domain; instead, they are replaced by prescribed values of air velocity, temperature, and water vapor concentration.
In the present work, axial exhaust fans are modeled using a pressure jump approach, calculated from polynomial expressions as a function of fan rotational speed, as detailed in Annex A. Evaporative pads are modeled as porous media that induce pressure losses, act as heat sinks, and serve as sources of water vapor, as described by Fidaros (2018). Their thermophysical properties and aerodynamic behavior are calculated analytically following Franco (2011). A detailed description of the evaporative pad model is provided in Annex A.

3.3 Simulation domains and numerical setup

3.3.1 2D and 3D computational domains

A total of six 2D grids and one 3D grid were used for the simulations. All computational domains include only the greenhouse structure and its internal components.
The first grid includes all six greenhouse chambers and is a 2D grid defined on a central cross-section. The objectives of this grid were:

a. to validate the basic modeling approach and the description of the optical properties of the cover, side walls, ground (including the energy balance model), benches, and radiation modeling; and
b. to develop a “transfer function” to define radiation boundary conditions on the side walls of internal chambers.

This first six-chamber cross-sectional grid is presented in Figure B.1 of Annex B. It consists of approximately 75,000 computational cells and is applicable to cases without crops and without PV panels installed inside the greenhouse. In this configuration, only the metal benches are present, without hydroponic substrates, except in the first chamber where electromechanical equipment is installed. Consequently, grid refinement is required only near the benches, ground, roof, and side walls. In these regions, a wall resolution of $y^+ < 40$ is ensured. The roof cover and side walls are modeled as solid materials with a thickness of 10 cm, requiring the calculation of equivalent material properties. Within the solid regions, at least 10 computational cells are used, ensuring appropriate cell aspect ratios.
The second set of grids includes only chambers 4 and 5, where the PV panels were installed. Three 2D grids were developed: one with the PV panels in a horizontal position, one with the panels inclined by 60° toward the east, and one with the panels inclined by 60° toward the west. This inclination corresponds to the maximum tilt of the tracking system. These grids are

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

presented in Figures B.2, B.3, and B.4 of Annex B, respectively, each consisting of approximately 45,000 computational cells.
In the initial six-chamber grid, each chamber contains approximately 12,500 computational cells. In contrast, in the two-chamber grids, each chamber contains approximately 22,500 computational cells, allowing for increased grid refinement near the solid boundaries of the PV panels.
Two additional 2D grids were developed for chambers 4 and 5: one including crops at an average height of 2.1 m (Figure B.5), and one including both crops and horizontally positioned PV panels (Figure B.6). The grid including only crops consists of approximately 75,200 computational cells (37,600 cells per chamber), while the grid including both PV panels and crops consists of approximately 100,000 computational cells (50,000 cells per chamber).
The 2D grids were primarily used for validation of the radiation models, although the continuity, momentum, and energy equations were also solved, yielding flow and temperature fields. However, since the simulated cross-sections do not include openings, axial fans, or evaporative pads, the resulting airflow is driven solely by thermal buoyancy. This configuration corresponds to winter operating conditions, during which ventilation openings are typically closed.
Finally, the 3D grid, presented in Figure B.7, includes a single chamber (chamber 4) equipped with PV panels and crops, and explicitly incorporates the axial exhaust fans and the evaporative pad within the computational domain. This grid consists of approximately 1,500,000 computational cells.

3.3.2 Key boundary conditions

The external cover surface is modeled as a semi-transparent wall with a mixed boundary condition. This boundary condition specifies the external air temperature, the convective heat transfer coefficient, the equivalent sky temperature, and the incident radiation. This surface is thermally and optically coupled to a solid layer representing the transparent cover material, in which both the energy and radiation equations are solved. The cover solid is coupled on its external side with the external cover surface and on its internal side with a semi-transparent wall representing the internal cover surface. The internal cover surface is, in turn, thermally and optically coupled to the cover solid material on one side and to the greenhouse air on the other side.
For the six-chamber geometry, the side walls are treated in the same manner. For the one- and two-chamber geometries, the side walls are modeled as adiabatic from a thermal standpoint, while prescribed radiation boundary conditions are applied.
The ground is modeled as opaque and isothermal. PV panels and benches are treated as semi-transparent solids (with optical properties ensuring radiation absorption) and are thermally coupled with both the greenhouse air and the solid materials. Plants are modeled as porous media, inducing pressure losses, acting as heat sinks and sources of water vapor, and exhibiting semi-transparent radiative behavior. Axial fans are modeled as pressure jumps. Evaporative pads are modeled as porous media, acting as pressure loss elements, heat sinks, and sources of water vapor.

3.3.3 Numerical solution strategy

The simulation procedure followed in this study can be summarized as follows:

a) Simulation of the 2D six-chamber geometry without accounting for shading effects.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

b) Simulation of the 2D six-chamber geometry including shading effects.
c) Comparison of the results from steps (a) and (b) to assess the importance of shading.
d) Validation of the radiation models up to this stage.
e) Simulation of the 2D six-chamber geometry to collect data for the development of the transfer equation.
f) Development of the transfer equation.
g) Simulation of the 2D two-chamber geometry including PV panels.
h) Validation of the radiation model including PV panel modeling.
i) Generation of results concerning the available PAR radiation at plant level, with and without PV panels, and for different PV inclination angles.
j) Simulation of the 2D two-chamber geometry including plants.
k) Validation of the radiation model including plant modeling.
l) Validation of the evapotranspiration model.
m) Simulation of the 2D two-chamber geometry including both PV panels and plants.
n) Simulation of the 3D one-chamber geometry.
o) Generation of results concerning flow and temperature fields.

All CFD simulations were performed using the commercial CFD software ANSYS Fluent (Software1). The Reynolds-Averaged Navier–Stokes (RANS) equations were solved using the finite volume method. For the 2D simulations, the flow was assumed to be steady-state, while unsteady simulations were performed for the 3D case. The flow was considered incompressible and turbulent.
Turbulence was modeled using the standard high-Reynolds-number k–ε model with wall functions. The governing equations solved include the continuity equation, momentum equations, energy conservation equation, radiation transfer equation (RTE), and a scalar transport equation for water vapor concentration. The energy equation was also solved within solid materials, where it reduces to the heat conduction equation.
Pressure–velocity coupling was achieved using the SIMPLEC algorithm. Radiation was modeled using the Discrete Ordinates (DO) method, considering four wavelength bands: UV (0.01–0.4 μm), PAR (0.4–0.76 μm), NIR (0.76–1.1 μm), and IR (1.1–1000 μm). The subdivision into four wavelength bands was adopted not only to better capture variations in optical properties, but also because plants primarily utilize PAR (Photosynthetically Active Radiation), while PV panels operate in the PAR and NIR ranges.
The discretization schemes employed were as follows:

a) Spatial discretization of convective terms:
  (i) second-order upwind scheme for momentum, turbulence, RTE, and scalar transport equations;
  (ii) third-order MUSCL scheme for the energy equation.
b) Spatial discretization of diffusive terms: second-order schemes for vector and scalar quantities, and first-order for radiation.
c) Temporal discretization: second-order implicit scheme with a time step of one minute.

The convergence criterion was set to a residual value of $\varepsilon \leq 10^{-6}$.

3.4 Simulated Scenarios

3.4.1 Overview of investigated cases

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

For the validation of the developed models, measurements and climatic conditions from the following days were used.

Table 3.3 CFD simulations validation days

Day	Conditions
28/04/2024	No crop No PV
14/05/2024	Crop Not PV
09/06/2024	No crop No PV
11/10/2024	Crop and PV
17/12/2024	No Crop with PV
18/04/2025	No Crop with PV

3.4.2 Seasonal conditions

Table 3.4 External climatic data 28/04/24

Hour, Zone time [h]	Solar Time [h]	Total radiation [W/m^2]	Diffusive radiation [W/m^2]	Temperature [C]	Wind velocity [m/s]	Relative Humidity [%]
9	7,56	308.37	63	14	0.73	
10	8,56	548.14	66.0	17.3	0.5	63.4
11	9,56	737.67	76.0	21.6	0.5	
12	10,56	846.35	570.0	24.4	0.53	41.9
13	11,56	888.46	859.0	24.2	0.77	
14	12,56	865.36	789.0	24.6	1.3	
15	13,56	816.45	719.0	24.8	1.07	
16	14,56	684	96.0	24.8	0.57	29.6
17	15,56	513.51	66.0	24.5	0.53	
18	16,56	322.64	54.0	23.7	0.27	37.6

Table 3.4 provides indicative external climatic data for selected hours used in the simulations for 28/04/2024, while Table 3.5 reports the external climatic conditions for 17/12/2024.

3.4.3 Rationale for scenario selection

The days selected for model validation were required to meet four main criteria:

(a) they had to correspond to the operating conditions of interest;
(b) they had to represent typical, yet distinct, climatic conditions;
(c) climatic variables—particularly solar radiation—had to exhibit a smooth temporal evolution consistent with the theoretical daily pattern. This requirement implies avoiding days characterized by alternating cloudiness and sunshine, and selecting days for which the total solar radiation shows an approximately sinusoidal daily variation; and
(d) complete datasets had to be available from both the external meteorological station and the internal greenhouse sensors.

An additional constraint concerned crop height, since different plant heights require different computational grids.
Based on these criteria, the following days were selected: summer days (09/06/2024) with relatively low radiation for the season; winter days with unusually high radiation levels (17/12/2024); representative spring days for which energy consumption measurements were also available, with both high (18/04/2025) and low (28/04/2024) radiation and low diffuse radiation; and days characterized by high diffuse radiation during spring (14/05/2024) and autumn (11/10/2024).

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

The autumn case was selected because, during this period, both crops and PV panels were present, with the PV panels in a horizontal position, as the automatic tracking system had not yet been activated. Consequently, all relevant operating configurations—no PV panels and no crops, crops only, PV panels only at different inclination angles, and combined PV panels and crops—were covered using days for which complete and reliable measurement datasets were available.

Table 3.5 External climatic data 17/12/24

Hour, Zone time [h]	Solar Time [h]	Total radiation [W/m²]	Diffusive radiation [W/m²]	Temperature [C]	Wind velocity [m/s]	Relative Humidity [%]
9	8,59	489	37	276,65	6	85,2
10	9,59	694	296,0	278,75	8	78,6
11	10,59	813	390,0	280,95	6	74
12	11,59	822	54,0	282,55	6	65,8
13	12,59	738	43,0	284,55	0,1	56,6
14	13,59	532	45,0	285,55	8	56,9
15	14,59	279	35,0	285,85	0,1	56,3
16	15,59	28	22,0	285,85	8	56,7

3.5 Model Validation

3.5.1 Validation methodology

A careful, step-by-step methodology was adopted to validate the developed models. For the purposes of the REGACE project, the validation focused primarily on the radiation model.
In the first step, the radiation model was validated for the six-chamber geometry in the absence of both PV panels and crops. To this end, simulations of one representative summer day and one representative winter day were performed, and the calculated radiation values were compared with measured radiation at the sensor locations. Since the simulations were two-dimensional, measurements from sensors located at the mid-plane of the chambers were used. For the winter case, the validation focused on measurements from sensors positioned above the PV panels (which were already installed) as well as from sensors located in chambers without PV panels.
After the radiation model was validated for empty chambers, the validation procedure was extended to chambers equipped with PV panels at various inclination angles. For this purpose, a winter day and a spring day were simulated, with particular emphasis on the spring day characterized by high external radiation levels. Subsequently, the radiation model was validated for the case of crops only, using a representative spring day and a crop height of 2 m.
When radiation measurements from the REGACE system sensors were used, the sum of PAR and NIR radiation was considered for validation. When measurements from the Sercom and Trisolar sensor systems were used, only PAR radiation was used for comparison.

3.5.2 Comparison with measured radiation and climate variables

The six-chamber model was validated against experimental measurements using simulations performed for two representative days.
Figures D.9 to D.12 present the comparison between measured and simulated PAR+NIR radiation for sensors I4t, I2, I4s, and I3 on 09/06/2024. This day corresponds to conditions with

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

no crop grown in the greenhouse and no PV panels installed. The validation is limited to the time periods when the shading curtain was not deployed. The accuracy achieved is very satisfactory, as the differences between measured and simulated values remain below 10%, with improved agreement around midday. These simulations correspond to the six-chamber geometry.

Figures D.22 and D.23 show the comparison between measured and simulated PAR radiation for sensors PAR8, PAR11, P35, and P5 on 17/12/2024. On this day, no crops were present, while PV panels were installed. For this reason, the validation focused on sensors either located in chambers without PVs or positioned above the PV panels, thus minimizing their shading influence. The accuracy achieved is lower, particularly around noon. It should be noted that the PAR sensors are not located exactly at the central cross-section of the greenhouse. Moreover, for sensors positioned above the PV panels (P35 and P5), the simulated PAR appears to be underestimated, likely because PV reflection effects are not yet accounted for in the model. Nevertheless, the temporal trends are well reproduced. These simulations correspond to the two-chamber geometry.

Figures D.9 to D.17 present the comparison between measured and simulated PAR and PAR+NIR radiation for sensors I2, I4s, I3, P35, P39, PAR10, P5, P9, and PAR11 on 18/04/2025, under conditions with PV panels installed at different inclinations and no crop present. The simulations correspond to the two-chamber geometry. The agreement is excellent, with differences generally below 5%, especially in areas located above the PV panels.

Finally, Figures D.18 to D.21 show the comparison between measured and simulated PAR+NIR radiation for sensors I2, I4s, I3, and I4t under conditions without PV panels but with crop present. The simulations were carried out using the two-chamber geometry, with plants modeled at a height of 2 m. The agreement is satisfactory, with differences limited to within 12%.

3.5.3 Model performance and main sources of uncertainty

Given that the comparison is performed against field measurements, the achieved accuracy can be considered satisfactory. The observed differences between measured and simulated values can be attributed to the following factors:

a) Different time steps and integration times between external and internal sensors. This can explain the discrepancies observed at the beginning and end of the day, as well as during episodes of sudden local shading.

b) Sensor positioning: the sensors are not located on the same vertical plane at the center of the chambers. While this does not significantly affect cases with empty chambers, it can influence cases with PV installations, since a sensor may be shaded in one vertical plane but not in a neighboring one.

c) Directional sensor response: the sensors measure radiation from a single direction and therefore do not capture reflected radiation. This effect may be particularly relevant for sensors positioned above the PV panels.

d) Greenhouse structure simplifications: the metal frame of the greenhouse is not included in the simulations, and thus transient shading effects caused by structural elements are not accounted for. However, given the 10-minute measurement integration time, this omission is not expected to have a significant impact.

e) Limited PV inclination configurations: due to the absence of a moving grid in the model, only specific PV inclinations are simulated. This may introduce discrepancies when the

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

actual PV orientation differs from the three modeled marginal positions (left inclination 60°, horizontal, and right inclination 60°).

f) Simplified crop geometry: plant height and width are represented using average values and therefore do not strictly correspond to the actual plant geometry.

3.6 Key Results and Discussion

3.6.1 Effects of shading and neighboring greenhouses

All Figures D.x.x referred to in the following text are presented in Annex D.

Figures D.22 to D.25 present the isocontours of PAR radiation for 09/06/24, a summer day with relatively low solar radiation for the season, at four characteristic hours (10:00, 12:00, 16:00, and 18:00). The simulations correspond to the six-chamber geometry, without PV panels, without crop, and without accounting for shading from neighboring roofs. Midday hours are not shown because during this period the greenhouse interior is covered by the shading curtain.

It should be noted that the reported times correspond to local time under daylight saving; therefore, the actual solar times are 08:54, 10:54, 14:54, and 16:54 h, respectively. In addition, the greenhouse orientation is NorthEast–SouthWest rather than North–South, resulting in non-symmetrical radiation patterns. These remarks apply to all simulations presented in this section.

During the morning hours, solar rays strike the left side of each roof as well as the left external side of the first chamber. However, with respect to roof radiation, each chamber behaves as if isolated, since shading from neighboring roofs is not considered. Above the benches, a reverse shadow appears, corresponding to the obstruction of ground-reflected radiation (the ground is covered with white geotextile). Near noon on a summer day, solar radiation becomes more uniformly distributed within the chambers, with decreasing intensity toward the ground. Later in the afternoon, solar rays impinge on the right side, and internal shadowed areas increase.

Figures D.26 to D.29 show the isocontours of PAR radiation for 17/12/24, a winter day with relatively high solar radiation for the season, at four characteristic hours (10:00, 12:00, 14:00, and 16:00), again for the six-chamber geometry. These simulations are performed without considering PV panels (although they were installed at the time), without crop, and without accounting for shading from neighboring roofs.

The high fraction of diffuse radiation allows illumination of both end chambers during the morning hours. Additionally, the low solar elevation angle enables deep penetration of beam radiation from the left side. Due to the greenhouse orientation, even during early afternoon hours the sun appears to originate predominantly from the left. However, for a day close to the winter solstice, daylight diminishes rapidly, and by early afternoon the interior illumination is already significantly reduced.

The same simulations were repeated while accounting for shading from neighboring roofs. Figures D.30 to D.33 present the corresponding PAR isocontours for 09/06/24, for the same characteristic hours and six-chamber configuration, without PV panels and without crop. In this case, shadowed regions appear near the neighboring roofs, where beam radiation is completely obstructed and only diffused radiation contributes. Nevertheless, on a summer day these shadowed areas remain limited to an extent.

Figures D.34 to D.37 present the corresponding results for 17/12/24, again without PV panels and without crops, but accounting for shading from neighboring roofs. During winter conditions, the shadowed regions increase significantly, affecting a larger portion of the chambers throughout the day.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

To quantify the observed differences, predicted radiation values at specific locations within the computational domain were compared. Figure D.38 shows the predicted PAR radiation at sensor locations I2, I3, I4s, I4t, and I7, comparing simulations with and without neighboring roof shading. At noon, no differences are observed, since shading effects are negligible on a summer day at this time. However, during morning and afternoon hours, notable differences emerge depending on sensor location.

At sensor I2, located near the left side of the greenhouse at a height of 2 m, the relative difference reaches a maximum of 13% in the morning, vanishes around noon, and increases again to approximately 7% in the afternoon. A symmetrical behavior is observed at sensor I3, located near the right side of the chamber, with differences of about 5% in the morning and a maximum of 15% in the afternoon. Sensors I4s and I4t, positioned near the center of the chamber, are less affected by shading; nevertheless, vertical variations are observed, with the upper sensor exhibiting differences up to 11% in the afternoon. The lower central sensor I4s shows smaller differences, as it is mainly influenced by the overall structure and air optical properties.

Figure D.39 presents the predicted PAR radiation at sensor locations I4t, I7, P35, and P5, again comparing simulations with and without neighboring shading. These locations were selected to represent positions above the PV panel level. On the winter day, the effect of shading is significant throughout the entire day. At central positions above the PV panels, differences begin at approximately 15% in the morning, decrease to a minimum of about 3% around noon, and increase again to nearly 10% in the afternoon. The I4t sensor, located lower than P5 and P35, exhibits smaller differences ranging from 6% to 14%, with a maximum around noon. Finally, sensor I7, located in a chamber without PVs but near the right side, shows differences of about 12% in the morning, reaching a maximum of 17% at noon and decreasing to approximately 1% in the afternoon, when overall radiation levels are already very low.

The quantitative analysis clearly demonstrates that accounting for shading from neighboring roofs is essential for accurate radiation simulations, particularly during winter conditions and in regions close to the chamber boundaries.

3.7. Impact of PV panels on radiation availability

Figures D.40 to D.43 present the isocontours of PAR radiation for 18/04/25, a spring day with relatively high solar radiation for the season, at four characteristic hours (10:00, 12:00, 16:00, and 18:00). The simulations correspond to the two-chamber geometry, equipped with PV panels in horizontal position, without crops, and accounting for shading from neighboring roofs. The hours around noon are not presented due to the deployment of the shading curtain.

In all figures, interference fringes are clearly observed, resulting from multiple reflections between PV panels, the ground, and the metallic benches. Such complex patterns would be impossible to predict using linear or simplified radiation models. Although the PV panels remain in a horizontal position, they do not induce significant shading for most of the day, except during the afternoon hours.

Figures D.44 to D.47 present the corresponding PAR isocontours for the same day (18/04/25) and hours, for the two-chamber geometry equipped with inclined PV panels, without crops and accounting for shading from neighboring roofs. The PV inclination angle is set to 60° toward the left during the morning hours and 60° toward the right during the afternoon hours. Compared to the horizontal configuration, PV inclination improves the radiation distribution inside the chambers, particularly during the afternoon.

Figures D.48 to D.50 show the isocontours of PAR radiation for 17/12/24, a winter day with relatively high solar radiation for the season, at three characteristic hours (10:00, 12:00, and

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

14:00). The hour 16:00 is omitted because the available radiation is already very low. The simulations correspond to the two-chamber geometry, equipped with horizontal PV panels, without crops, and accounting for shading from neighboring roofs.

Figures D.51 to D.53 present the corresponding results for inclined PV panels on the same winter day and hours. PV inclination improves radiation levels during the morning and around noon, while its benefit diminishes during the afternoon due to the very low solar elevation and reduced incoming radiation.

Figure D.54 presents the variation of PAR radiation across the width of chamber 4 for 18/04/25, with PV panels in horizontal position, at three heights above the ground (1 m, 2 m, and 3 m), at zone time 12:00. Figure D.55 shows the corresponding results for the same day and time with inclined PV panels. In both configurations, the average PAR value at 1 m height is slightly lower than at 2 m, which in turn is slightly lower than at 3 m.

At all heights and for both PV configurations, a reduction of available PAR is observed across the chamber width. This is explained by the fact that zone time 12:00 during daylight saving corresponds to a solar time before solar noon, causing solar rays to strike the roof predominantly from the left side. Additionally, at all heights and in both PV configurations, fluctuations in PAR radiation are evident due to the combined effect of PV shading and reflections from the metallic benches. These fluctuations are most pronounced at the 2 m height, while they are reduced in the case of inclined PV panels, which also provide slightly higher average PAR values. Nevertheless, in all cases and across the entire chamber width, PAR radiation remains above 250 $W \cdot m^{-2}$, which can be considered a threshold for crop growth. Therefore, PV panels, regardless of their position, are not expected to negatively affect crop growth under these conditions.

Figure D.56 presents the PAR radiation variation across the width of chamber 4 for 17/12/24, with PV panels in horizontal position, at heights of 1 m, 2 m, and 3 m, at zone time 12:00. Figure D.57 presents the corresponding results for inclined PV panels. On this winter day, differences between the three heights are more pronounced, and radiation fluctuations are stronger, with the highest variability again observed at the 2 m height. PV inclination at 12:00 allows a significantly higher amount of PAR radiation to reach the crop level, marginally exceeding the threshold range of 200–250 $W \cdot m^{-2}$, compared to the horizontal PV configuration.

Figure D.58 shows the PAR radiation distribution across the greenhouse width at a height of 2 m for 18/04/25 at 10:00, for three cases:

a) without PV panels,
b) with PV panels in horizontal position, and
c) with inclined PV panels.

Although the presence of PV panels leads to an average PAR reduction of approximately 35% compared to the case without PVs, the remaining radiation levels are still sufficient and are not expected to negatively affect crop growth.

Figure D.59 presents the corresponding comparison for 17/12/24 at zone time 10:00. The previously observed trend, namely the significant difference between horizontal and inclined PV configurations, is confirmed. While the inclined PV configuration yields PAR levels close to those obtained without PVs, the inclination plays a crucial role in ensuring adequate PAR availability when compared to the horizontal PV configuration, particularly under winter conditions.

3.8 Interaction between PVs, crops and microclimate

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

Figures D.60 to D.63 present the isocontours of PAR radiation for 14/05/24, a spring day characterized by a high fraction of diffuse radiation throughout most of the day. Results are shown for four characteristic hours (10:00, 12:00, 16:00, and 18:00) for the two-chamber geometry, equipped with cucumber plants of 2 m height and without PV panels. Although the simulations correspond to a two-chamber configuration, only one chamber is depicted in the figures for clarity.
Due to the high percentage of diffuse radiation on this day, no clear dominant light transmission direction can be identified, except during the late afternoon (18:00). In all cases, it is evident that PAR radiation is strongly absorbed within the crop canopy, while the regions between the plants are shaded by the plants themselves.
Although the present 2D simulations correspond to an enclosure flow configuration, since no openings are included in this section, it was considered useful to also present the thermal and flow fields. These fields represent a theoretical flow driven exclusively by thermal buoyancy, which can occur during periods when the greenhouse is fully closed, and neither axial fans nor roof openings are operating.
Figures D.64 to D.67 present the temperature isocontours for 14/05/24 at the same four characteristic hours. In all cases, two distinct regions are clearly identified:

(i) the region inside and between the plants, where temperature decreases due to evapotranspiration and shading, respectively, and
(ii) the region above the plant canopy, where temperature increases, as expected in greenhouse environments.

As the available solar radiation increases, the temperature differences between the plant canopy and the surrounding air become more pronounced, reflecting the intensification of evapotranspiration processes.
Figure D.68 presents the velocity magnitude isocontours at 12:00 on 14/05/24. The presence of plants is clearly identifiable through the reduced velocity magnitude within the crop region. The flow exhibits a circular pattern, consistent with motion driven solely by thermal buoyancy. Figure D.69 presents the corresponding streamlines, highlighting the air circulation within the enclosure and the impact of the porous crop on flow structure.
Figures D.70 to D.73 present the PAR radiation isocontours for 11/10/24, an autumn day with a high fraction of diffuse radiation, at four characteristic hours (10:00, 12:00, 14:00, and 16:00). The simulations correspond to the two-chamber geometry, with cucumber plants of 2 m height and PV panels in horizontal position. Although this is an autumn day, it is evident that the PV panels do not cause localized excessive shading directly beneath them; instead, they reduce the available PAR in a relatively homogeneous manner. Nevertheless, for most of the day, PAR levels at plant height remain sufficient. This conclusion is consistent with previous observations and becomes even clearer under predominantly diffuse radiation conditions.
Figures D.74 to D.77 present the temperature isocontours for the same day and hours. Once again, a temperature decrease is observed within the crop canopy due to evapotranspiration and in the space between plants due to shading. Although the temperature of the PV panels increases, their thermal boundary layer remains confined very close to their surfaces and does not significantly affect the temperature of the rest of the domain, at least until noon. From noon onwards, however, the PV panels clearly divide the chamber into two distinct thermal regions:

(i) a warmer region above the PV panels, and
(ii) a cooler region below them.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

When only plants are present, this thermal boundary is located above and around the crop canopy. Since the elevated surface temperature of the PV panels alone is insufficient to fully explain this behavior, the flow field must also be examined.
Figures D.78 to D.81 present the velocity magnitude isocontours and streamlines for 11/10/24 at the same four characteristic hours. A consistent flow pattern is established throughout the day. The PV panels divide the domain into two main regions:

(i) a lower region beneath the panels, characterized by a large recirculation influenced by the porous nature of the plants, and
(ii) an upper region above the panels, where two recirculation zones develop.

The structure of these upper recirculations depends on the angle of incidence of the incoming solar radiation. During the morning hours, a small recirculation is confined to the upper left region, while a larger one occupies the central and right parts above the PVs. Around noon, the two recirculation become comparable in size, each covering approximately half of the upper region. In the afternoon, the left recirculation expands to cover most of the space above the panels, while the right recirculation is restricted to the upper right corner.
As a result, the air above the PV panels becomes trapped within these recirculation and does not mix efficiently with the cooler air below, leading to overheating during the day. Depending on the season and ambient temperature, this phenomenon may either benefit or hinder crop growth. However, it is unequivocally detrimental for PV operation, since elevated temperatures reduce photovoltaic efficiency.
This flow pattern appears to break down when the PV panels are inclined, as shown in Figure D.82 for 18/04/25, corresponding to a simulation with inclined PV panels and no plants.

Conclusions

The developed computational fluid dynamics (CFD) model was satisfactorily validated against field measurements for multiple seasons and for different greenhouse configurations, including empty chambers, chambers equipped with PV panels, and chambers with crops. The validation covered key physical processes such as radiation propagation and reflection, as well as plant evapotranspiration.
A notable capability of the model is its ability to simulate interference fringes during light propagation (electromagnetic wave behavior), a phenomenon that cannot be captured using simplified or linear radiation models.
A transfer function was successfully developed to map external climatic conditions onto the vertical surfaces of the greenhouse. This approach is particularly important for 3D simulations, where modelling the entire greenhouse would otherwise be computationally prohibitive. The transfer function enables accurate boundary condition definition while significantly reducing computational cost.
The simulations demonstrated that shading from neighboring chamber roofs is an important factor and must be accounted for:

- during morning and afternoon hours,
- during the winter period, and
- for the outer rows of plants,

as it can lead to a reduction in available PAR of up to 20%.
Conversely, in the central areas of the greenhouse, during midday hours in summer, the effect of neighboring shading can be safely neglected.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

The determination of optimal shading conditions depends on several parameters, including crop height, time of day, day of the year, and greenhouse orientation. During the summer months, PV panels—whether in horizontal or inclined positions—do not reduce the available solar radiation to prohibitive levels. In contrast, during the winter months, careful management of the PV tilt angle is crucial, as it can allow a significantly larger fraction of solar radiation to reach plant level.

Above the PV panels, the available radiation increases significantly due to reflections.

The thermal boundary layer associated with the elevated PV surface temperature is confined to a few centimeters around the panel surface and does not directly affect the rest of the greenhouse volume. However, the presence of the PV panels indirectly alters the thermal field in a decisive way.

Although horizontally mounted PV panels do not always strongly affect the available radiation—particularly during summer—they significantly modify the flow and thermal fields. By trapping air in the space between the panels and the roof, they inhibit downward convective heat transfer. This effect is undesirable during summer (when heat removal is needed) but may be beneficial during winter. At the same time, it leads to overheating above and around the PV panels, reducing their electrical performance. This adverse effect can be mitigated by adjusting the panel inclination.

Since changing the PV inclination influences both PV energy production and greenhouse energy demand (for heating or cooling), its impact on the flow field should be explicitly incorporated into the responsive control software that determines the optimal panel slope.

Overall, the developed computational model provides a comprehensive spatial description of:

- radiation distribution at multiple wavelength bands,
- temperature fields, and
- air velocity fields

throughout the greenhouse volume, rather than only at discrete sensor locations. This enables both qualitative insight and quantitative assessment of greenhouse microclimate behavior.

As a numerical experiment platform, the model substantially reduces the time and cost of parametric studies, supports informed decisions regarding the design and spatial arrangement of PV systems, and offers valuable guidance for the development of responsive control strategies.

Finally, since the model has been successfully validated, it can be reliably used both for extensive parametric investigations and for the generation of high-quality datasets for training a digital twin system.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

References

J.A.Duffie and W.A. Beckman, Solar Engineering of Thermal Processes, 3rd Edition, Wiley, New Jersey, 2006

Baxevanou C, Fidaros D., Bartzanas Th. And C.Kittas, 'Numerical simulation of solar radiation, air flow and temperature distribution in a naturally ventilated tunnel greenhouse', CIGR Journal, Volume 12, No 3, Pages 48-67, 2010

W.J.Parton, J.A. Logan, 'A model for diurnal variation in soil and air temperature', Agricultural Meteorology, 23(0) (1981) 205-216

Hubbell J.H., Seltzer S.M. 'Tables of x-Ray mass attenuation coefficients and mass energy absorption coefficients' National Institute of Standards and Technology (NIST)

Rita Mehra, 'Application of refractive index mixing rules in binary systems of hexadecane and heptadecane with n-alkanols at different temperatures'

R.Guzzi and R.Rizzi, 'Water vapor absorption in the visible and near infrared results of field measurment'

F. Villarreal-Guerrero, M. Kacira, E. Fitz-Rodríguez, C. Kubota, G.A. Giacomelli, R. Linkec, A. Arbel, 'Comparison of three evapotranspiration models for a greenhouse cooling strategy with natural ventilation and variable high pressure fogging', Scientia Horticulturae 134 (2012) 210–221

Thierry Boulard, Jean-Claude Roy, Jean-Baptiste Pouillard, Hicham Fatnassi, Ariane Grisey, Modelling of micrometeorology, canopy transpiration and photosynthesis in a closed greenhouse using computational fluid dynamics, Biosystems engineering 158 (2017) 110 – 133.

Hicham Fatnassi, Pierre Emmanuel Bournet, Thierry Boulard, Jean Claude Roy, Francisco D. Molina-Aiz f, Rashyd Zaaboul, Use of computational fluid dynamic tools to model the coupling of plant canopy activity and climate in greenhouses and closed plant growth systems: A review, Biosystems engineering 230 (2023) 388 – 408.

Katsoulas N., Stanghellini C., Modelling Crop Transpiration in Greenhouses: Dierent Models for Dierent Applications, Agronomy 2019, 9, 392; doi:10.3390/agronomy9070392

Fidaros D.K., Baxevanou C.A, Bartzanas T. and C.Kittas, 'Numerical simulation of thermal behavior of a ventilated arc greenhouse during a solar day', Renewable Energy Volume: 35 Issue: 7 Pages: 1380-1386, 2010

C. Baxevanou, D. Fidaros, Th. Bartzanas, C. Kittas, 'Yearly numerical evaluation of greenhouse cover materials', Computers and Electronics in Agricultrure, Vol. 149, pp.54-70, 2018

C. Baxevanou, D. Fidaros, N. Katsoulas, E. Mekeridis, Chr. Varlamis, A. Zachariadis and St. Logothetidis, Simulation of Radiation and Crop Activity in a Greenhouse Covered with Semitransparent Organic Photovoltaics, Appl. Sci. 2020, 10, 2550

Molina-Aiz FD, Valera DL, Alvarez AJ, Maduen A. A wind tunnel study of airflow through horticultural crops: determination of the drag coefficient. Biosystems Engineering 2006;93:447–57

D. Fidaros, C. Baxevanou, Th. Bartzanas, C. Kittas, 'Numerical study of mechanically ventilated broiler house equiped with evaporative pads', Computers and Electronics in Agricultrure, Vol. 149, pp.101-109, 2018

Franco, A., Valera, D.L., Pena, A., Perez, A.M., 2011.Aerodynamic analysis and CFD simulation of several evaporative cooling pads used in Mediterranean greenhouses. Computers and Electronics in Agriculture. 76, 218-230

Carmago, J.R., Ebinuma, C.D., Cardoso, S., 2003. A mathematical model for direct evaporative cooling air conditioning system. Engenharia Termica. 4, 30-34.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Kloppers, J.C., Kröger, D.G., 2005. The Lewis factor and its influence on the performance prediction of wet-cooling towers. International Journal of Thermal Sciences. 44, 879–884.

Kröger, D.G., 2004. Air-cooled heat exchangers and cooling towers, thermal –flow performance evaluation and design. Volume I, PennWell Corporation, Tulsa, Oklahoma

Cengel Heat Transfer

Jiayu Zhang , Elias Kaiser , Hanyi Zhang , Leo F. M. Marcelis and Silvere Vialet-Chabrand, A simple new method to determine leaf specific heat capacity, Zhang et al. Plant Methods (2025) 21:6, https://doi.org/10.1186/s13007-025-01326-3

S. J. Mousavizadeh1, K. Mashayekhi1, A. Daraei Garmakhany2, A. Ehteshamnia1 and S. M. Jafar, Evaluation of Some Physical Properties of Cucumber, Aug. 2010, Volume 4, No.4 (Serial No.29)

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Chapter 4. Modelling of the PV modules performance

Introduction

This chapter addresses the modelling of the PV modules performance, which focuses on developing a reliable and physically consistent representation of PV modules behavior under real operating conditions inside greenhouse-integrated agrivoltaic systems. Unlike the outdoor installations, PV modules inside greenhouses operate under partial sky obstruction, as structural elements and covering materials limit the direct view of the sky dome. Moreover, solar radiation experiences attenuation while passing through the greenhouse cover due to reflection, absorption, and scattering processes. These effects, together with modified irradiance spectra, non-uniform spatial irradiance, dynamic structural shading, and enhanced rear-side contributions due to internal reflections, create operating conditions that deviate significantly from PV systems installed outdoors (Zainali et al., 2025). Consequently, dedicated PV performance modelling approaches are required to accurately reproduce the real irradiance–temperature–power relationship of PV modules in greenhouse environments.

From the data collected and presented is evident that there might be some interactions inside the greenhouse that might affect PV performance. In particular, measured irradiance attenuation through the greenhouse envelope, time-dependent shadow patterns from structural elements, and reflective interactions within the enclosure indicate that commonly-used PV models may not fully capture the real power response of bifacial PV (BFPV) modules installed in greenhouse conditions. This motivates the development and validation of a customized modelling framework specifically adapted to these environments.

Within this context, this chapter aims to answer the question that how the greenhouse interacts BPFV system. Another question which this chapter is going to address is whether PVlib could provide accurate estimation of BFPV performance inside the greenhouse. In addition, this chapter investigates to what extent the modelling accuracy improves when adopting the Dynamic Bifacial Power Gain (DBPG) approach (Sohani et al., 2025a). The DBPG approach is a physical data-driven modelling framework designed to capture the dynamic and time-varying operating conditions of BFPV systems under real operational constraints.

A structure with three sections is considered for this chapter. Section 4.1 investigates irradiance attenuation through the greenhouse cover. This analysis is essential to quantify how much the incoming solar radiation is reduced while passing through the greenhouse envelope and to determine whether this attenuation remains constant under different outdoor irradiance conditions. Understanding the indoor irradiance level and its relationship with outdoor radiation is crucial, since it defines the effective radiative input to the PV modules and crops, and directly governs their operating conditions. Therefore, this analysis provides the fundamental basis for interpreting PV performance inside the greenhouse and for establishing realistic irradiance inputs for subsequent performance evaluation and modelling.

Section 4.2 quantifies the operational behavior of the greenhouse-integrated PV systems using the Performance Ratio (PR) and compares the results with an outdoor benchmark obtained from the reference installation at the ESTER Laboratory of the University of Rome Tor Vergata. This comparison reveals the impact of greenhouse-

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

specific conditions on PV performance. Since PR is a key indicator widely used to assess system efficiency, energy yield, and overall operational quality, an accurate determination of PR is essential for reliable energy assessment as well as for subsequent economic and environmental evaluations of greenhouse-integrated PV systems.
Section 4.3 evaluates PVlib as a widely used model for BFPV performance estimation and examines its applicability under greenhouse conditions. The DBPG model is then applied to estimate the power production, and its accuracy is assessed through comprehensive error analysis to determine the reliability of these two approaches for PV power modelling in greenhouse-integrated systems. Accurate power estimation is of fundamental importance, as predicted PV energy yield directly affects techno-economic performance indicators, environmental impact assessments, and system-level optimization studies. Therefore, evaluating modelling accuracy under real greenhouse operating conditions is a critical step toward trustworthy performance prediction.

4.1 Solar irradiance inside the greenhouse

When solar radiation passes through the greenhouse cover, the irradiance coming from the outside is inevitably reduced due to reflection, absorption, and scattering mechanisms within the covering material. Quantifying this attenuation is essential because the transmitted irradiance directly governs the microclimatic conditions inside the greenhouse and the energy available for PV systems and agricultural processes. Therefore, analyzing indoor–outdoor irradiance relationships, temporal profiles, and cumulative metrics provides critical insight into both the magnitude and consistency of irradiance losses.

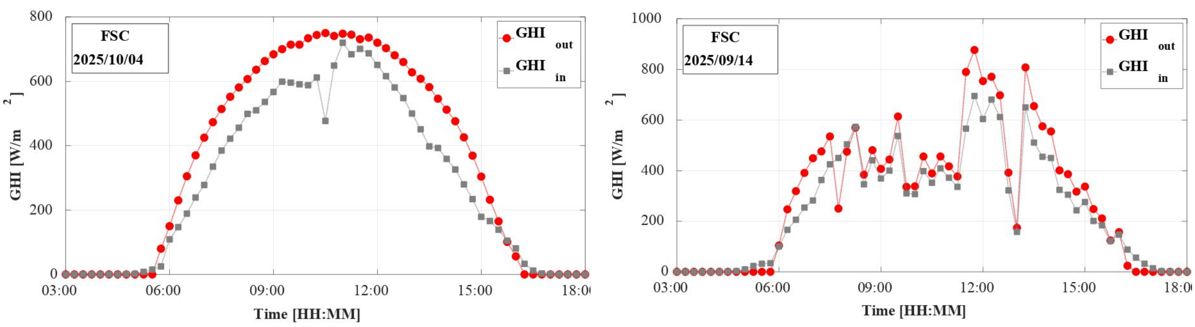
4.1.1 GHI_{in} versus GHI_{out}

Fig. 4.1a presents the profiles of global horizontal irradiance (GHI) outside and inside the greenhouse, denoted as GHI_{out} and GHI_{in}, respectively, for a sample clear-sky day and Fattoria Solidale del Circeo (FSC). As shown in Fig. 4.1a, from early morning onward, the difference between the GHI outside and inside the greenhouse gradually increases as solar irradiance rises. This growing gap between GHI_{out} and GHI_{in} indicates an increasing level of radiation attenuation caused by the greenhouse envelope. Around the time of maximum irradiance, a distinct drop is observed in the GHI_{in} profile, while GHI_{out} remains smooth. This drop is most likely attributed to a temporary shading event, caused by greenhouse structural elements casting a shadow on the pyranometer measuring GHI_{in}.
Following this drop, GHI_{in} increases again and approaches the magnitude of GHI_{out} more closely than during the morning hours. Compared to the morning behavior, this suggests an additional contribution from reflected radiation, likely originating from surrounding objects such as PV panels, since the indoor pyranometer is installed above the panels. This effect explains the observed asymmetry in the GHI_{in} profile, whereas GHI_{out} exhibits a nearly symmetric shape around solar noon. The observed asymmetry in GHI_{in} highlights the role of internal greenhouse geometry and reflective surfaces, which can locally modify the radiative environment beyond pure transmission losses.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

As the system moves away from peak irradiance in the afternoon, the contribution of reflected irradiance gradually diminishes, and the GHI_{in} profile progressively resembles its morning behavior, eventually decreasing toward zero together with GHI_{out}. This behavior suggests that midday indoor irradiance cannot be interpreted solely as a function of cover transmittance but also depends on secondary optical effects such as reflections and temporary shading. Such effects are particularly relevant for agrivoltaic greenhouses, where structures interact with the radiation field inside the enclosure.

The profiles of GHI_{in} and GHI_{out} are compared for a sample overcast day in Fig. 4.1b. In contrast to the clear-sky conditions, both irradiance profiles exhibit pronounced short-term fluctuations throughout the day. Under these conditions, a major part of the irradiance is diffuse and the difference between GHI_{out} and GHI_{in} remains relatively more uniform over time, indicating that radiation attenuation inside the greenhouse is governed primarily by the cover transmittance rather than by solar geometry. The intermittent peaks and sharp drops observed simultaneously in both GHI_{in} and GHI_{out} profiles are associated with transient changes in the irradiance level (Due to the issues like cloud passages), which leads to the change in the diffuse and global irradiance in a similar manner inside and outside the greenhouse. This behavior suggests that, during overcast days, the greenhouse radiative environment is dominated by diffuse radiation, leading to a more proportional and stable relationship between GHI_{in} and GHI_{out}, despite the overall lower irradiance levels.



(a) **(b)**

Fig. 4.1. The profiles of global horizontal irradiance outside and inside the greenhouse (GHI_{out} and GHI_{in}) for (a) a clear-sky day and (b) an overcast day for FSC.

Figure 4.2 illustrates the relationship between GHI_{in} and GHI_{out} for the FSC site. In the plotted scatter plot, both values are normalized by dividing by 1000 W/m^2, which is for the Standard Test Condition (STC). A strong linear correlation is observed between the two variables, indicating that the indoor irradiance closely follows the outdoor irradiance under a wide range of operating conditions. This confirms that the greenhouse cover acts primarily as a transmissive medium, affecting the incoming solar radiation in a consistent manner.

The linear regression fitted to the data yields a slope of 0.8566, which quantitatively represents the average transmission factor of the greenhouse envelope. This implies that, on average, approximately 85.8% of the outdoor irradiance is transmitted inside the greenhouse, corresponding to a 14.2% reduction in irradiance due to the combined

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

effects of the cover material, structural elements, and optical losses. This level of attenuation is significant and highlights the non-negligible impact of the greenhouse envelope on the available solar resource for both crop growth and agrivoltaic applications.
Despite the strong overall correlation, there are some points that are significantly far from the fitted line. Data points located above the fitted line, where GHI_{in} is relatively high compared to GHI_{out}, can be associated with conditions characterized by enhanced internal reflections. These reflections may originate from PV module surfaces, crops, leaves, metallic structures, or other reflective elements inside the greenhouse, which locally increase the irradiance measured by the indoor pyranometer. This effect is particularly relevant given that the sensor is installed above the PV panels, making it sensitive to reflected radiation components.
Conversely, the data points below the regression line, corresponding to high outdoor irradiance and comparatively low indoor irradiance. Such a condition can come from greenhouse structural components, support beams, or temporary blocking of reflected radiation paths, leading to short-term reductions in the irradiance reaching the indoor sensor.
Overall, the scatter plot demonstrates that while the greenhouse cover imposes a clear and measurable attenuation on solar irradiance, secondary optical effects such as reflections, shading, and geometric interactions play a relevant role in shaping the instantaneous indoor radiative environment. These effects should therefore be taken into account when modeling solar availability and energy performance in agrivoltaic greenhouse systems.

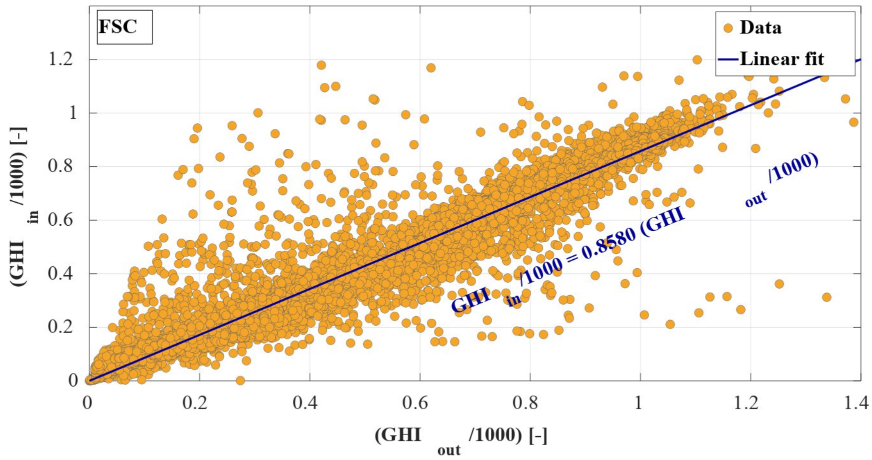


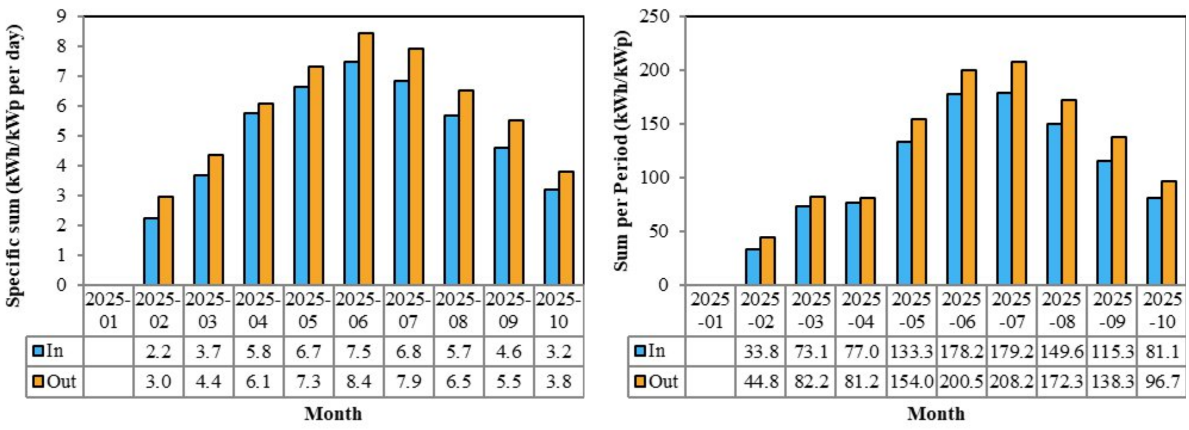
Fig. 4.2 Scatter plot of $\boldsymbol{GHI_{in}}$ versus $\boldsymbol{GHI_{out}}$ at the FSC site, with the fitted linear regression line.

Figure 4.3a and Figure 4.3b compare the monthly received solar energy on the horizontal plane measured outside and inside the greenhouse, normalized by the bifacial standard test condition (BSTC), corresponding to a front irradiance of 1000 W/m² and a rear irradiance of 135 W/m². This normalization allows a consistent comparison of seasonal irradiance levels independently of absolute radiation intensity.
Figure 4.3a presents the raw monthly sums, showing a clear seasonal trend with increasing the received energy from winter to summer, followed by a gradual decline toward autumn. The normalized outdoor irradiance rises from values below 50

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

kWh/kW$_p$ in winter months (e.g., about 45 kWh/kW$_p$ in February) to peak levels of approximately 200–210 kWh/kW$_p$ in June and July, while the corresponding indoor irradiance increases from around 35 kWh/kW$_p$ to about 175–180 kWh/kW$_p$ over the same period. In all months, the normalized outdoor irradiance remains systematically higher than the indoor one, reflecting the attenuation imposed by the greenhouse cover, as previously discussed for instantaneous GHI profiles and the GHI_{in}–GHI_{out} correlation. The largest absolute differences are observed during late spring and summer months, when solar irradiance levels are the highest and transmission losses accumulate over longer daylight periods.

In order to better account for differences in data availability and effective irradiance duration, Figure 4.3b shows the specific monthly values, obtained by dividing the monthly sums by the number of effective days with irradiance (Number of effective days with irradiance is obtained from dividing the number of total hours with available irradiance data by 24). This representation provides a more representative comparison of the average daily radiative conditions within each month. In particular, this normalization is especially relevant for months such as May, where data filtering and consequently, partial month availability could otherwise bias the interpretation with the raw data.



(a) **(b)**

Fig. 4. 3 Comparing the monthly profiles of the received energy on the horizontal surface inside and outside the greenhouse of the FSC site (a) Raw values; (b) The specific values, which obtains by dividing the total monthly values by the equivalent days with available irradiance data.

The specific values introduced in Figure 4.3b highlight a smoother and more physically meaningful seasonal evolution. Both indoor and outdoor irradiance increase steadily from winter toward early summer, reaching their maximum around June, with specific normalized values of approximately 7.5 kWh/kW$_p$ per day indoors and 8.4 kWh/kW$_p$ per day outdoors, and then, decrease toward autumn, falling to about 3.2 kWh/kW$_p$ per day and 3.8 kWh/kW$_p$ per day, respectively, in October. The relative difference between GHI_{out} and GHI_{in} remains fairly consistent throughout the year, typically on the order of 10–15%, which is fully consistent with the attenuation level derived from the linear regression between GHI_{in} and GHI_{out}. This agreement confirms that the average transmission behavior of the greenhouse cover is stable across seasons. However, a slightly larger relative gap is observed during high-irradiance months, which can be attributed to the increased influence of structural shading, angle-of-

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

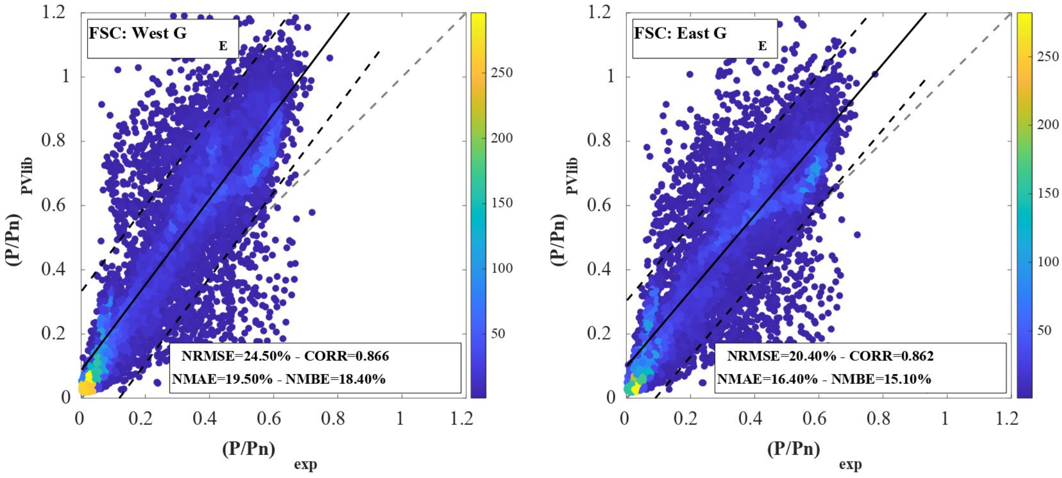
Funded by the European Union

incidence effects, and internal optical interactions (like reflections from the objects) mostly under clear-sky conditions, as previously identified.
Overall, the comparison between raw and specific monthly values confirms that while absolute radiative energy strongly depends on season and daylight duration, the normalized indoor–outdoor relationship remains stable. This indicates that the greenhouse envelope causes an almost constant reduction in solar irradiance throughout the year. In other words, the average transmission of the cover does not change significantly from season to season. Secondary effects, such as reflections and temporary shading, mainly influence short-term variations and periods with high irradiance. However, these effects do not significantly affect the overall seasonal trend of the irradiance inside the greenhouse.

4.1.2 Selection of the representative effective irradiance

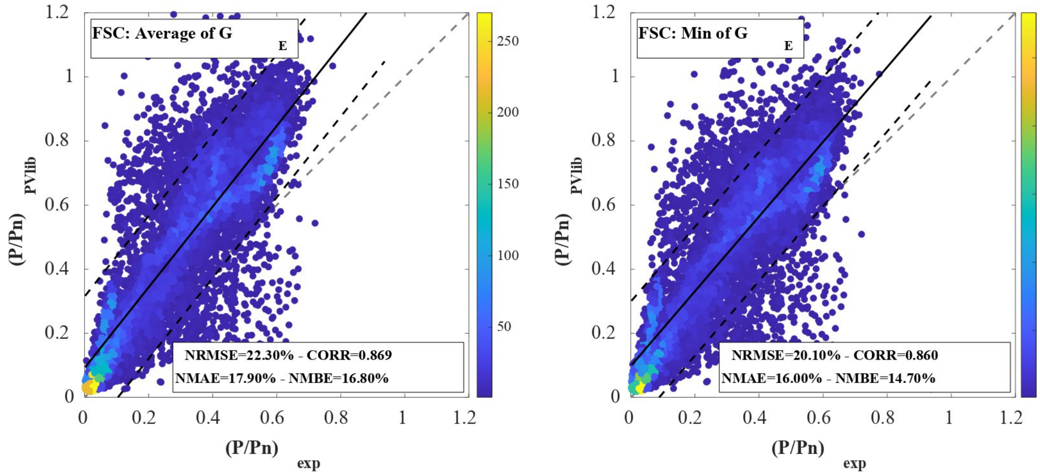
The PV system of FSC is electrically integrated; that is, the East- and West-side PV sub-arrays are electrically interconnected, and the power reported by the inverter corresponds to the sum of the power produced by both sides. In contrast, irradiance measurements are available separately for the East and West greenhouses. For several analyses—including the calculation of PR and developing the DBPG model—it is therefore necessary to define a single value as a representative effective irradiance (G_E) for the whole FSC system.
With this objective, four alternative definitions were examined:

- West G_E
- East G_E
- Average of G_E (West and East)
- Minimum of G_E



(a) (b)

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union



(c) **(d)**

Fig. 4.4 Comparison of the observed and PVlib estimated normalized power for the FSC site using the four representative effective irradiance definitions (a) West $\boldsymbol{G_E}$; (b) East $\boldsymbol{G_E}$; (c) Average $\boldsymbol{G_E}$,and (d) minimum $\boldsymbol{G_E}$.

For each case, the system power was computed using PVlib, taking the corresponding G_E as the input and using the measured back-of-module temperature (T_{bom}). The accuracy of PVlib under these four definitions is reported in Figure 4.4a to Figure 4.4d. As can be observed, the calculated power shows systematic overestimation in all the conditions, primarily due to structural shading effects (and other potential factors such as spectral changes) that are not captured in PVlib. This is reflected by the consistently positive bias values and the relatively large errors. Among the four options, using West G_E leads to the highest error (Figure 4.4a; NRMSE = 24.50%), whereas East G_E performs much better (Figure 4.4b; NRMSE = 20.40%). As expected, using the average G_E results in an intermediate accuracy (Figure 4.4c; NRMSE = 22.30%). Nevertheless, the lowest error is obtained when using the minimum $\boldsymbol{G_E}$ (Figure 4.4d: NRMSE = 20.10%), which is very close to the East-side result. This indicates that, in most instances, the minimum G_E corresponds to the East greenhouse—likely due to differences in reflections, and the shading patterns induced by surrounding objects and structures. This finding is also physically consistent: in a series-constrained PV system (as in FSC), the overall performance is effectively limited by the lowest irradiance condition, making the minimum irradiance a more appropriate representative input for system-level power estimation.

4.2 Evaluation of the PV system performance

The performance of the PV systems installed in the FSC and UTH greenhouses was evaluated using the PR. PR is a widely adopted metric for assessing PV system quality. It is a dimensionless ratio between the final yield (Y_f) and the reference yield (Y_r). Y_f represents the electrical energy delivered by the PV system, normalized by the installed nominal power of the PV array. It reflects the actual usable energy output of the system over a defined period. Moreover, Y_r shows the available solar energy resource at the site, also normalized by the reference irradiance level. Here, Y_r is calculated using the measured plane-of-array irradiance and is referenced to the BSTC irradiance level.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

Table 4.1 The performance evaluation of the PV system in FSC site

Year	Month	Y_f (kWh/kWh$_p$)	$Y_{r,front}$ (kWh/kWh$_p$)	$Y_{r,rear}$ (kWh/kWh$_p$)	Y_r (kWh/kWh$_p$)	PR
2025	3	63.53	101.49	13.60	115.09	0.5520
2025	4	89.05	135.47	25.78	161.25	0.5523
2025	5	116.83	185.11	28.28	213.39	0.5475
2025	6	129.57	216.22	26.37	242.59	0.5341
2025	7	129.16	209.24	28.90	238.15	0.5424
2025	8	108.02	172.82	24.80	197.62	0.5466
2025	9	79.31	132.83	17.57	150.40	0.5273
2025	10	52.40	88.61	10.38	98.99	0.5293
The whole period		**790.32**	**1279.57**	**180.26**	**1459.83**	**0.5414**

The available data was utilized to calculate the monthly PR values for both greenhouses from March to October. The results are reported in Table 4.1 and Table 4.2 for the FSC and UTH sites, respectively. Furthermore, PR values obtained from the tests conducted at the ESTER Laboratory of University of Rome Tor Vergata (ESTER Lab) are presented in Table 4.3 as the reference benchmark of the same PV performance outside the greenhouse. The ESTER Lab measurements were performed during January and February on PV modules identical to those installed in the FSC and UTH greenhouses, providing a reliable benchmark under well-monitored reference conditions.

Table 4. 2 The performance evaluation of the PV system in UTH site

Year	Month	Y_f (kWh/kWh$_p$)	$Y_{r,front}$ (kWh/kWh$_p$)	$Y_{r,rear}$ (kWh/kWh$_p$)	Y_r (kWh/kWh$_p$)	PR
2025	3	62.15	91.96	21.04	113.01	0.5499
2025	4	102.26	148.79	37.17	185.96	0.5499
2025	5	125.03	184.50	49.47	233.98	0.5344
2025	6	134.10	208.48	52.26	260.74	0.5143
2025	7	133.85	215.53	51.55	267.08	0.5012
2025	8	122.51	194.21	53.51	247.73	0.4945
2025	9	81.12	128.46	30.19	158.64	0.5114
2025	10	50.65	73.52	17.72	91.24	0.5551
The whole period		**811.67**	**1245.45**	**312.92**	**1558.37**	**0.5208**

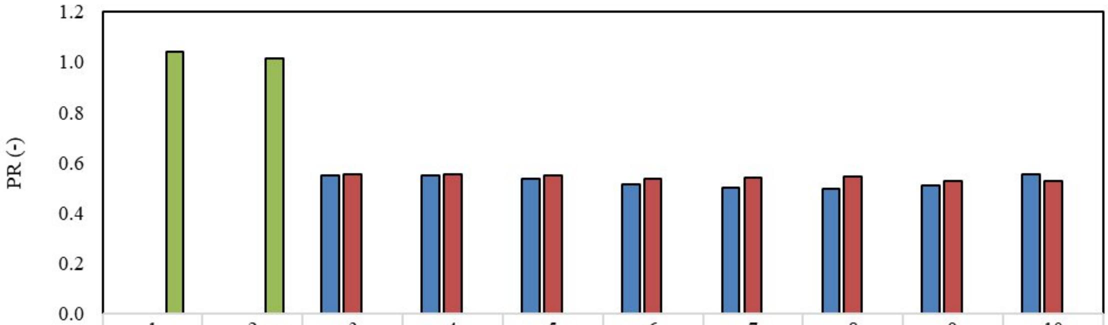
In addition to tabular presentation, Fig. 4.5 is also provided. This figure offers a visual comparison of the monthly PR values for the FSC, UTH, and ESTER Lab cases, enabling a clearer interpretation of the relative performance of the systems.

Table 4.3. The performance evaluation of the PV system in ESTER Lab.

Year	Month	Y_f (kWh/kWh$_p$)	$Y_{r,front}$ (kWh/kWh$_p$)	$Y_{r,rear}$ (kWh/kWh$_p$)	Y_r (kWh/kWh$_p$)	PR
2025	1	11.71	9.82	1.44	11.25	1.0403
2025	2	86.43	74.34	10.88	85.22	1.0141
The whole period		**98.14**	**84.16**	**12.32**	**96.48**	**1.0172**

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union



Month	1	2	3	4	5	6	7	8	9	10
UTH			0.5499	0.5499	0.5344	0.5143	0.5012	0.4945	0.5114	0.5551
FSC			0.5520	0.5523	0.5475	0.5341	0.5424	0.5466	0.5273	0.5293
ESTER	1.0403	1.0141								

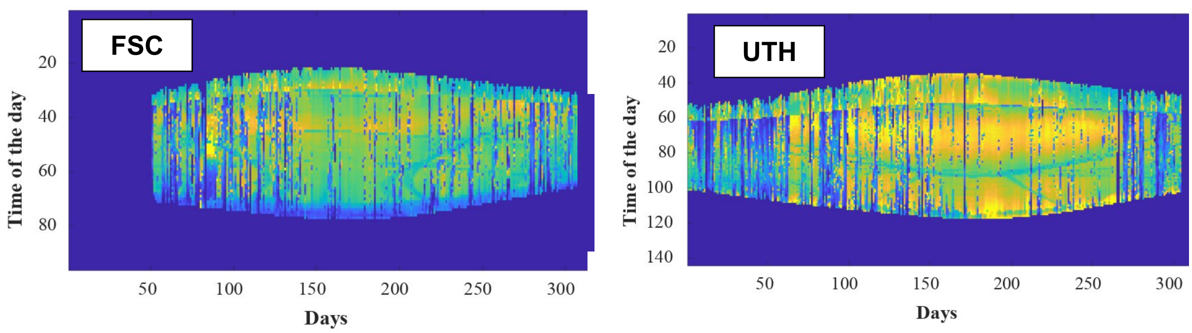
Month

Fig. 4. 5 Comparing the monthly values of PR for FSC, UTH, and ESTER Lab.

According to Table 4.1 and Table 4.2, PR values for FSC greenhouse vary between 0.5273 and 0.5523, while for the UTH greenhouse they range between 0.4945 and 0.5551. These results indicate a lower PV system performance under greenhouse-integrated operating conditions compared to working outside. Based on Table 4.3, ESTER Lab measurements show significantly high PR values, lying in the range of 1.0141 to 1.0403. The high PR values observed in ESTER Lab confirm that the PV modules operate properly and that no intrinsic performance degradation or structural defects exist at the module level.

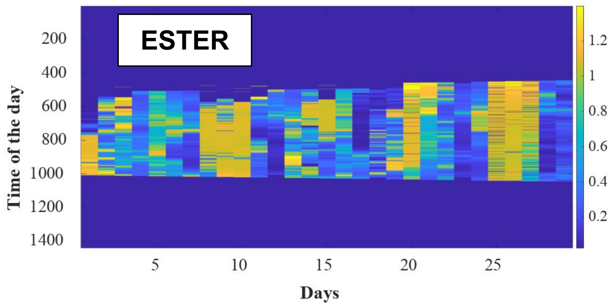
During the day, the shadow of the structure falls on the PV modules, which leads to reduction in the produced power of them. The contour plots of the clearness index shown in Fig. 4.6a and Fig. 4.6b, generated using irradiance data measured at a reference point (Global tilted irradiance (GTI) pyranometer), clearly confirm the presence of such shading. The gray arc-shaped regions embedded within the brighter irradiance distribution indicate time intervals where shading occurs at the measurement point. On the other hand, such shadowing is not seen for ESTER Lab, as Fig. 4.6c shows

The computed PR values for FSC and UTH also indicate that there are relatively small monthly fluctuations of PR compared to conventional outdoor PV installations (Fig. 4.5). It can be attributed to the more stable thermal environment inside the greenhouse. The reduced temperature variations lead to lower temperature effect, and consequently, steadier module operating conditions, thereby moderating seasonal performance changes.



by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

(a) (b)



(c)

Fig. 4.6 The contour plots of the clearness index, as the ratio between GTI and clear-sky GTI, for different cases; (a) FSC; (b) UTH; (c) ESTER Lab.

A number of experiments were conducted at ESTER Lab to confirm the influence of shadowing on PV power production. For this purpose, four conditions were evaluated (Figure 4.7):

1. Single vertical shadow (Figure 4.7a),
2. Single horizontal shadow (Figure 4.7b),
3. Single diagonal shadow (Figure 4.7c)
4. Double shadowing (Figure 4.7d).

Figure 4.8 presents the results. Experiments started by single vertical shadow. For this experiment, the shadow mask was put around 10. As shown in Figure 4.8a, before putting the shadow mask, both normalized profiles was following the same trend. However, by applying the shadow mask, the normalized power suddenly drops and is reduced by almost the half. The same profiles as the single vertical shadow are observed for horizontal and diagonal shadow experiments days (Figure 4.8b and Figure 4.8c).

(a) (b)

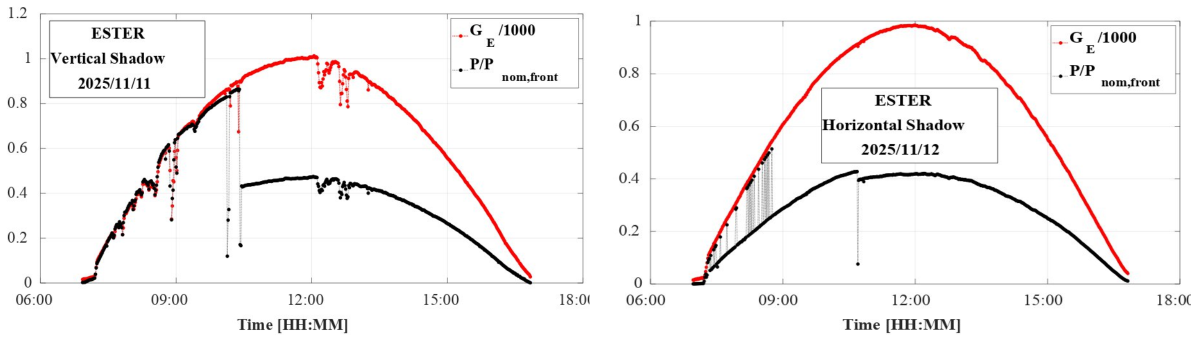
by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

(c) (d)

Fig. 4.7 The conducted experiments at ESTER Lab to confirm the influence of shadowing on PV power production; (a) Single vertical shadow; (b) Single horizontal shadow; (c) Single diagonal shadow; (d) Double shadowing.

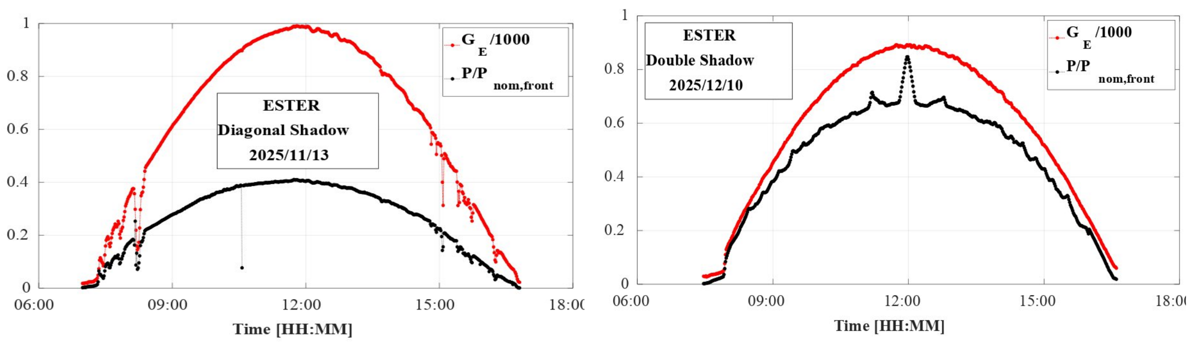
For all the three single-shadow experiments (vertical, horizontal, and diagonal), the shadow mask was positioned very close to the module surface. Having reached the initial confirmation of the shading effect on power production, the shadow source mask was placed farther from the module in the double-shadow experiment to better approximate real operating conditions. Accordingly, in contrast to the single-shadow case where nearly no light reached the region under the mask, in the double-shadow configuration a portion of the irradiance can still reach the shaded cell surfaces.
The results of the double shadowing experiment are presented in Figure 4.8d. As shown, a significant reduction in the normalized power is again observed. Around the peak generation period, the normalized power drops to approximately 60%, providing a final confirmation of the negative impact of structural shading on the PV system performance inside the greenhouse. Around noon, a sudden increase is observed due to the sun reaching a position where the shading structure casts much smaller shadow on PV modules than other times of the day. In real greenhouse installations, shading is typically caused by multiple structural columns, much more complicated than the tested double-shadow configuration, which can lead to even greater power losses. This observation is consistent with the discussion presented in the PR analysis section.



(a) (b)

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union



(c) (d)

Fig. 4.8 The profiles for the normalized power and effective irradiance for the shadow experiments at ESTER Lab; (a) Single vertical shadow; (b) Single horizontal shadow; (c) Single diagonal shadow; (d) Double shadowing.

4.3 PV production modelling

In the previous project deliverable (Deliverable 4.1), the modeling framework, underlying assumptions, and mathematical formulations for bifacial PV (BFPV) performance evaluation were comprehensively introduced and documented. In that report, PVlib (Holmgren et al., 2018) was introduced as a widely adopted tool in industry practice for BFPV power output. Subsequently, in view of the limitations of PVlib, the Dynamic Bifacial Power Gain (DBPG) models were proposed to achieve more accurate and physically consistent power estimations. The DBPG approach is a data-driven physical modeling framework, developed in two versions: the first DBPG model (DBPG1) (Sohani et al., 2025b), which requires rear tilted irradiance (RTI) as an input and evaluated in Deliverable 4.1 for the utility scale BFPV. The second DBPG model (DBPG2), which has been proposed after the DBPG1 model (Sohani et al., 2025a) and eliminates the need for RTI or ground albedo data. The previous deliverable therefore established the theoretical background, mathematical formulation, and preliminary validation of PVlib and DBPG-based modeling frameworks (The DBPG1 model).
In this part of the present deliverable, both PVlib and the DBPG models are briefly recalled, and their application is extended to PV systems in FSC and UTH greenhouses. Among two versions of the DBPG2 models, the DBPG2 is utilized here, as it has been shown to have higher accuracy for the cases with the number of PV and albedo conditions like the investigated greenhouse cases (Sohani et al., 2025a).

4.3.1. A brief description of modelling framework

In PVlib, BFPV power is computed using the same model as monofacial PV (MFPV) driven by an effective irradiance that accounts for rear-side contribution, expressed as:

$$^{F,PVlib}_{t} = P_{est}^{MF,PVlib}(G_E,T_{amb}) \quad (4.1)$$

T_{amb} is the ambient temperature. G_E also denotes the effective irradiance. It is determined by:

$$= GTI + \varphi\ RTI \quad (4.2)$$

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

In Eq. (4.2) GTI and RTI are the global tilted irradiance and rear tilted irradiance. Moreover, φ represents the bifaciality coefficient, which is the ratio between Standard Test Condition (STC) power values of rear and front sides (Stein et al., 2021).
As introduced in details in Deliverable 4.1, in the DBPG2 model, BFPV power is estimated as the summation of two terms:

$$F = P^{MF} + P_n^{BF,front}\ DBPG \qquad (4.3)$$

The first term is the power output for an MFPV with the same specifications if it is exposed to the same GTI and T_{amb} as the investigated BFPV. The second term is also dynamic bifacial power gain ($DBPG$), which describes how much power is different than an MFPV with the identical condition as BFPV. $DBPG$ is expressed in normalized form with respect to the BFPV front side nominal power. Since installed BFPV plants typically consist only of BFPV modules, the monofacial reference power under identical operating conditions is obtained through simulation. Here, PVlib approach is employed for this purpose.
In the DBPG2 modeling framework, $DBPG$ is estimated as a polynomial function. For fixed-tilt (FT) systems, DBPG is expressed as a function of the cosine of the angle of incidence ($\cos(AOI)$) and the clearness index (K_{CS}), while for single-axis tracking (SAT) systems, solar azimuth and elevation angles (AZ and EL) are additionally included to account for tracking-induced geometric effects (Eqs. (4.4) and (4.5), respectively).

$$\begin{aligned} & BPG_{est}^{DBPG2\ model}\Big)_{FT} \\ & _0 + a_1 \cos(AOI) + a_2 K_{CS} + a_{12} \cos(AOI) K_{CS} + a_{11} (\cos(AOI))^2 \\ & _{22} K_{CS}^2 \end{aligned} \qquad (4.4)$$

$$\begin{aligned} & BPG_{est}^{DBPG2\ model}\Big)_{SAT} \\ & a_0 + a_1 \cos(AOI) + a_2 K_{CS} + a_3\ AZ + a_4\ EL + a_{12} (AOI) K_{CS} \\ & _{13} \cos(AOI) AZ + a_{14} \cos(AOI) EL + a_{23} K_{CS} AZ + a_{24} K_{CS} EL \\ & _{34} AZ\ EL + a_{11} (cosd(AOI))^2 + a_{22} K_{CS}^2 + a_{33} AZ^2 + a_{44} EL^2 \end{aligned} \qquad (4.5)$$

The power of BFPV in both FSC and UTH greenhouses is estimated to use both the PVlib and DBPG2 models and the results are compared with ESTER Lab, a reference BFPV installation operating outside greenhouse. This comparison aims to assess how each modeling approach performs in greenhouse and non-greenhouse environments and to quantify the differences in their ability to capture bifacial power behavior under structural shading effects. The BFPV systems in FSC and UTH are SAT, while the one in ESTER Lab is FT.

4.3.2. Comparison of PVlib and DBPG2 models

Figure 4.9 presents the scatter plots of normalized estimated power versus normalized measured power for the PVlib and the DBPG2 models in the FSC and UTH greenhouse-integrated PV systems, as well as in the ESTER Lab open-field installation. The Normalized Root Mean Square Error (NRMSE), Normalized Mean Absolute Error (NMAE), and Normalized Mean Bias Error (NMBE) values are reported in each subfigure to quantitatively assess the estimation accuracy.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

In FSC, PVlib model shows a considerable dispersion around the one-to-one line (Figure 4.9a). NRMSE and NMBE values are 20.20% and 14.80%, respectively. These relatively high errors mainly originate from the fact that PVlib estimates BFPV power through the effective irradiance concept, without explicitly incorporating dynamic shading patterns and spatially non-uniform bifacial irradiance typical of greenhouse structures. As a result, a high overestimation is observed.

By contrast, the DBPG2 model significantly improves the estimation accuracy (Figure 4.9b). NRMSE decreases to 6.90%, corresponding to a significant relative reduction of approximately 66%. At the same time, NMBE is reduced from 14.80% to 0.18%, indicating an almost complete removal of the bias. This enhancement is achieved because the DBPG2 model works with the historical data and solar-geometry-related variables, enabling simultaneous consideration of bifacial gain and time-dependent shading effects inherent in greenhouse-integrated BFPV systems.

A similar behavior is obtained in the UTH greenhouse case. As shown in Figure 4.9c, using PVlib is accompanied by NRMSE and NMBE values of 19.10% and 14.80%, respectively. Again, the relatively large errors are attributed to the absence of an explicit treatment of dynamic shading and spatially varying bifacial irradiance in the PVlib framework.

Applying the DBPG2 model (Figure 4.9d) leads to a pronounced improvement. NRMSE is reduced to 5.86%, which represents a considerable relative enhancement of about 69%. Moreover, NMBE decreases to 0.18%, demonstrating that the systematic overestimation observed with PVlib is almost entirely eliminated. Similar to in FSC case, this improvement stems from the capability of the DBPG2 to capture the combined influence of structure dynamic shadowing and reflection variations, which are the dominant physical phenomena in greenhouse PV installations.

ESTER Lab case corresponds to an open-field BFPV system where structural shading and geometric obstruction effects are minimal. Under these conditions, the PVlib model already achieves a good agreement with measurements (Figure 4.9e), yielding NRMSE of 1.87% and NMBE of 0.84%. This confirms that in conventional unshaded configurations, the effective irradiance-based approach of PVlib provides reliable power estimation.

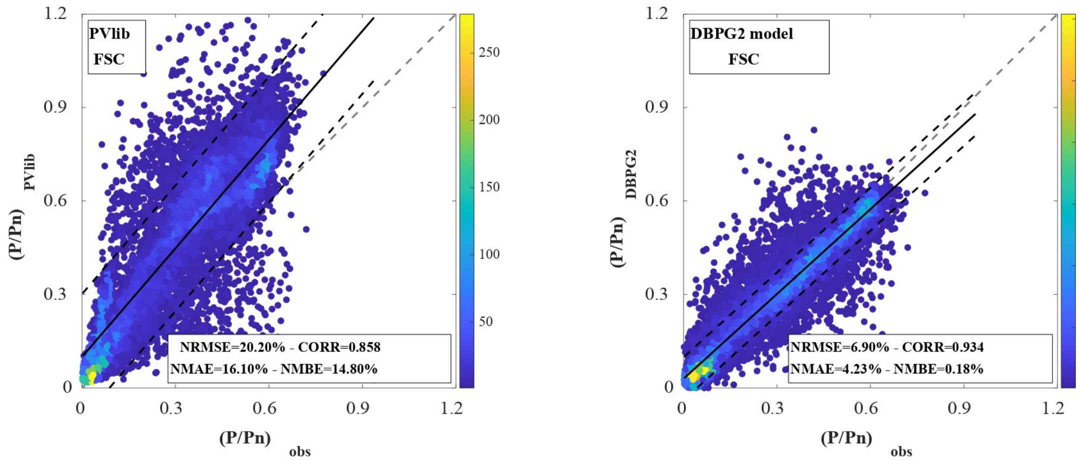
Nonetheless, the DBPG2 model enhances the estimation accuracy even in this case. According to Figure 4.9f, NRMSE decreases from 1.87% to 1.42%, corresponding to a relative improvement of about 24%. Additionally, NMBE is reduced from 0.84% to 0.01%, indicating an almost unbiased estimation. This result highlights that even when shading is not a dominant factor, the DBPG2 model offers a more precise estimation of bifacial irradiance contribution by embedding site-specific bifacial gain behavior into the model structure.

Across all investigated cases, the DBPG2 model consistently demonstrates superior estimation performance compared to PVlib, in terms of both NRMSE and NMBE. The improvement is especially significant in greenhouse-integrated BFPV systems, where dynamic shading and bifacial gain coexist and strongly influence power production. Consequently, the DBPG2 model is particularly suitable for PV performance estimation in agrivoltaic and greenhouse applications, where PVlib application imposes very high errors due to not capturing the complex effect of shadow and structure. Such high errors cause PVlib not to be a reliable model for BFPV estimation in these cases. On the other hand, as observed, the DBPG2 model utilization substantially reduces both random error and systematic bias, leading to more reliable power estimation outcomes.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Fig. 4.10 compares the normalized daily power output profiles between measured observations and simulations obtained from PVlib and the DBPG2 model for the FSC, UTH, and ESTER sites under representative operating days. Fig. 4.10.a and Fig. 4.10.b correspond to two selected days at the FSC site. In both cases, the DBPG2 model closely reproduces the measured power evolution from sunrise to sunset, whereas PVlib exhibits visible deviations during midday and afternoon hours, where rear-side irradiance contribution and spatial non-uniformity across the module rows become more pronounced.
Fig. 4.10.c and Fig. 4.10.d show the daily profiles for the UTH site. Once again, DBPG2 maintains strong agreement with experimental data throughout the daytime period and successfully captures short-term power fluctuations caused by variable sky conditions. In contrast, PVlib shows larger discrepancies during intervals of rapid irradiance variation, resulting in a less accurate reconstruction of the measured curve.
Fig. 4.10.e and Fig. 4.10.f present the ESTER site under two different daily conditions. The DBPG2 model consistently demonstrates a closer overlap with the observed power profiles, especially around peak production hours, while PVlib tends to misestimate the power magnitude when rear-side effects are dominant.
Overall, across all cases in Fig. 4.10.a–f, the DBPG2 model systematically provides a more faithful representation of the measured daily power profiles than PVlib. This outcome is fully consistent with the previously presented scatter-plot analysis, confirming the superior predictive accuracy and robustness of the DBPG2 model across different sites, sky conditions, and system configurations.
To enable a fair comparison of model performance across different locations, the assessment is conducted using PR rather than the absolute energy yield. Although the absolute energy production inherently varies with local climatic conditions, PR provides a normalized indicator that isolates the influence of system performance and modeling accuracy from site-dependent irradiance availability. Accordingly, the monthly PR profiles for the three investigated cases—FSC, UTH, and ESTER—are illustrated in Fig. 4.11, while the cumulative PR over the entire monitoring period is summarized in Table 4.4.



(a) (b)

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

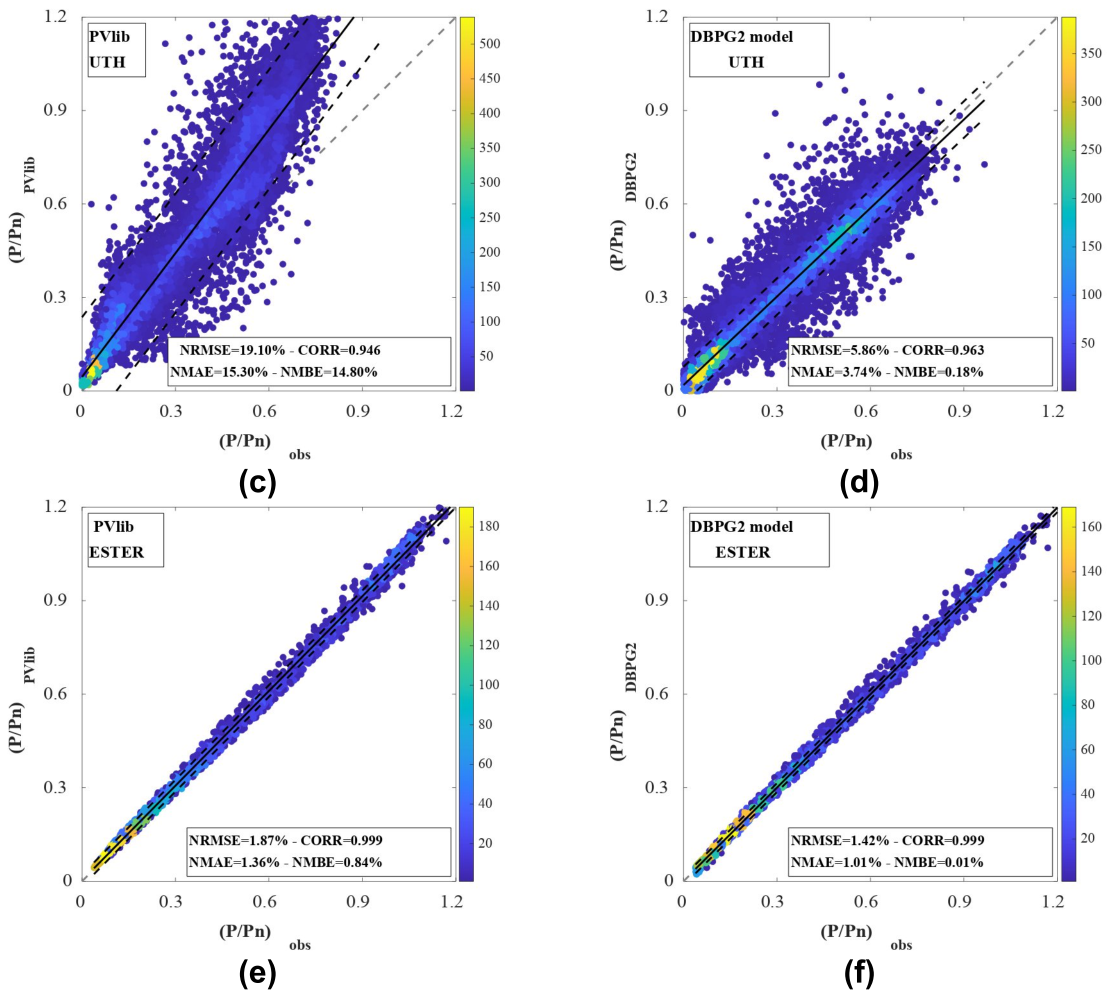


Fig. 4.9 Checking the power estimation accuracy of DBPG2 model compared to PVlib using the test data for (a and b) FSC; (c and d) UTH; (e and f) ESTER Lab.

As previously demonstrated in the assessment of the power estimation models (Fig. 4.9 and Fig. 4.10), PVlib exhibited a pronounced positive bias in estimating power output for the greenhouse-integrated systems (FSC and UTH), whereas the DBPG2 model substantially reduced this bias by incorporating historical data and dynamic bifacial gain behavior. In addition to bias (NMBE), the DBPG2 model was able to offer much better NRMSE. Since the reference yield in PR computation is derived from measured irradiance and remains identical for all models, any systematic overestimation in simulated power directly propagates into an inflated PR. Therefore, PR serves here as an effective indicator of cumulative modeling accuracy.

Based on Fig. 4.11a, PVlib consistently overestimates PR for the FSC greenhouse throughout the year. The observed monthly PR ranges between 0.5273 and 0.5520, while PVlib estimates values between 0.7698 and 0.8117, corresponding to a relative overestimation of approximately 42% to 54% (absolute deviations of +0.2274 to +0.2844 PR units, with the maximum error in September and the minimum error in July).

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

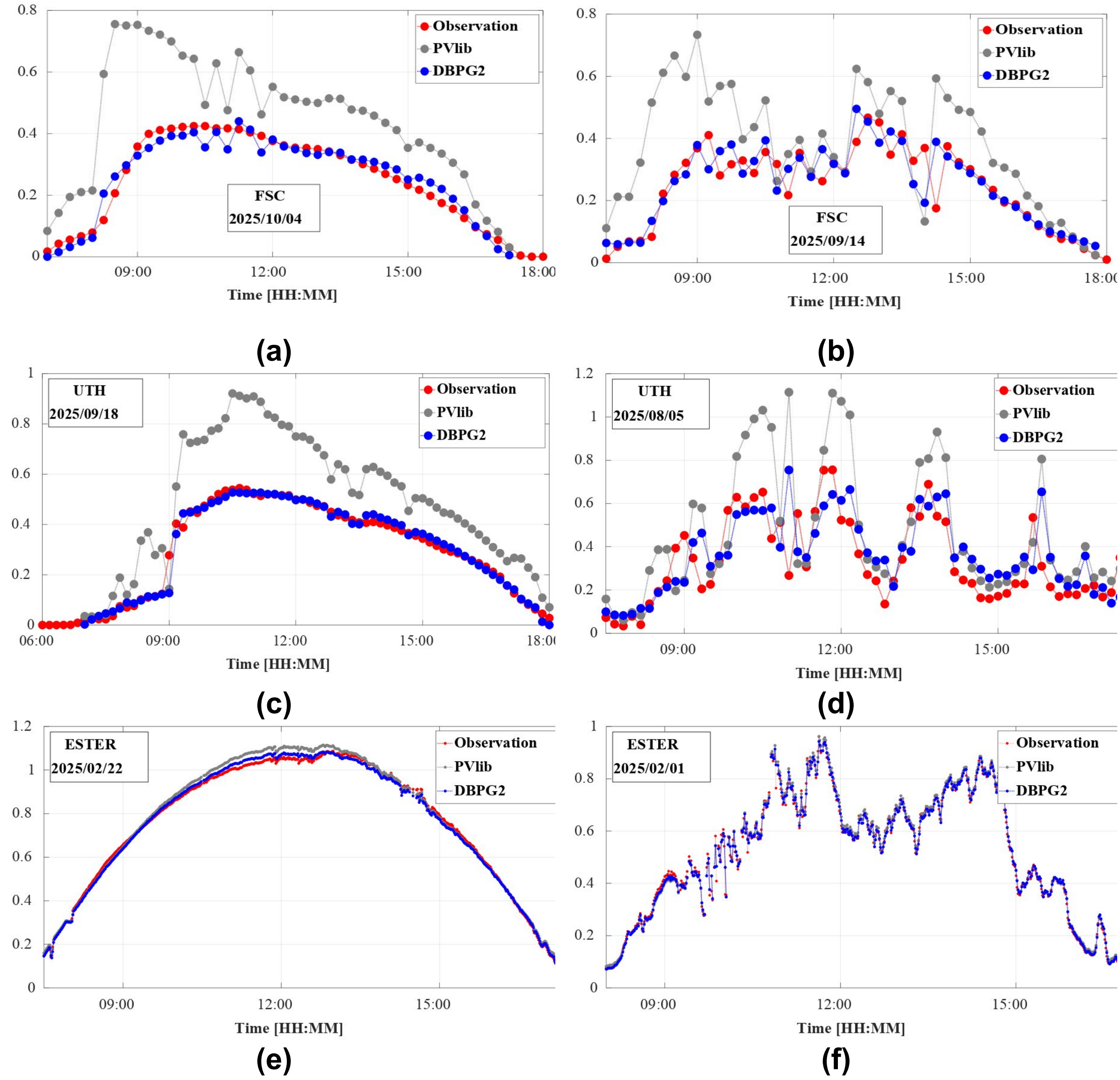


Fig. 4.10 Comparison of normalized power output profiles throughout the day between measured observations and simulated results obtained from PVlib and DBPG2 models: (a and b) FSC (c and d) UTH; (e and f) ESTER sites on selected dates.

In contrast, the DBPG2 model offers much closer PR estimated values to the measured ones across all months, with a relative deviation limited to approximately −6.3% to +0.3% (absolute deviations of −0.0345 to +0.0014 PR units, where the maximum underestimation occurs in March and the best agreement in August).

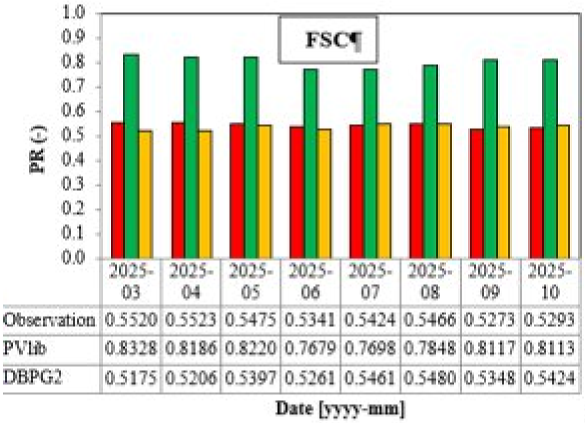
This demonstrates that the improved power estimation capability of the DBPG2 model directly translates into highly accurate performance ratio estimation in greenhouse-integrated bifacial PV systems.

The persistent PVlib overestimation indicates that greenhouse-specific phenomena—such as dynamic structural shading, partial obstruction of the sky view, and multiple internal reflections—substantially reduce the effective bifacial irradiance, yet remain unaccounted for in conventional bifacial modeling frameworks. Consequently, PVlib interprets the greenhouse system as operating under ideal favorable conditions (like an open field), leading to systematically inflated PR values.

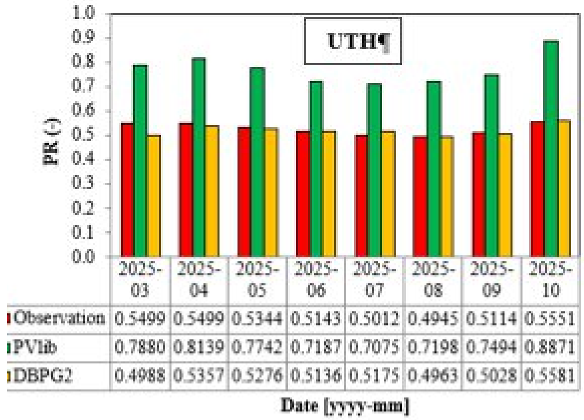
by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

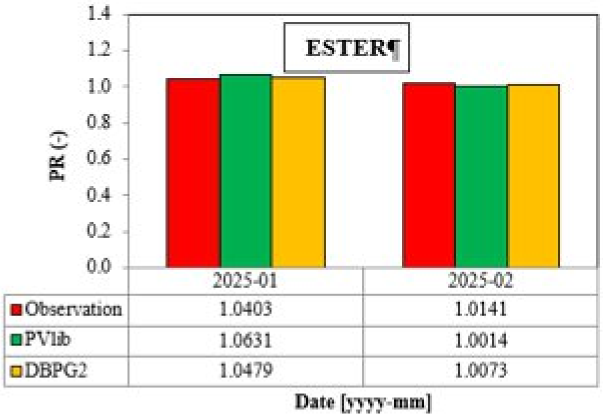
On the other hand, the high accuracy achieved by the DBPG2 model originates from its data-informed dynamic formulation, which reconstructs power of a BFPV system based on historical system response data rather than relying solely on simplified irradiance assumptions. By embedding the combined effects of structural shading, time-varying sky obstruction, and internal reflection losses into its learned response function, the DBPG2 model implicitly captures greenhouse-specific radiative behavior in a simple but accurate way.



(a)



(b)



(c)

Fig. 4.11 Comparing the monthly values of the observed PR with PVlib and the DBPG2 model estimation for each case.

As a result, the DBPG2 model is able to estimate the real irradiance–power relationship of greenhouse-integrated BFPV systems accurately, enabling reliable estimation of both instantaneous power output and cumulative performance ratio.

Such a large PVlib bias raises serious concerns regarding the use of PVlib in greenhouse applications, since PR is widely adopted as a key benchmark indicator for system performance evaluation, contractual guarantees, and techno-economic evaluation. An overestimation exceeding 40% would directly lead to misleadingly optimistic projections of the annual energy yield, economic profitability, and environmental impact mitigation. Meanwhile, the DBPG2 model eliminates this bias and delivers PR estimation within approximately much lower deviation from observations, thereby providing a robust and credible foundation for subsequent energy, economic, and environmental assessments of greenhouse-integrated BFPV systems.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

A similar trend is observed for the UTH greenhouse, as illustrated in Fig. 4.11b. The measured monthly PR values range between 0.5143 and 0.5551, whereas PVlib predicts values between 0.7187 and 0.8871, corresponding to a relative overestimation of approximately 40% to 60% (absolute deviations of +0.2044 to +0.3320 PR units, with the maximum error in October and the minimum error in June). This confirms that the systematic overestimation of greenhouse performance by PVlib is not site-dependent but inherent to its modeling assumptions, as discussed.
The DBPG2 model again shows strong agreement with measurements, with relative deviations limited to approximately −9.3% to −0.1% (absolute deviations of −0.0511 to −0.0007 PR units, where the maximum underestimation occurs in March and the best agreement in June). The consistency of the DBPG2 model accuracy across two independent greenhouse systems demonstrates that the model reliably generalizes to different greenhouse geometries and climatic conditions to provide high accuracy.
Similarly, in the UTH case, greenhouse-specific phenomena such as structural shading, sky obstruction, and internal reflection losses remain present; therefore, the same conclusion drawn for the FSC case continues to hold. PVlib, as a model relying on simplifying assumptions, exhibits not high accuracy in representing the real PV performance inside the greenhouse, which leads to substantial overestimation of power output and PR. The DBPG2 model, by reconstructing the effective response of the BFPV system from historical operational data, captures these complex interactions more realistically and thus provides reliable performance estimation for greenhouse-integrated BFPV systems.
A different behavior is observed for the ESTER Lab case, representing the outdoor reference system, as illustrated in Fig. 4.11c. The measured monthly PR values range between 0.9934 and 1.0502, while PVlib and DBPG2 predictions remain within a narrow deviation band throughout the year. For instance, in January, the measured PR is 1.0403, compared to 1.0631 estimated by PVlib and 1.0479 by the DBPG2 model, corresponding to relative deviations of +2.2% and +0.7%, respectively.
The close agreement between the estimated and measured PR values in the ESTER Lab case indicates that, in open-field conditions where greenhouse enclosure effects, structural shading, and internal reflection losses are absent, PVlib, with the simplifying assumptions, is able to provide high accuracy. Even in this case, the DBPG2 model maintains more accurate than PVlib, although its structural advantage becomes most evident in greenhouse-integrated configurations.
The PR values have been obtained for the whole period for each case by using the BFPV power estimation of PVlib and the DBPG2 model. The results are then compared with the observation. For all the three conditions, similar to the monthly values, the measured irradiance data is utilized for obtaining the reference yield.PR values in addition to the estimation error for modelling approaches are reported in Table 4.4. According to this table, for the FSC case, the measured cumulative PR equals 0.5417, while PVlib predicts 0.7975, corresponding to a relative error of 47.2%, which shows a substantial overestimation of greenhouse-integrated BFPV system performance, consistent with the monthly findings. The DBPG2 model estimates a cumulative PR of 0.5353, resulting in a relative error of only 1.2%, indicating excellent agreement with measurements.

Table 4.4. Comparing the observed PR for the whole period with PVlib and the DBPG2 model estimation for each case.

Case	Observation	PVlib	DBPG2 model	Pvlib error (%)	DBPG2 model error (%)

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

FSC	0.5417	0.7975	0.5353	47.2	1.2
UTH	0.5208	0.7546	0.5167	44.9	0.8
ESTER	1.0172	1.0086	1.0120	0.8	0.5

A similar outcome is observed for the UTH case. In this case, the measured cumulative PR has the value of 0.5208, whereas PVlib brings an estimation of 0.7546, yielding a relative error of 44.9%. The DBPG2 model, however, estimates a cumulative PR of 0.5167, corresponding to a relative error of only 0.8%. These results demonstrate that PVlib consistently overestimates the long-term performance of greenhouse-integrated BFPV systems, while DBPG2 reproduces the observed cumulative PR with high fidelity. This behavior directly reflects what has been previously observed in the monthly PR analysis and in the power estimation results, confirming that the long-term performance deviation originates from the simplifiying modelling assumptions utilized in PVlib, which leads to overestimation of power, yield, and consequently, PR of greenhouse-integrated BFPV systems, whereas the DBPG2 model effectively reduces this deviation by accurately reconstructing the real system response.

Conclusion

In this chapter, the performance of bifacial photovoltaic (BFPV) modules in a greenhouse environment was comprehensively modelled and evaluated. First, the radiative behaviour inside the greenhouse was analysed in order to quantitatively determine the effects of solar irradiance attenuation by the greenhouse cover, partial sky-view obstruction, internal reflections, and structural shading. The results showed that, on average, approximately 14–15% of the incoming solar radiation is reduced when passing through the greenhouse envelope, while internal reflections and dynamic shading lead to non-uniform irradiance conditions on the PV modules. This confirmed that conventional open-field PV models are unable to reproduce the real behaviour of greenhouse-integrated systems.
Next, the real operational performance of the PV systems installed in the FSC and UTH greenhouses was assessed using the Performance Ratio (PR) and compared with a reference open-field installation at the ESTER Laboratory. The results indicated that the PR of the greenhouse systems (approximately 0.52–0.55) is significantly lower than that of the open-field reference system (approximately 1.02). Controlled shadowing experiments further provided experimental confirmation of the severe reduction in power output in the presence of structural shading and highlighted dynamic shading as the primary cause of performance degradation.
Subsequently, two modelling frameworks, PVlib and DBPG2, were examined for power estimation. The results showed that PVlib, due to its reliance on the effective irradiance concept and its inability to account for dynamic shading and internal reflection effects, systematically overestimates both power output and PR, leading to cumulative PR errors of approximately 45–47% for greenhouse systems. In contrast, the DBPG2 model, which is based on historical data and solar-geometry-related parameters, was able to implicitly incorporate the combined effects of bifacial gain, structural shading, and sky-view obstruction. Consequently, the PR estimation error was reduced to below 1–2%, and model bias was almost entirely eliminated. Even for the open-field ESTER system, DBPG2 exhibited higher accuracy than PVlib.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

Overall, this chapter demonstrated that accurate modelling of BFPV system performance in greenhouse environments is not achievable without explicitly accounting for shading physics and enclosed-environment radiative conditions. The DBPG2 model therefore provides a reliable tool for predicting the power output and performance of greenhouse-integrated BFPV systems, which establishes a solid foundation for the energy, economic, and environmental analyses to be conducted in the subsequent stages of the project.
Based on the obtained results, the following items could be taken into account as the key achievements:

- Quantitative characterization of greenhouse radiative conditions. The indoor–outdoor irradiance relationship was experimentally quantified, identifying stable average cover transmittance and revealing secondary optical phenomena (internal reflections and dynamic shading). This delivers essential boundary-condition knowledge for the PV, microclimate, and digital twin models developed.
- Experimental evidence of greenhouse-specific PV performance losses. Real operational PR values were established for two full-scale greenhouse demonstrators and benchmarked against an outdoor reference system. Controlled shadowing experiments provided direct physical validation of structural shading as the dominant loss mechanism. This closes the experimental validation gap required for model calibration.
- Critical assessment of existing industry PV modelling tools. The applicability of PVlib for greenhouse-integrated BFPV systems was systematically evaluated. The analysis demonstrated that standard effective-irradiance-based models lead to large systematic overestimations of power and PR under greenhouse conditions. This finding provides an evidence-based justification for the development of dedicated greenhouse PV models within REGACE.
- Development and validation of a greenhouse-adapted BFPV performance model. The DBPG2 data-driven physical model was successfully adapted, trained, and validated for greenhouse-integrated systems. The model accurately reproduced real power output under complex dynamic shading and bifacial irradiance conditions, reducing prediction errors from ~20% to below 7% and eliminating systematic bias.
- Reliable PR estimation for greenhouse-integrated PV systems. DBPG2 achieved cumulative PR estimation errors below 1–2%, compared to ~45% overestimation by PVlib. This delivers a trustworthy performance modelling tool, directly supporting downstream techno-economic and environmental analyses.
- Transferable modelling framework across installation typologies. The DBPG2 model demonstrated consistent accuracy for both greenhouse-integrated and open-field reference systems, proving its robustness and generalisation capability — a key requirement for scalable digital twin development.
- Direct contribution to the digital twin and AI integration roadmap. The validated PV performance model developed in provides a calibrated, site-aware PV sub-model ready for integration with greenhouse microclimate, CFD, and crop-growth models, forming a core component of the holistic PV-greenhouse digital twin envisioned.

As the limitation, while the DBPG2 model achieved high predictive accuracy, it remains fundamentally data-driven and therefore requires an initial training period using historical system measurements for each new greenhouse installation. Additionally, spectral modifications induced by greenhouse covers and long-term ageing or soiling effects of the cover and module surfaces were not explicitly modelled and are only implicitly embedded through training data. The current formulation also represents

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

shading and reflection effects through learned response functions rather than explicit three-dimensional geometric ray-tracing, which may limit extrapolation to radically different structural designs without re-training.

Future developments should focus on enhancing the generalization capability of the DBPG framework so that accurate predictions can be achieved for new greenhouse geometries with minimal site-specific training data. Integrating explicit spectral transmission models of greenhouse covers and incorporating ageing and soiling dynamics will further improve long-term predictive accuracy. Coupling the DBPG model with three-dimensional geometric shading and radiative transfer models will allow hybrid physical–data-driven formulations capable of extrapolating across design configurations. Finally, embedding the validated DBPG-based PV model into the broader digital twin architecture — together with microclimate, CFD, and crop-growth models — will enable system-level optimisation of greenhouse design, energy management, and crop production under the REGACE project vision.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

References

Holmgren, W. F., Hansen, C. W. & Mikofski, M. A. 2018. Pvlib Python: A Python Package For Modeling Solar Energy Systems. *Journal Of Open Source Software,* 3**,** 884.

Sohani, A., Pierro, M., Bovesecchi, G., Moser, D. & Cornaro, C. 2025a. An Accurate Data-Driven Physical Model For Bifacial Pv Power Estimation. *Energy***,** 139048.

Sohani, A., Pierro, M., Moser, D. & Cornaro, C. 2025b. Comparison Of Physical Models For Bifacial Pv Power Estimation. *Energy Conversion And Management,* 327**,** 119515.

Stein, J., Reise, C., Castro, J. B., Friesen, G., Maugeri, G., Urrejola, E. & Ranta, S. 2021. Bifacial Photovoltaic Modules And Systems: Experience And Results From International Research And Pilot Applications. Sandia National Lab.(Snl-Nm), Albuquerque, Nm (United States); Fraunhofer ….

Zainali, S., Lu, S. M., Fernández-Solas, Á., Cruz-Escabias, A., Fernández, E. F., Zidane, T. E. K., Honningdalsnes, E. H., Nygård, M. M., Leloux, J. & Berwind, M. 2025. Modelling, Simulation, And Optimisation Of Agrivoltaic Systems: A Comprehensive Review. *Applied Energy,* 386**,** 125558.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Chapter 5. Modeling Crop Growth and Yield Formation

Introduction

Within the broader context of developing a digital twin for an agrivoltaic greenhouse, there was a need to integrate a model capable of representing crop development. The selected model had to be able to simulate both the physical variables required to provide feedback to the other sub-models, as well as those necessary for economic and food safety assessments. Furthermore, the model needed to be capable of simulating a wide range of crops in order to ensure applicability to contexts beyond the specific experimental conditions addressed in the present project. Given these premises, the FAO AquaCrop model (Salman et al., 2021; Smith and Steduto, 2012; Steduto et al., 2009, 2007) was selected, as it has been extensively used in the literature across a variety of contexts and for different crops. AquaCrop was employed to simulate the experimental cropping cycles carried out at the Fattoria Solidale del Circeo between 20 May 2025 and 1 September 2025. The crops selected for the experiment were tomato and eggplant, for which microclimatic data were available, including temperature, relative humidity, photosynthetically active radiation (PAR), CO_2 concentration, and solar radiation, all logged at 5-minute intervals. Irrigation data, also recorded every 5 minutes, were available, together with data from an in situ pedological survey. In addition, yield data were monitored at discrete intervals based on fruit ripening stages, along with plant development indicators, specifically the leaf area index (LAI).
Furthermore, a statistical model was developed to derive an indicator of the crop production status, which will be discussed in Section 5.3. The formulation of the multi-state model was based on data collected in the Volos greenhouse (Greece) during the period from 10 October 2024 to 10 December 2024. During this time span, a cucumber cultivation experiment was carried out in hydroponic greenhouses. The data available for the above-mentioned experiment were as follows: a) microclimate measurements logged every 15 minutes and aggregated to daily means of CO_2, RH, PAR and temperature, b) daily irrigation and drainage volumes, and c) weekly harvest weights for each crop line.

5.1. Overview of the AquaCrop model

AquaCrop goes beyond the previous models (Raes et al., 2023a) by separating the evapotranspiration (ET) into soil evaporation (E) and crop transpiration (Tr):

$$ET = E + Tr \tag{5.1}$$

and by modelling the yield (Y) as the product of biomass (B) and harvest index (HI):

$$Y = HI \cdot B \tag{5.2}$$

The equation at the core of the AquaCrop model to calculate the biomass is:

$$B = WP \cdot \sum Tr \tag{5.3}$$

where WP is the water productivity, and the Tr is the crop transpiration previously introduced.
To simulate crop growth, Eq. 5.3 requires "a set of additional model components including: the soil, with its water balance; the crop, with its development, growth and yield processes; and the atmosphere, with its thermal regime, rainfall, evaporative demand and carbon dioxide concentration. Additionally, some management aspects are explicitly considered (e.g.,

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

irrigation, fertilization, mulches, weeds, etc.), as they will affect the soil water balance, crop development and therefore final yield", as Fig. 5.1 illustrate (Raes et al., 2023a).

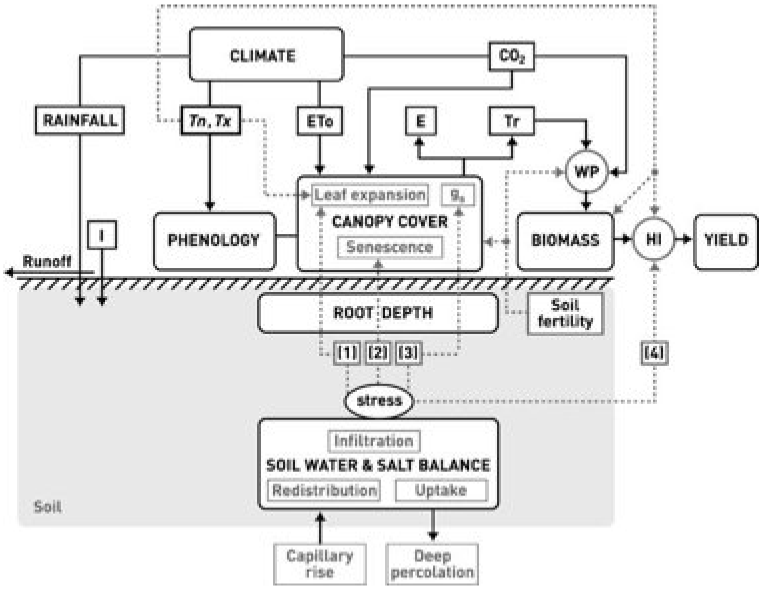


Fig. 5.1 Chart of AquaCrop indicating the main components of the soil–plant–atmosphere continuum and the parameters driving phenology, canopy cover, transpiration, biomass production, and final yield. [I, irrigation; Tn, minimum air temperature; Tx, maximum air temperature; ET_o, reference evapotranspiration; E, soil evaporation; Tr, canopy transpiration; g_s, stomatal conductance; WP, water productivity; HI, harvest index; CO_2, atmospheric carbon dioxide concentration; (1), (2), (3), (4), water stress response functions for leaf expansion, senescence, stomatal conductance and harvest index, respectively]. Continuous lines indicate direct links between variables and processes. Dashed lines indicate feedbacks". (Raes et al., 2023a)

The calculations are performed by the model through 4 sequential steps (see Fig. 5.2) at each daily timestep (Food and Agriculture Organization of the United Nations, 2023a):

1. Development of green canopy cover (CC): AquaCrop simulates the foliage development by means of green canopy cover (CC), which is the fraction of the soil surface covered by the canopy. Canopy cover ranges from zero (sowing - 0 % of the soil surface covered by the canopy) to a maximum value at mid-season, which can reach the value of 1 (100 % of the soil surface covered by the canopy);
2. Crop transpiration (Tr): Under the hypothesis of well-watered conditions, Tr is calculated by multiplying the reference evapotranspiration (ET_o) with a crop coefficient. The latter is proportional to CC and hence varies over time during the life cycle of the crop;
3. Above-ground biomass (B): The biomass produced by a crop is proportional to the cumulative amount of crop transpiration (ΣTr). The proportional factor, which modulates the biomass production, is the biomass water productivity (WP). AquaCrop makes use of a normalized biomass water productivity (WP*, that is the WP normalized for the effect of the climatic conditions) which makes WP* valid for diverse locations, seasons, and CO_2 concentrations;
4. Crop yield (Y): The above ground biomass (B) includes all mass generated by the crop during its development. Crop yield (Y) is obtained by means of a Harvest Index (HI), which represent the fraction of biomass that is the actual harvestable product. The HI is obtained adjusting the reference Harvest Index (HI_o) with a factor that accounts for stress effects.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

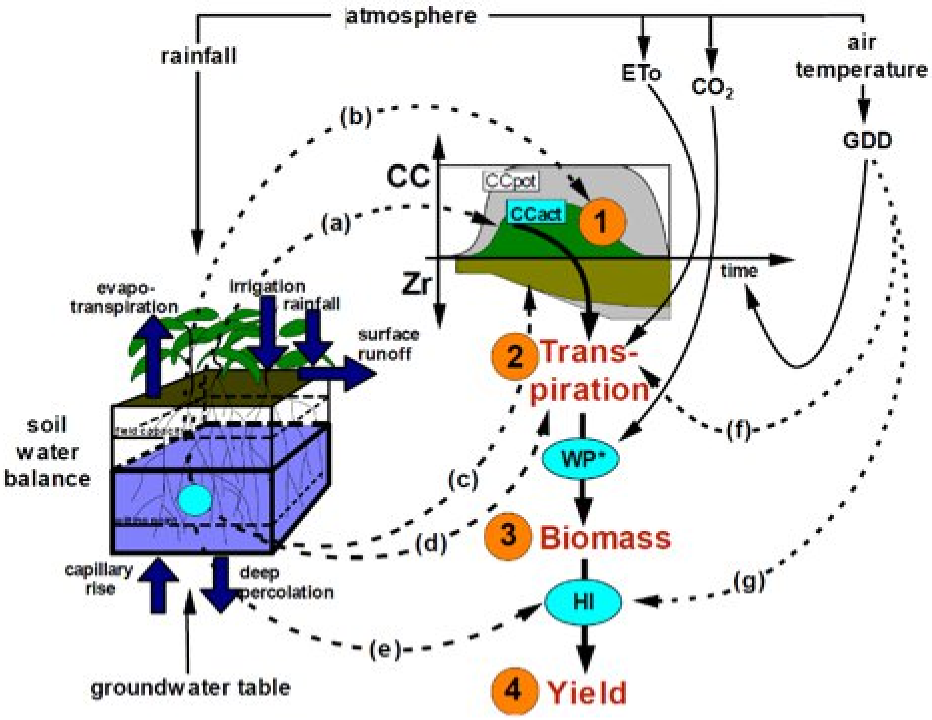


Fig. 5.2 "Calculation scheme of AquaCrop with indication of the 4 steps, and the processes (dotted arrows) affected by water stress (a to e) and temperature stress (f to g). CC is green canopy cover; Zr, rooting depth; ET_o, reference evapotranspiration; WP*, normalized biomass water productivity; HI, harvest index; and GDD, growing degree day. Water stress: (a) slows canopy expansion, (b) accelerates canopy senescence, (c) decreases root deepening but only if severe, (d) reduces stomatal opening and transpiration, and (e) affects harvest index. Cold temperature stress (f) reduces crop transpiration. Hot or cold temperature stress (g) inhibits pollination and reduces HI". (Food and Agriculture Organization of the United Nations, 2023a)

5.1.1 Required inputs

The inputs necessary for the preliminary configuration of the AquaCrop model fall into five primary categories:

1. Weather data;
2. Crop calendar;
3. Crop characteristics
4. Soil characteristics
5. Management practices.

5.1.1.1 Weather data

AquaCrop, for each step of the simulation (1 day), requires the minimum and maximum air temperature, rainfall, and the reference evapotraspiration (ET_o) as a proxy of the evaporative demand of the environment of interest. Moreover, the mean CO_2 concentration of the whole period has to be known as it affects water productivity biomass, WP*. Temperature is of main interest because it affects crop development, to the extent that, when limiting, growth and biomass accumulation can slow down. On the other hand, rainfall and ET_o are required for the water balance of the soil.

5.1.1.2 Crop calendar

The simulation obviously requires both the sowing/planting date and the harvesting date. The planting/sowing date can influence the duration of the crop life cycle due to the different

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

accumulation of growing degree days (GDD) during the various phases of the year. Although these are necessary inputs, they do not require further discussion.

5.1.1.3 Crop characteristics

The biological and biophysical processes involved in plant development, which AquaCrop is designed to model, are highly complex. To this end, the FAO model employs a set of crop parameters to describe the characteristics of different crops. In order to pursue its objectives, the FAO has calibrated these parameters for crops of greatest interest and has made these data available through the model itself. The model parameters are classified into: a) conservative parameters, which do not change over time, geographical location, or crop management practices; and b) non-conservative parameters, which require tuning when the cultivar or environmental conditions differ from those used by the FAO for calibration. These parameters will be discussed in greater detail in subsequent sections of this document.

5.1.1.4 Soil characteristics

The soil can be modeled using a structure of horizontal layers, each of which is characterized by its own physical properties. The physical properties of interest are those related to water retention; in particular, each layer is defined by the soil water content at saturation (θ_{sat}), at field capacity (θ_{FC}), and at permanent wilting point (θ_{WP}), as well as by the saturated hydraulic conductivity of the soil (K_{sat}). These quantities can be measured through in situ analyses, or indicative values provided by AquaCrop documentation (Tab. 5.1) for various soil texture classes (Soil Science Division Staff et al., 2017) can be used.

Table 5.1 Default soil characteristics. Based on Table 3.1 of AquaCrop Handbook I. (Food and Agriculture Organization of the United Nations, 2023a)

Soil textural class	Bulk density, ρ_b [ton·m^3]	θ_{sat} [vol %]	θ_{FC} [vol %]	θ_{WP} [vol %]	TAW [mm·m^{-1}]	Ksat [mm·day^{-1}]
Sand	1.71	36	13	6	70	3,000
Loamy sand	1.63	38	16	8	80	2,200
Sandy loam	1.56	41	22	10	120	1,200
Loam	1.42	46	31	15	160	500
Silt loam	1.42	46	33	13	200	575
Silt	1.52	43	33	9	240	500
Sandy clay loam	1.40	47	32	20	120	225
Clay loam	1.32	50	39	23	160	125
Silty clay loam	1.27	52	44	23	210	150
Sandy clay	1.32	50	39	27	120	35
Silty clay	1.21	54	50	32	180	100
Clay	1.19	55	54	39	150	35

In addition, the presence of gravel – which reduces the amount of water available to plants – can be specified, as well as soil penetrability, which affects and may limit root expansion within the soil. The soil profile is further characterized by the depth and salinity of a potentially present groundwater table and by its temporal variability.

5.1.1.5 Management practices

Management practices can be divided into: a) field management, and b) irrigation management. Field management refers to practices that affect the soil water balance (e.g., mulching, soil

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

bunds), fertility levels (with effects on water productivity, canopy growth, etc.), and weed management.
On the other hand, if the crop is not (or not exclusively) rainfed, detailed information on irrigation can be specified. The irrigation schedule (timing and applied water amount) can be defined, and different application methods can be selected (e.g., sprinkler, drip irrigation) (Food and Agriculture Organization of the United Nations, 2023b)

5.1.2 Running simulation

The FAO provides a stand-alone software program that implements the AquaCrop model, which can be downloaded and installed. The most recent version is 7.1, available on the dedicated portal ("Software | AquaCrop | FAO," n.d.). The following Fig. 5.3 shows the main menu of the software.

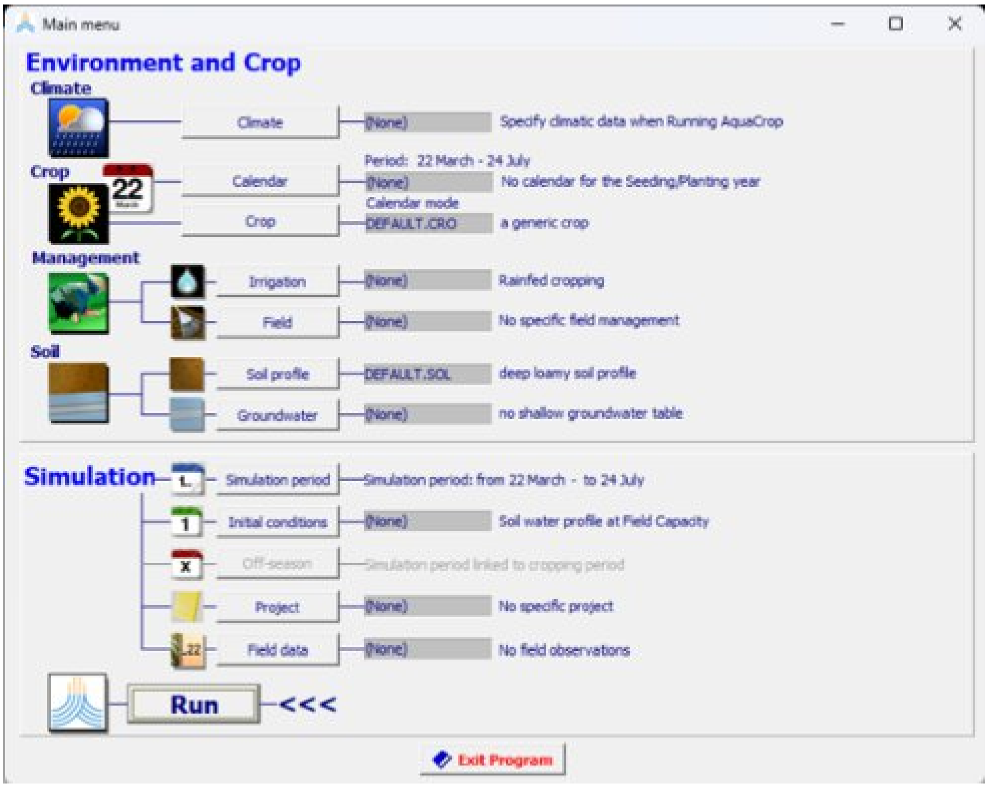


Fig. 5.3 FAO AquaCrop software.

Furthermore, in August 2016, a Python library (AquaCrop-OSPy) was released, resulting from a collaboration between researchers at the University of Manchester, the Water for Food Global Institute, the U.N. Food and Agriculture Organization, and Imperial College London (Kelly and Foster, 2021). The simulations carried out for this project were primarily conducted using the Python environment (version 3.13.4) and AquaCrop-OSPy (version 3.0.12). ("AquaCrop-OSPy," n.d.).

5.2. AquaCrop simulation at FSC

5.2.1 Soil

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

The area in which the Fattoria Solidale del Circeo is located falls within the Tyrrhenian coastal plains of central Italy, consisting of reclaimed alluvial surfaces with fluvio-palustrine and peaty deposits (Fig. 5.4).
Furthermore, the geological map of the Lazio Region generally identifies in this area predominantly silt-clay deposits in palustrine, lacustrine, and brackish facies, or black humic soils and sediments of lacustrine and palustrine environments, reclaimed with freshwater mollusks from the Pontine Plain (Fig. 5.5).

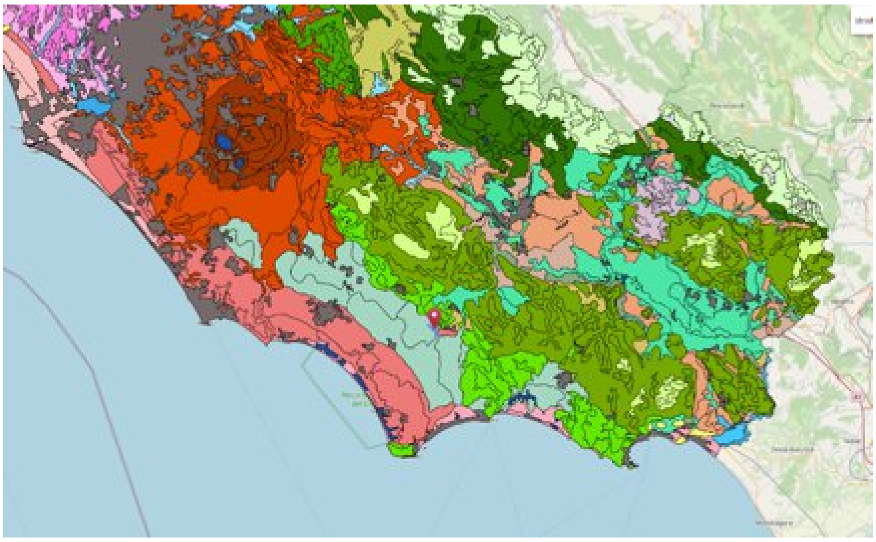


Fig. 5.4 Soil map of the Lazio Region. (“Soil Atlas of Lazio - geoportale.regione.lazio.it,” n.d.)

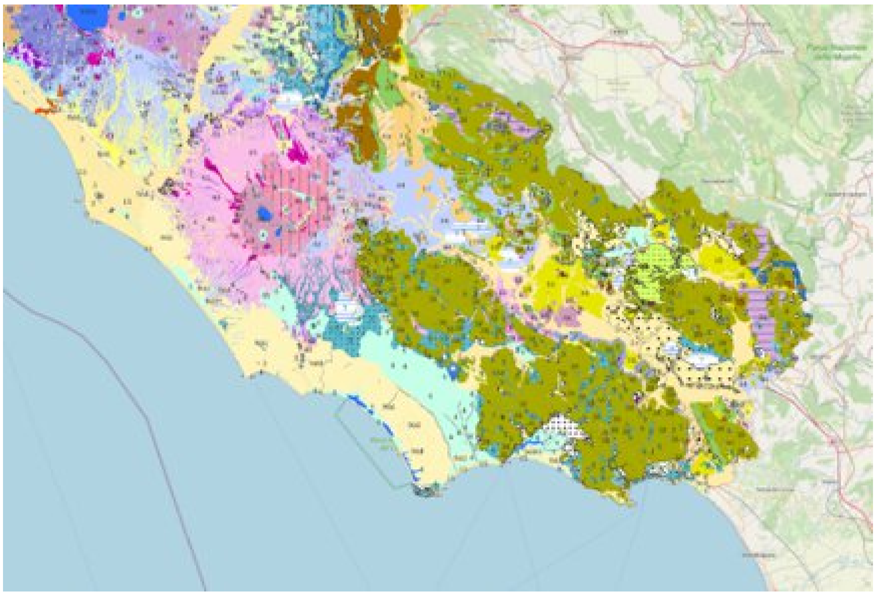


Fig. 5.5 Geological Map of the Lazio Region. (“Geological map of Lazio - geoportale.regione.lazio.it,” n.d.)

To better characterize the soil, between June and July 2025, pedological analyses were conducted on four soil samples collected in situ from a depth of 0 to 100 cm, with each layer having an identical thickness of 25 cm. The results of the analyses are presented in Table 5.2.

Table 5.2 Soil pedological analyses for FSC site.

Sample	**Bulk density, ρ_b [ton·m^3]**	**w_{sat} [% w./dry w.]**	**w_{FC} [% w./dry w.]**	**w_{WP} [% w./dry w.]**

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

P1-1 0/25	1.01	25.2	21.8	11.8
P1-2 25/50	1.05	28.6	24.1	13
P1-3 50/75	1.13	28	23.2	12.5
P1-4 75/100	1.10	25.5	20.1	10.9

Since the gravimetric water content is related to the volumetric water content through the following relationship,

$$\theta = w \cdot \rho_b \tag{5.4}$$

where θ is the volumetric water content, ρ_b is the bulk density and w is the gravimetric water content, the data can be expressed in terms of volumetric water content (Tab. 5.3).

Table 5.3 Soil pedological analyses: volumetric water contents.

Sample	**Bulk density, ρ_b [ton·m^3]**	**θ_{sat} [vol %]**	**θ_{FC} [vol %]**	**θ_{WP} [vol %]**
P1-1 0/25	1.01	25.5	22.0	11.9
P1-2 25/50	1.05	30.0	25.3	13.7
P1-3 50/75	1.13	31.6	26.2	14.1
P1-4 75/100	1.10	28	22.1	12.0

Based on Tables 5.1 and 5.3, field observations, and general information on the area derived from the maps provided by the Lazio Region, it is possible to formulate hypotheses to identify one of the twelve soil textural classes. The measured data do not exactly match any of the standard classes; however, the most plausible appear to be "Sandy loam" for the field capacity (FC) and wilting point (WP) values, and "Silt loam," which closely approximates the wilting point conditions and the general description of the geological area, although it seems to overestimate the saturation and field capacity values. Since no data on soil particle size distribution are available, it is not possible to perform evaluations using models such as the SPAW model by Saxton, provided by the USDA Agricultural Research Service. (Saxton and Rawls, 2006)

5.2.2 Weather data

To proceed with the simulations, the model requires four climate inputs at a daily time step:

1. minimum temperature;
2. maximum temperature;
3. precipitation;
4. reference evapotranspiration.

Furthermore, the software requires the average carbon dioxide (CO_2) concentration for the period. For this study, the average CO_2 concentration was calculated as the mean of all sensors available for the area under analysis over the entire reference period (from transplanting to harvest), thus making full use of the collected data.
The internal greenhouse microclimate temperature data are available and were used after being extracted from the historical series at a daily time step. The precipitation variable, representing rainfall, was considered uniformly zero throughout the period, as the greenhouse is covered. Moreover, during the summer period in Italy, rainfall is rare and would not have significantly affected the water availability in the surface layers of the soil inside the greenhouse.
The reference evapotranspiration (ET_o) variable was uniquely defined as the evapotranspiration of a reference surface: "A hypothetical reference crop with an assumed crop height of 0.12 m,

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

a fixed surface resistance of 70 s·m^{-1} and an albedo of 0.23" (Allen et al., 1998). Reference evapotranspiration is typically estimated using the Penman-Monteith formulation:

$$ET_o(\mathrm{mm}\cdot day^{-1}) = \frac{0.408\cdot\Delta\cdot(R_n - G) + \gamma\frac{900}{T+273}u_2\cdot(e_s - e_a)}{\Delta + \gamma\cdot(1 + 0.34\cdot u_2)} \quad (5.5)$$

where R_n is the net radiation at the crop surface [MJ·m^{-2}·day^{-1}], G is the soil heat flux density [MJ·m^{-2}·day^{-1}], T is the mean daily air temperature at 2 m height [°C], u_2 is the wind speed at 2 m height [m·s^{-1}], e_s is the saturation vapor pressure [kPa], e_a is the actual vapor pressure [kPa], e_s-e_a is the saturation vapor pressure deficit [kPa], Δ is the slope vapor pressure curve [kPa·°C^{-1}], and γ is the psychrometric constant [kPa·°C^{-1}].
The use of Eq. 5.5 requires knowledge of numerous variables that are not always available and has been calibrated for application in open-field conditions. Moreover, the aerodynamic conductance depends on the air velocity around and within the crop canopy (Katsoulas and Stanghellini, 2019), which in a closed greenhouse is very low (leading to numerical difficulties in estimating aerodynamic resistances) and is generally unknown. A model that is employed when not all the necessary variables are available is that of Hargreaves (Hargreaves and Allen, 2003; "Hargreaves, G.H. and Samani, Z.A. (1982) Estimating potential evapotranspiration. Journal of Irrigation and Drainage Engineering, 108, 223-230. - References - Scientific Research Publishing," n.d.; Hargreaves and Samani, 1985):

$$ET_o(\mathrm{mm}\cdot day^{-1}) = 0.0023\cdot(T_{mean} + 17.8)\cdot(T_{max} - T_{min})^{0.5}\cdot R_a \quad (5.6)$$

where T_{mean} [°C] is the mean air temperature, T_{max} [°C] and T_{min} [°C] are the daily maximum and minimum air temperature, respectively, and Ra [mm·day^{-1}] is the extraterrestrial radiation. This model, relying on extraterrestrial radiation, is not suitable for detecting differences related to the presence or absence of photovoltaic panels.
To overcome the difficulties encountered in quantifying ET_o within experimental greenhouses using the available variables, a simplified model (Baille et al., 1994; Katsoulas and Stanghellini, 2019) proposed by Baille et al. was employed:

$$\lambda ET = \mathrm{A}\cdot R_s + \mathrm{B}\cdot VPD \quad (5.7)$$

where λET [W·m^{-2}] is the canopy evapotranspiration, R_s [W·m^{-2}] is the solar radiation, VPD [kPa] is the vapor pressure deficit, A [-], and B [W·m^{-2}·kPa^{-1}]. The coefficients A and B can be found in the literature, and for the present study, those corresponding to tomato were used. (Carmassi et al., 2013). In the greenhouse, solar radiation measurements were available only above the panel plane, making the measurement insensitive to the shading effect of the installation. For this reason, it was decided to use PAR data measured at canopy height to indirectly derive the solar radiation driving the evapotranspiration process. Photosynthetically Active Radiation (PAR) is defined as the radiative flux in the spectral region between 400 and 700 nm, and corresponds to approximately 42% of the solar spectrum at sea level (Moretti and Marucci, 2019). PAR is often measured in μmol·s^{-1}·m^{-2}, and the conversion factor to W·m^{-2},considering the Sun as the source, is 4.57 (Thimijan and Heins, 1983), leading to the following relationship:

$$R_s(W\cdot m^{-2}) = \frac{PAR}{4.57\cdot 0.42} \quad (5.8)$$

where PAR unit is μmol·s^{-1}·m^{-2} as previously stated.
Finally, the vapor pressure deficit (VPD) can be easily calculated knowing the daily minimum and maximum values of temperature and relative humidity (Allen et al., 1998):

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

$$VPD\,(kPa) = e_s - e_a \tag{5.9}$$

where e_a [kPa] is the actual vapor pressure, and e_s [kPa] is the mean saturation vapor pressure. By substituting Eqs. 5.8 and 5.9 into Eq. 5.7, it is possible to calculate the evapotranspiration to be used in the simulations. The daily trend of the calculated reference evapotranspiration is shown in

Fig. 5.6 Reference for the evapotranspiration of the two plots cultivated with tomato.

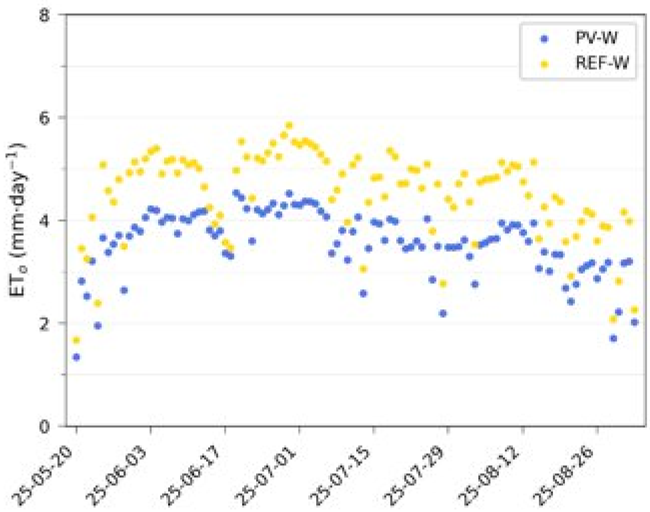


Fig. 5.7 Reference for the evaporation of the two plots cultivated with eggplant.

Figs. 5.6 and 5.7. As expected, evapotranspiration in the reference plot is consistently higher, regardless of orientation, due to the greater availability of solar radiation compared with plots covered by the photovoltaic system and affected by the resulting shading.

5.2.3 Supplementary relevant quantity

5.2.3.1 Growing Degree Days

In the field of agricultural sciences, the concept of heat units has been successfully used to describe the timing of biological processes since 1730, when Réaumur first introduced it. In the context of crop phenology and development the concept is often measured in growing degree days (GDD). The canonical form to calculate GDD is as follows:

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

$$GDD = \left(\frac{T_{max} + T_{min}}{2}\right) - T_{base} \quad (5.10)$$

where T_{max} and T_{min} are the daily maximum and minimum air temperature, respectively, and T_{base} is the temperature below which the process of interest does not progress (McMaster and Wilhelm, 1997).
In the present study, the calculation of GDD follows the following logic:

$$if\ \left(\frac{T_{max} + T_{min}}{2}\right) \leq T_{base} : GDD = 0 \quad (5.11a)$$

$$if\ \left(\frac{T_{max} + T_{min}}{2}\right) > T_{upper} : GDD = T_{upper} - T_{base} \quad (5.11b)$$

$$else\ : GDD = \left(\frac{T_{max} + T_{min}}{2}\right) - T_{base} \quad (5.11c)$$

where T_{upper} is the temperature above which the rate of the process of interest no longer increases, halts, or slows down (Bonhomme, 2000).
Moreover, following a literature review (Aldubai et al., 2022; Calado and Portas, 1987; Humphrey et al., 2024; Smith and Steduto, 2012; Thornley and Johnson, 1990), for tomato the GDD calculation was performed setting T_{base} a 8 °C (Thornley and Johnson, 1990), while T_{upper} was set at 28 °C (Smith and Steduto, 2012), whereas for eggplant the temperatures were set at 10 °C and 35 °C (Paredes et al., 2025). The progression of GDD throughout the crop development cycle is shown in Figs. 5.8 and 5.9.
It can be observed that for tomato (Fig. 5.8) the microclimate often saturated the relationship, which remained almost constant at the maximum daily range of 20 °C, whereas eggplant (Fig. 5.9) shows that the temperature inside the greenhouse rarely reached its T_{upper} which is higher than that of tomato.

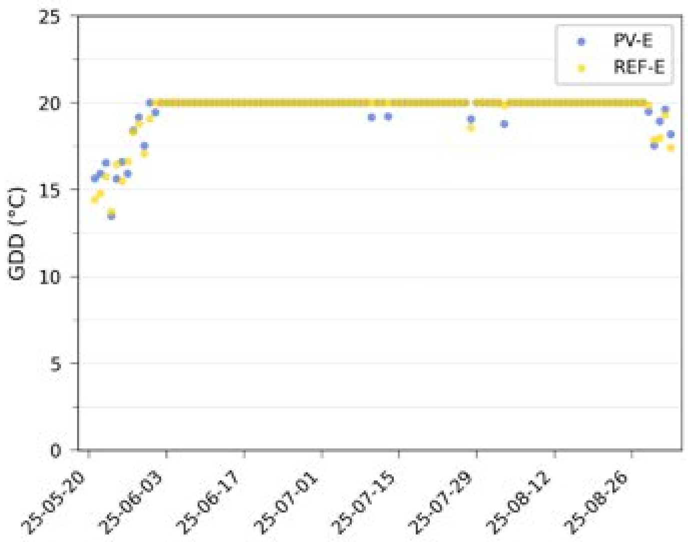


Fig. 5.8 GDD for the two plots cultivated with tomato.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

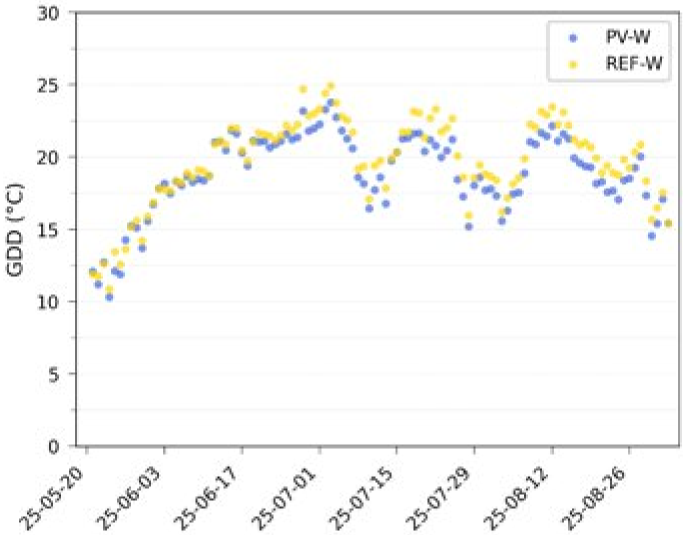


Fig. 5.9 GDD for the two plots cultivated with eggplant.

5.2.3.2 Leaf Area Index

One of the most relevant variables in the context of crop development is the Leaf Area Index (LAI), defined as "the leaf area per unit area of land" (Watson, 1947). This variable was monitored at intervals during the crop cycle for both tomato and eggplant, using the AccuPAR LP-80 ceptometer manufactured by METER Group. The instrument is designed to measure Photosynthetically Active Radiation (PAR) both above and below-canopy. The fraction of transmitted PAR can be linked to LAI with different theoretical models, and the instrument uses one of them to calculate LAI from the PAR measurements.

The adopted measurement protocol involved taking measurements during the central hours of the day and on mostly sunny days, performing multiple samplings per plot in areas representative of the average conditions of the measurement area. Two samplings were carried out for each plot, which were then averaged. In the literature, a relationship is known between LAI and the cumulative GDD during crop development (Carmassi et al., 2007). This correlation was used to interpolate the available LAI data and reconstruct the daily progression of LAI for each of the four plots. The Weibull function was employed to fit the data, and once the parameters of the function were determined, it was used to recalculate the daily LAI based on the cumulative GDD up to that date.

$$LAI = A \cdot \frac{k}{\lambda} \cdot \left(\frac{GDD}{\lambda}\right)^{k-1} \cdot e^{-\left(\frac{GDD}{\lambda}\right)^{k}} \tag{5.12}$$

Any slightly negative values generated were replaced with zero values. The results are shown in Figs. 5.10 and 5.11.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

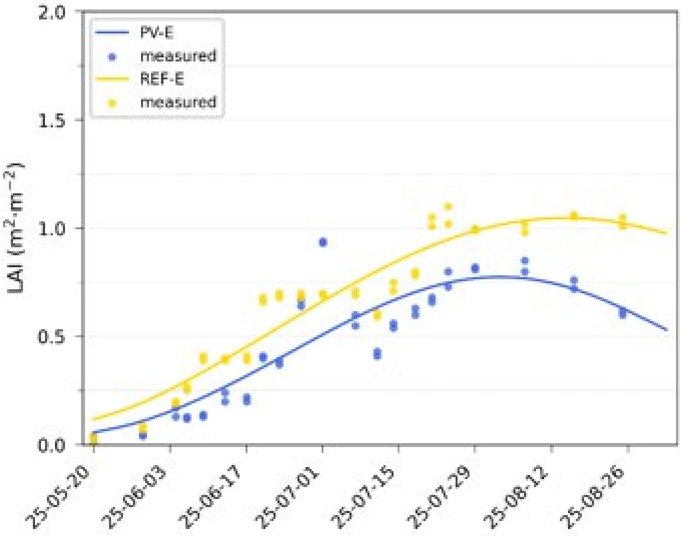


Fig. 5.10 Measured and interpolated LAI for each of the two tomato plots.

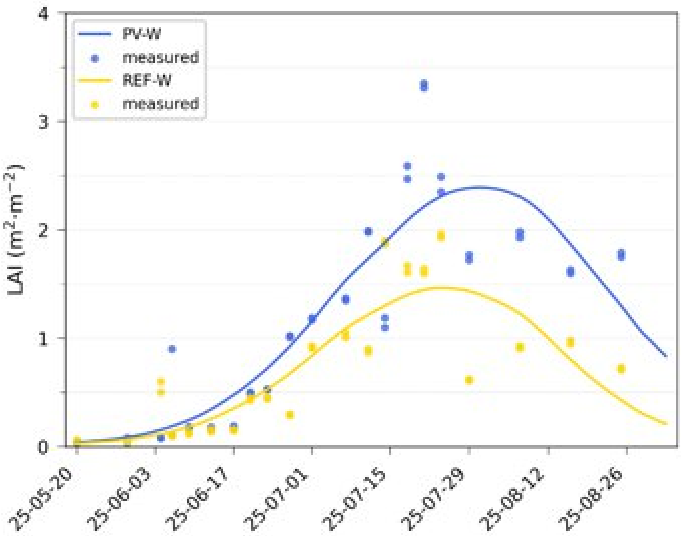


Fig. 5.11 Measured and interpolated LAI for each of the two eggplant plots.

Fig. 5.10 shows that tomato plants in the reference plot exhibit more vigorous growth compared with plants cultivated beneath the photovoltaic panels. In addition, the observed trends also indicate a faster rate of canopy contraction toward the end of the crop cycle, as if the plant life cycle had been shortened. Conversely, Fig. 5.11 shows an opposite pattern, with the reference plot consistently exhibiting lower LAI values than plants grown beneath the photovoltaic system. This may be attributed to the fact that DLI levels beneath the panels reached values optimal for eggplant development, whereas the opposite occurred for tomato, whose optimal light conditions were recorded in the unshaded reference area.

5.2.3.3 Canopy Cover

In an effort to find a measure of plant canopy development, researchers in the field have defined green Canopy Cover (CC) as a less costly alternative to measuring LAI. CC is defined as the portion of soil covered by the canopy projection relative to the total area (Jennings et al., 1999; Stevens et al., 2020). From this definition, it is clear that, unlike LAI – which can exceed unity depending on leaf morphology, canopy architecture, and ecological strategy – CC can reach at most a value of one. Moreover, the FAO AquaCrop model utilizes CC rather than LAI directly.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

The two variables are often correlated, and relationships linking them can be found in the literature. One such relationship, developed by Hsiao et al. (2009), assumes the following form (Hsiao et al., 2009):

$$CC = 1.005 \cdot \left(1 - e^{-06 \cdot LAI}\right)^{1.2} \tag{5.13}$$

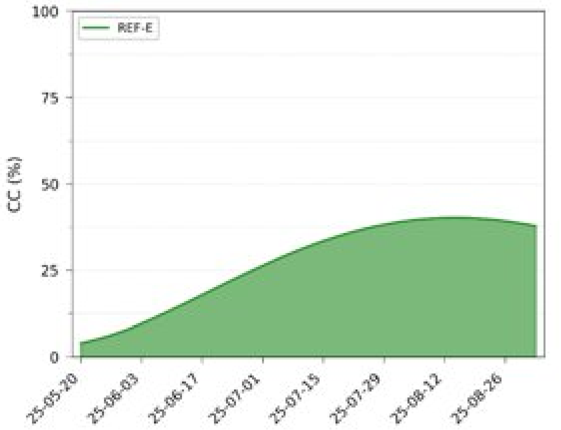


Fig. 5.12 Canopy Cover (CC) for the tomato reference plot.

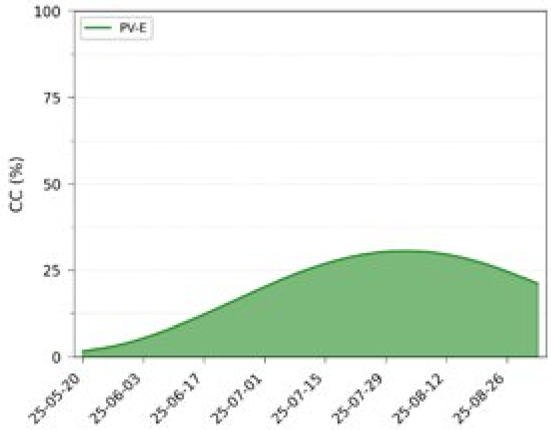


Fig. 5.13 Canopy Cover (CC) for the tomato PV plot.

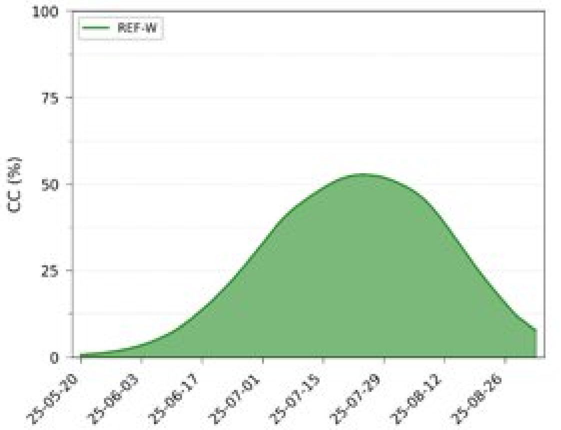


Fig. 5.14 Canopy Cover (CC) for the eggplant reference plot.

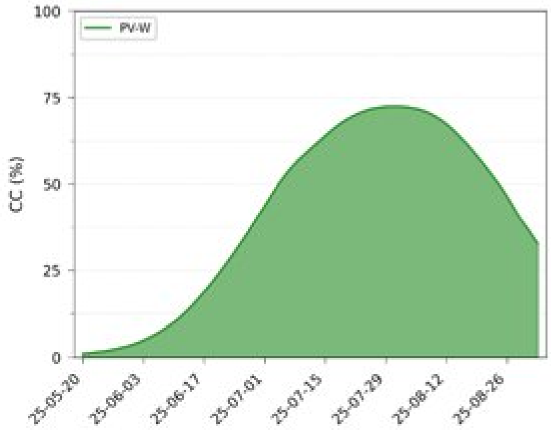


Fig. 5.15 Canopy Cover (CC) for the eggplant PV plot.

Eq. 5.13 has already been used in the literature for tomato (Cheng et al., 2022), and it will be employed in the present study to estimate CC from LAI for both tomato and eggplant. The results of the calculations based on the available data for the two plots where tomatoes and eggplants were grown are shown in Figs. 5.12 to 5.15.

5.2.4 Irrigation

In Italy, the summer season is hot and predominantly dry. Moreover, the greenhouse crops could not benefit from rainfall, nor was groundwater available. For this reason, a drip irrigation system was installed and monitored as illustrated in section 1.3. The cumulative daily irrigation water applied to the plot surface, in liters, was converted to mm·day^{-1} and input into the software as the irrigation schedule. Irrigation data were recorded separately for each of the four plots: 1) tomato reference (REF-E), 2) tomato PV (PV-E), 3) eggplant reference (REF-W), and 4) eggplant PV (PV-W).

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

5.2.5 Crop data

As previously mentioned, the input parameters for the FAO AquaCrop model are divided into conservative and non-conservative, the latter requiring field calibration. The values of these parameters and other information for the most relevant crops have been the subject of studies published by the FAO. (Raes et al., 2023b; Steduto et al., 2012).
Conservative parameters were generally kept unchanged, while non-conservative parameters were calibrated using data from the reference plot, i.e., the one without the photovoltaic installation. Once the parameters were calibrated, the simulation was repeated using the climate and irrigation data from the plot affected by the photovoltaic system. The model response was evaluated in terms of fresh yield, a variable that was monitored during the experiments. The crops involved in the experiments, which were conducted when the monitoring systems were fully operational, were tomato and eggplant. Finally, field observations indicated that the root system of the crops expanded more horizontally than vertically, as the soil below the first 30 cm was very compact. For this reason, a maximum rooting depth of 30 cm was set for all crops.

5.2.5.1 Tomato

During calibration, an improvement in the simulation results was observed by adopting a "Silt loam" soil class. Regarding crop parameters, in addition to calibrating the timing of phenological stages, adjustments were made to the reference harvest index in order to model a drastic yield reduction, which was recorded throughout the area and likely overlapped with the stresses experienced by the plants. Moreover, a delay of approximately two weeks in the onset of harvest was observed, which was modeled by shifting the parameter related to the start of production (time from transplanting to flowering) and by shortening the parameter associated with the duration of yield formation. The observed delay in the onset of the productive phase of the crop could have been caused by multiple factors, including initial water stress, difficulties for pollinating insects to access the greenhouse due to the use of an anti-avian net, and/or the disease itself, which affected productivity. The parameters used for the simulations are specified in Table 5.4.

Table 5.4 Input parameters in simulating the response of tomato using sing AquaCrop.

Parameter[1]	Default value	Value	Unit	Remarks
Base temperature	7	7	°C	Default
Upper temperature	28	28	°C	Default
Soil surface covered by an individual seedling	5 to 20	6	$cm^2 \cdot plant^{-1}$	Measured
Number of plants per hectare	15000 to 80000	10417	$plants \cdot ha^{-1}$	Estimated
Transplant to recovery	4	4	day	Default
CGC	0.12286	0.12286	day^{-1}	Default
CCx	75	Ref 40 – PV 30	%	Estimated
Time from transplant to start senescence*	Recovery + 1300-1600	55	day	Estimated
CDC	0.07238	0.07238	day^{-1}	Default
Time from transplant to maturity*	Recovery + 1500-2000	90	day	Estimated
Time from transplant to flowering*	Recovery + 250-400	48	day	Estimated
Length of the flowering stage*	600-900	42	day	Estimated
Crop determinancy linked with flowering	no	no	-	Default
Minimum effective rooting depth	0.3	0.3	m	Default
Maximum effective rooting depth	1	0.3	m	Estimated

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

Water productivity normalized for ET_o e CO_2	18	18	$g \cdot m^{-2}$	Default
Reference harvest index	55-65	40	%	Estimated
Duration of the yield formation	58	44	day	Estimated
Excess of potential fruits	100	0	%	Estimated

[1]: Conservative parameters not discussed here were kept at their default settings (Raes et al., 2023b)
*: Values in growing degree days (GDD) (Raes et al., 2023b)

5.2.5.2 Eggplant

Table 5.5 reports the parameters used for the simulations performed with AquaCrop for eggplant cultivation. For the simulation, the 'Silt loam' soil class, identified as optimal in the previous simulations, was maintained. Regarding crop parameters, in addition to calibrating the timing of phenological stages, adjustments were made to the reference harvest index and water productivity to adapt the model to the yield observed in the greenhouse. In the case of eggplant, the issues observed for tomato were not encountered. Moreover, no delay of approximately two weeks in the onset of harvest, as observed for tomato, was recorded. The use of the default profile prepared for tomato is justified by the fact that the two crops belong to the same family and genus, being different species of C_3 Solanaceae plants, and no FAO-published preset exists for eggplant.

Table 5.5 Input parameters in simulating the response of eggplant using sing AquaCrop.

Parameter[1]	**Default value**	**Value**	**Unit**	**Remarks**
Base temperature	7	10[2]	°C	Default
Upper temperature	28	35[2]	°C	Default
Soil surface covered by an individual seedling	5 to 20	20	$cm^2 \cdot plant^{-1}$	Measured
Number of plants per hectare	15000 to 80000	10417	$plants \cdot ha^{-1}$	Estimated
Transplant to recovery	4	4	day	Default
CGC	0.12286	0.12286	day^{-1}	Default
CCx	75	Ref 53 – PV 72	%	Estimated
Time from transplant to start senescence*	Recovery + 1300-1600	95	day	Estimated
CDC	0.07238	0.07238	day^{-1}	Default
Time from transplant to maturity*	Recovery + 1500-2000	114	day	Estimated
Time from transplant to flowering*	Recovery + 250-400	42	day	Estimated
Length of the flowering stage*	600-900	58	day	Estimated
Crop determinancy linked with flowering	no	no	-	Default
Minimum effective rooting depth	0.3	0.3	m	Default
Maximum effective rooting depth	1	0.3	m	Estimated
Water productivity normalized for ET_o e CO_2	18	23	$g \cdot m^{-2}$	Default
Reference harvest index	55-65	70	%	Estimated
Calendar days from sowing to start of yield formation	34	34	day	Estimated
Excess of potential fruits	100	40	%	Estimated

[1]: Conservative parameters not discussed here were kept at their default settings (Raes et al., 2023b)
*: Values in growing degree days for tomato (GDD) (Raes et al., 2023b)
[2]: Values from literature (Paredes et al., 2025)

5.2.6 Simulation results

The metrics used to evaluate the model and its fit to the experimental data were calculated using the following equations:

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{5.14}$$

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{5.15}$$

$$MBE = \frac{1}{n} \cdot \sum_{i=1}^{n} \hat{y}_i - y_i \tag{5.16}$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum \left(y_i - \overline{y}_i\right)^2} \tag{5.17}$$

$$nRMSE = \frac{RMSE}{y_{max}} \cdot 100 \tag{5.18}$$

where y_i are observed values, $\hat{y}_i$ are predicted values, and y_{max} is the maximum observed value. The choice of using the maximum value as the normalization parameter stems from the fact that the production curve is cumulative and starts from values that are practically zero. In this case, using the maximum is essentially equivalent to using the range for normalization. Moreover, the maximum production value is a meaningful metric, as it represents the total production over the entire cropping cycle. The following sections present the results related to tomato cultivation first, followed by eggplant.

5.2.6.1 Simulation of tomato production

The results of the simulations carried out after calibration for the REF-E plot, where the plants were not shaded by the photovoltaic installation, are presented graphically in Figs. 5.16 and 5.17.
The metrics calculated to evaluate the simulation are reported in Table 5.6.

Table 5.6 Performance metrics of the fresh yield simulation for the tomato reference plot.

Metric	Plot	Value	Unit
RMSE	REF-E	0.68	$ton \cdot ha^{-1}$
MAE		0.53	$ton \cdot ha^{-1}$
MBE		-0.21	$ton \cdot ha^{-1}$
R^2		0.92	-
nRMSE		9.56	%

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

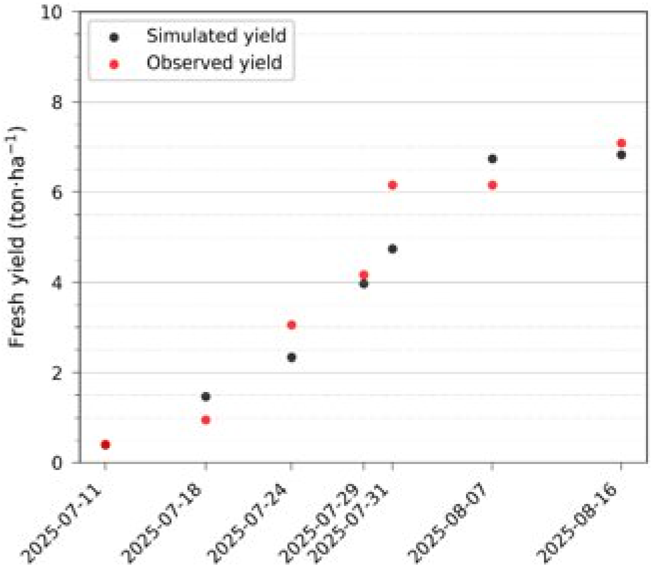


Fig. 5.16 Cumulative fresh yield for the tomato reference plot: observed (red) vs simulated (black).

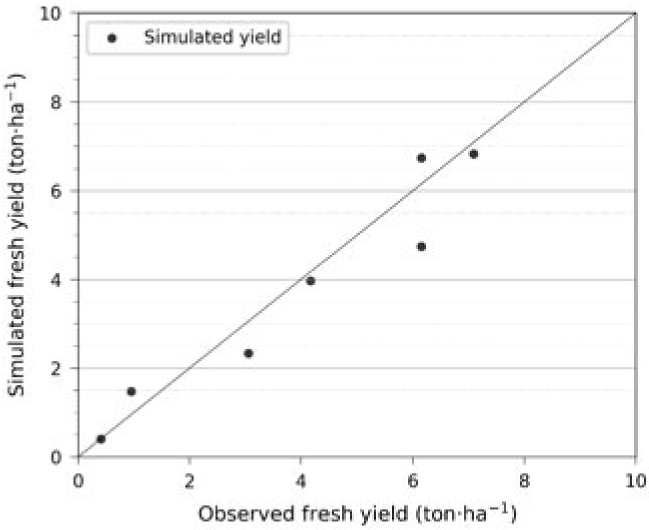


Fig. 5.17 Correlation between observed and simulated cumulative fresh yield for the tomato reference plot.

The metrics show a generally good agreement between simulated and observed data, with no outliers penalizing the Root Mean Square Error (RMSE) compared to the Mean Absolute Error (MAE). The negative value of the Mean Bias Error (MBE) indicates a slight underestimation by the model relative to the observed values, though of limited magnitude. The coefficient of determination (R^2) further indicates that the model can capture the data trend very well.

The simulation was then repeated without further modifying the model parameters, using the climate and irrigation data from the PV-E plot, which is located directly under the photovoltaic installation. The results are shown in Figs. 5.18 and 5.19.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

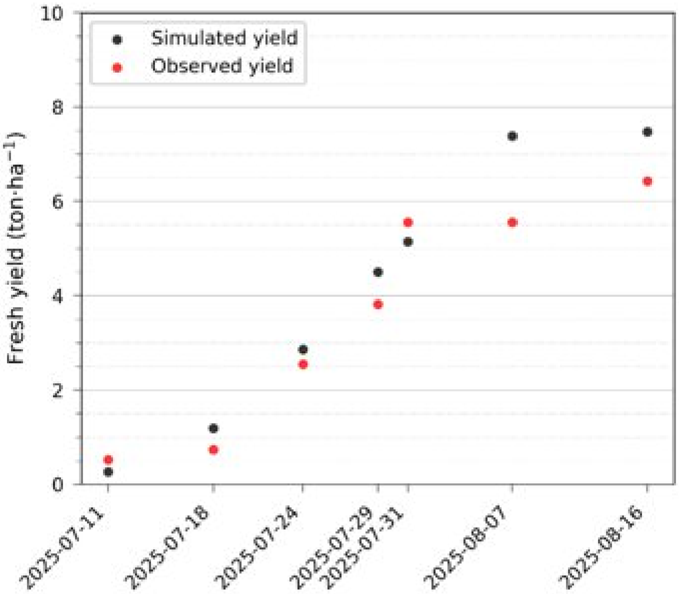


Fig. 5.18 Cumulative fresh yield for the tomato PV plot: observed (red) vs simulated (black).

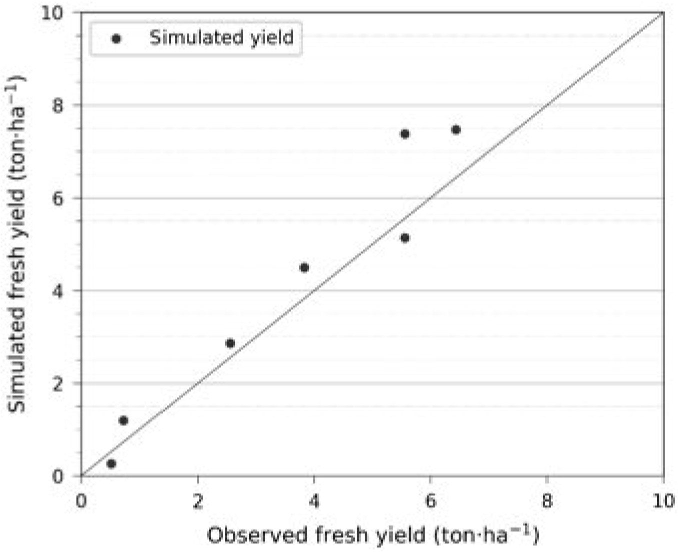


Fig. 5.19 Correlation between observed and simulated cumulative fresh yield for the tomato PV plot.

The metrics obtained from the results of the latest simulation are reported in Table 5.7, where the metrics from the reference plot simulation are also included to facilitate comparison.

Table 5.7 Performance metrics of the fresh yield simulation for the tomato plot below the PV panels.

Plot	RMSE [ton·ha^{-1}]	MAE [ton·ha^{-1}]	MBE [ton·ha^{-1}]	R^2 [-]	nRMSE [%]
REF-E	0.68	0.53	-0.21	0.92	9.56
PV-E	0.88	0.71	0.52	0.84	13.72

Analyzing the metrics, it is worth noting the consistency of the results when compared to those obtained during calibration, with the metrics remaining good but showing a slight deterioration. The most interesting datum is the MBE, which became positive, indicating a slight overestimation of the simulated fresh yield. This can be partly attributed to the generally slight decline in productivity observed in crops grown under the photovoltaic installations. It is

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

difficult to draw definitive conclusions, but one hypothesis is that the AquaCrop model may not fully capture the reduction in yield, likely linked to lower PAR availability, as this limitation is represented in the model only as a variation in potential evapotranspiration. Finally, it should be noted that toward the end of the experiment, both plots exhibited a peculiar halt in productivity. On August 7, 2025, no harvest was recorded, while the previous data show a peak in production.

5.2.6.2 Simulation of eggplant production

Similarly, the results of the simulations carried out after calibration for the REF-W plot, where the plants were not shaded by the photovoltaic installation, are presented graphically in Figs. 5.20 and 5.21.
Table 5.8 presents the key metrics that provide a comprehensive assessment of the simulation results.

Table 5.8 Performance metrics of the fresh yield simulation for the eggplant reference plot.

Metric	Plot	Value	Unit
RMSE		2.54	ton·ha^{-1}
MAE		1.61	ton·ha^{-1}
MBE	REF-W	-0.54	ton·ha^{-1}
R^2		0.98	-
nRMSE		5.57	%

Similarly, for the reference eggplant cultivation, after calibration, the metrics show a generally good agreement between simulated and observed data. The negative value of the Mean Bias Error (MBE) indicates a slight underestimation by the model compared to the observed values, although of limited magnitude. The coefficient of determination (R^2) further demonstrates that the model is able to capture the data trend very well. It is worth noting that only on two dates (late July and early August) a more pronounced deviation between simulated and observed values was observed.

The simulation was then repeated without further modifying the model parameters, using the climate and irrigation data from the PV-W plot, located directly under the photovoltaic installation. The results are shown in Figs. 5.22 and 5.23.

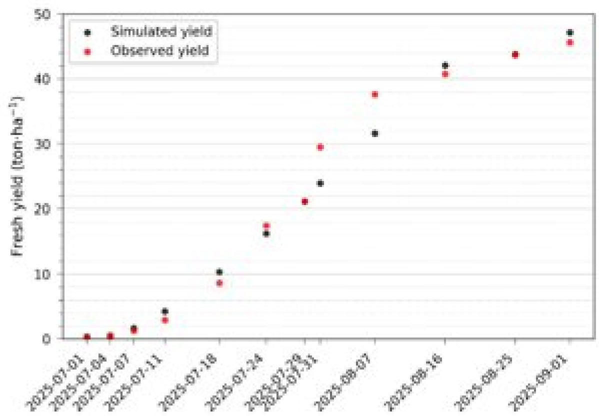


Fig. 5.20 Cumulative fresh yield for the eggplant reference plot: observed (red) vs simulated (black).

Funded by the European Union
by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

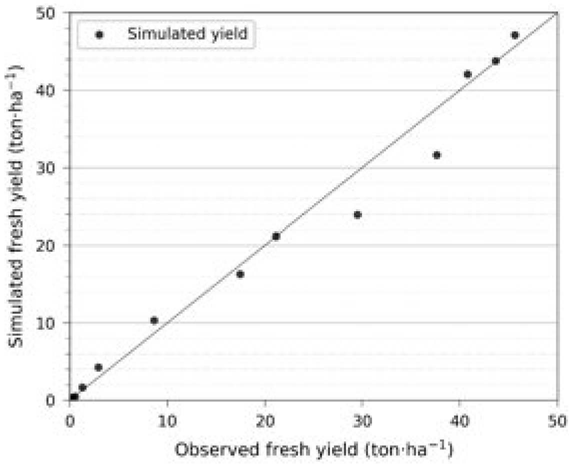


Fig. 5.21 Correlation between observed and simulated cumulative fresh yield for the eggplant reference plot.

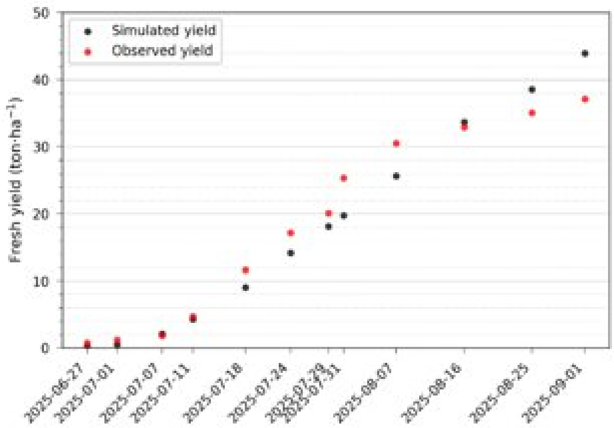


Fig. 5.22 Cumulative fresh yield for the eggplant PV plot: observed (red) vs simulated (black).

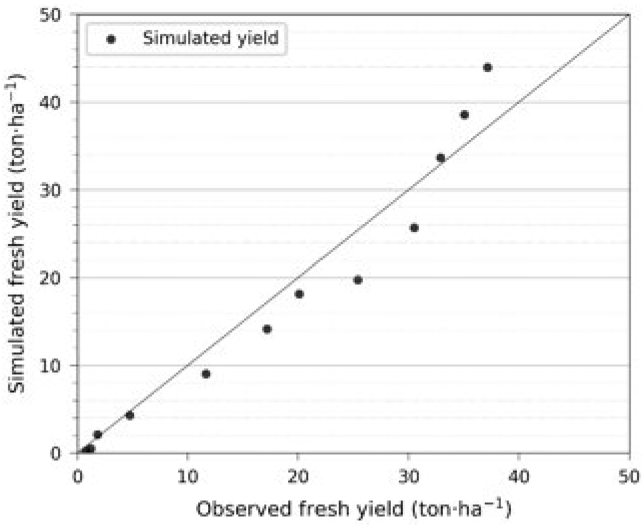


Fig. 5.23 Correlation between observed and simulated cumulative fresh yield for the eggplant PV plot.

The metrics obtained from the simulation results are reported in Table 5.9, where the metrics from the reference plot simulation are also included to facilitate comparison.

Table 5.9 Performance metrics of the fresh yield simulation for the eggplant plot below the PV panels.

Plot	RMSE [ton·ha^{-1}]	MAE [ton·ha^{-1}]	MBE [ton·ha^{-1}]	R^2 [-]	nRMSE [%]

Funded by the European Union
by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

REF-W	2.54	1.61	-0.54	0.98	5.57
PV-W	3.36	2.59	-0.69	0.94	9.05

From the comparison of the metrics obtained from the simulations for eggplant crops, it is worth highlighting the agreement between the two simulations and noting once again that the model provides consistent results even in this context. Once more, the metrics for the crop under the photovoltaic installation show a slight decline in performance. Furthermore, unlike the case of tomato, it can be observed that the model bias remained negative, despite the final cumulative values slightly overestimating the observed data in both cases. As with tomato cultivation, the model proved sensitive to the microclimatic differences measured between the two plots, producing different simulated values for the reference and below-PV plots. Nevertheless, the slight deterioration in the metrics may indicate that the inevitable difference in PAR levels between the two zones influences crop productivity that is not captured by the model. This hypothesis is supported by the fact that PAR, and therefore the overall radiation available at the crop level, influences not only evapotranspiration (considered by the model) but also photosynthesis.

5.3 A statistical approach: multi-state modeling of greenhouse cucumber yield dynamics under microclimate effects

5.3.1 Background and objectives

Modern greenhouses generate dense sensor streams (e.g., CO_2 concentration, relative humidity, and radiation), but day-to-day agronomic decisions are still often supported by static thresholds or models that treat yield as a single, continuous response (Matis, 1989). In practice, however, production frequently switches between distinct regimes (e.g., low, medium, and high output), and the speed of entering or existing these regimes is exactly what matters for short-horizon greenhouse steering.

The study proposes a continuous-time multi-state (Markov) modelling framework to convert routine monitoring data into operational indicators. Daily cucumber yield is discretized into three ordered “yield states”, and microclimate variables are linked to the transition rates between states.

The main objective is to quantify how day-to-day deviations in relative humidity (RH), photosynthetically active radiation (PAR), and CO_2 accelerate or delay upgrades (low→medium, medium→high) and regressions (5.24), and to translate the estimated dynamics into interpretable quantities such as waiting times and 7–30 day transition probabilities.

Funded by the European Union
by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

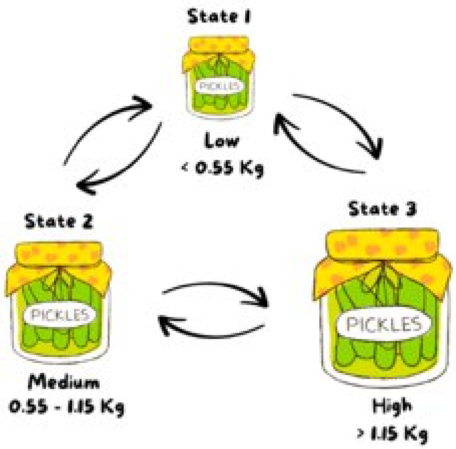


Fig. 5.24 Conceptual three-state space for daily cucumber yield (low, medium, high) with bidirectional transitions. Direct low↔high jumps were not observed and are constrained in the fitted model.

5.3.2 Methodology

The modelling approach treats each crop line as a subject observed repeatedly over time, whose yield class evolves as a continuous-time Markov chain. Let S(t) denote the state occupied at time t, with states {1,2,3} corresponding to low, medium, and high yield. The model is defined through transition intensities $q_{rs}(t)$, which represent the instantaneous rate of moving from state r to state $s, (r \neq s)$. All intensities are collected in a generator matrix Q, whose diagonal entries ensure that each row sums to zero. Sojourn (residence) times in a state are exponentially distributed, and the expected time spent in state r is $1/\left(-q_{rr}\right)$.
Microclimate covariates enter through a log-linear regression on each allowed transition intensity. In practice, exponentiated regression coefficients are interpreted as hazard ratios (HRs): the multiplicative change in the transition rate associated with a +1 standard deviation increase in a covariate. For example, $HR = 2$ on the 1→2 transition means that the move from low to medium yield is twice as fast under a one-SD higher covariate value.
The work estimates the model by maximum likelihood using the *msm* package in R (Christopher, 2011). Because direct transitions between the lowest and highest states were empirically absent, the fitted generator constrains 1↔3 intensities to zero, forcing movement between low and high yield to proceed via the intermediate state. Transition probabilities over a horizon Δt (e.g., 7, 14, or 30 days) are obtained via the matrix exponential $P(\Delta t) = exp(Q\Delta t)$, providing decision-ready forecasts.

5.3.3 Data

The study is based on a hydroponic cucumber experiment conducted in four adjacent greenhouse compartments in Volos (Greece) during 10 Oct–10 Dec 2024 (62 days). Each greenhouse hosts six crop lines, giving 24 experimental units observed daily.
Greenhouses differ by treatment: one control compartment (reference microclimate), one with CO_2 enrichment, one with a photovoltaic (PV) roof, and one combining PV with CO_2 enrichment. Three complementary data streams were used:

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

- Microclimate measurements logged every 15 minutes and aggregated to daily means (CO_2, RH, PAR; temperature was also available).
- Daily irrigation and drainage volumes, combined into a net water balance.
- Weekly harvest weights for each crop line, forward carried to obtain a daily yield series (reference greenhouse's daily yield trajectories are shown in Fig. 5.26).

Pre-processing includes (i) filling occasional gaps in the 15-minute log with a short neighborhood smoother that preserves the daily cycle, (ii) screening covariates through correlations computed on detrended within-greenhouse anomalies to reduce spurious seasonality-driven associations (Fig. 5.25), and (iii) standardizing CO_2, RH, and PAR within each greenhouse (z-scores) so that effects represent day-to-day departures from the typical conditions of each compartment. Temperature and net water balance were excluded from the final models due to strong collinearity with PAR and limited transition-event counts.

Table 5.10 Greenhouse code – treatment.

Greenhouse code	Treatment
GH-3	Control (reference microclimate)
GH-2	CO_2 enrichment
GH-4	PV roof
GH-5	PV roof + CO_2 enrichment

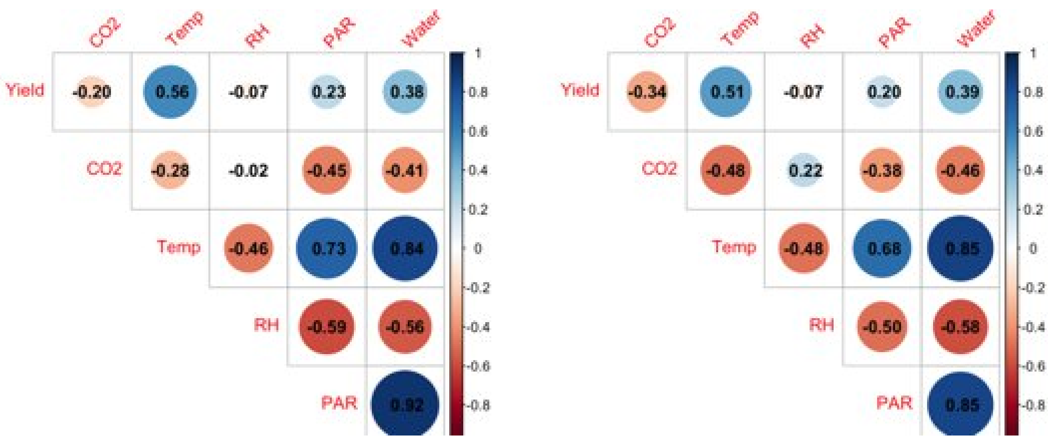


Fig. 5.25 Correlation patterns among detrended within-greenhouse anomalies (GH-3 only, left; all greenhouses pooled, right). Temperature and water balance are strongly related to PAR, motivating a parsimonious covariate set.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

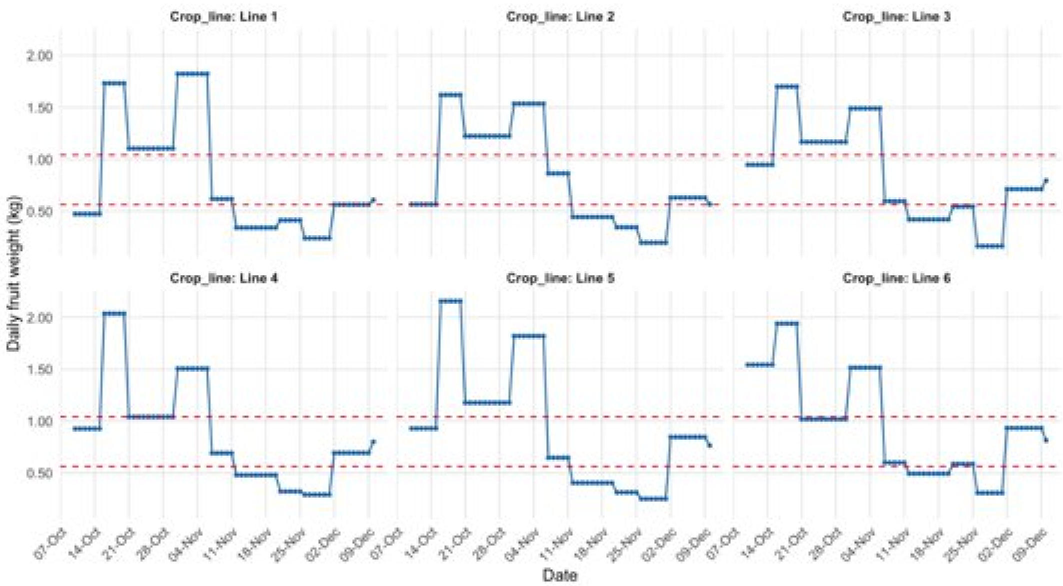


Fig. 5.26 Example of daily yield trajectories (control greenhouse, six crop lines) showing plateaus and the thresholds used to discretise yield into three ordered states.

5.4 Results

The model is first fitted on the control greenhouse (GH-3) to validate the workflow, and then on a pooled dataset across all four compartments with greenhouse fixed effects. Across analyses, the estimated dynamics display a clear "through-the-middle" pattern: the intermediate state acts as the main gateway between low and high production, and direct low↔high jumps are negligible.
Single-greenhouse (GH-3) results indicate that low-yield periods can be persistent under baseline conditions, while reaching the high-yield state is associated with strong short-term persistence. In 30-day forecasts, a crop line starting in the low state remains low with high probability, whereas a line starting in the medium state has substantial probability of upgrading to high yield.
Pooling the four greenhouses (24 crop lines; 1,488 line-days) sharpens inference on microclimate effects and shows consistent, agronomically meaningful patterns:

- RH is a dominant lever: higher-than-usual RH substantially speeds upgrades (1→2 and 2→3) and strongly reduces regressions (2→1 and 3→2).
- PAR has a similarly strong effect, accelerating both upgrade transitions and slowing downgrades.
- Day-to-day deviations in CO_2 show no clear pooled effect once RH and PAR are included.

These patterns are consistent with greenhouse cucumber studies in which total solar radiation is the dominant driver of yield, and with the classic rule-of-thumb that a 1% increase in light yields approximately 1% more product (Ding, 2019) and (Marcelis, 2005). Quantitatively (pooled model), a +1 SD increase in RH multiplies the 1→2 intensity by about 5.10 and the 2→3 intensity by about 2.64, while decreasing the 3→2 intensity to roughly 0.13 of its baseline. PAR increases the 1→2 and 2→3 intensities by about 2.63 and 2.78, respectively, and reduces the 2→1 intensity to about 0.12. Greenhouse fixed effects (CO_2 enrichment and/or PV roof) are comparatively uncertain once microclimate covariates are controlled for, suggesting that short-horizon yield-regime dynamics are largely explained by daily RH and PAR conditions. These estimates can be translated into actionable probabilities. For example, as shown in Fig.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

5.27, under average covariate conditions in the pooled model, a line in the low state today has roughly a 47% chance to remain low after 30 days, a 42.5% chance to move to medium, and a 10.5% chance to reach high yield. The overall interpretation is that raising RH can help lines escape and avoid falling back into low productivity, while increasing available light (natural or supplemental) is key to reaching and sustaining peak yield.

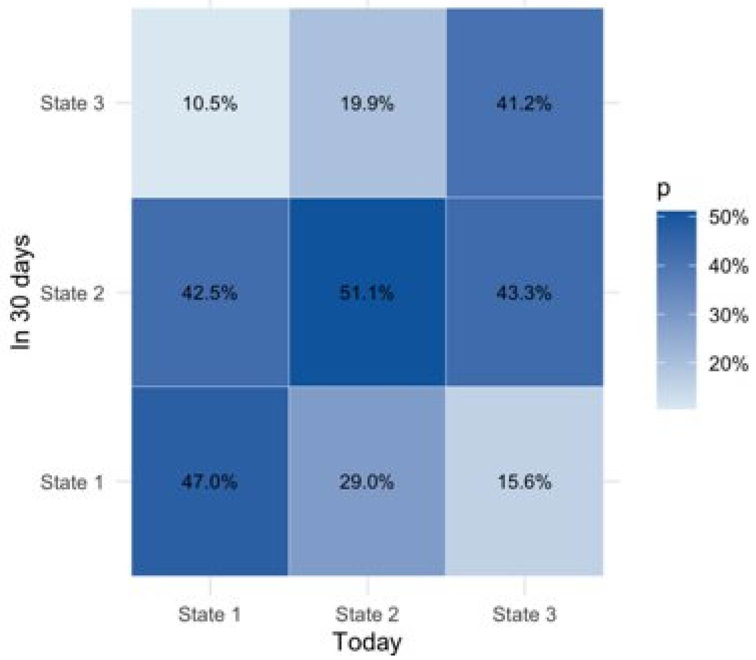


Fig. 5.27 Pooled model: 30-day transition probabilities between yield states under average covariate conditions (columns: state today; rows: state in 30 days).

Conclusion

In conclusion, the FAO AquaCrop model appears to be successfully applicable to the context of agrivoltaic greenhouses; however, further investigations are needed to explore the complexity of the physical and biological processes involved. It is worth emphasizing the criticalities encountered, related to the large number of parameters that must be measured in order to achieve accurate modeling. Moreover, many of these parameters require dedicated sensor systems or are difficult to monitor under field conditions. As previously discussed, the FAO AquaCrop model can be applied to different crop types, but it requires a preliminary calibration effort that cannot be overlooked. Future research developments could focus on the need to develop simplified and more general models capable of interfacing with a digital twin and that can be readily used by practitioners with commonly available resources. Overall, the work based on multi-state model demonstrates that a lightweight multi-state model can turn greenhouse monitoring data into interpretable, probabilistic indicators (expected waiting times and upgrade/downgrade risks) that support near-term climate-control decisions and can be embedded as a decision layer in greenhouse digital twins. Although several limitations qualify the conclusions of the multi-state model results (e.g., weekly harvests forward-carried to daily values), the analysis still yields simple, defensible, and implementable rules of thumb for growers. In conclusion, the multi-state model is a well-established statistical framework in literature and has been applied in a variety of contexts; however, in this study it is employed in an innovative manner within the agrivoltaic context.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

References

Aldubai, A.A., Alsadon, A.A., Migdadi, H.H., Alghamdi, S.S., Al-Faifi, S.A., Afzal, M., 2022. Response of Tomato (Solanum lycopersicum L.) Genotypes to Heat Stress Using Morphological and Expression Study. Plants 11, 615. https://doi.org/10.3390/PLANTS11050615/S1

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop Evapotraspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56, FAO - Food and Agriculture Organization of the United Nations. FAO - Food and Agriculture Organization of the United Nations.

AquaCrop-OSPy [WWW Document], n.d. URL https://aquacropos.github.io/aquacrop/ (accessed 12.9.25).

Baille, M., Baille, A., Delmon, D., 1994. Microclimate and transpiration of greenhouse rose crops. Agric. For. Meteorol. 71, 83–97. https://doi.org/10.1016/0168-1923(94)90101-5

Bonhomme, R., 2000. Bases and limits to using 'degree.day' units. European Journal of Agronomy 13, 1–10. https://doi.org/10.1016/S1161-0301(00)00058-7

Calado, A.M., Portas, C.M., 1987. BASE-TEMPERATURE AND DATE OF PLANTING IN PROCESSING TOMATOES. Acta Hortic. 185–194. https://doi.org/10.17660/ACTAHORTIC.1987.200.17

Carmassi, G., Bacci, L., Bronzini, M., Incrocci, L., Maggini, R., Bellocchi, G., Massa, D., Pardossi, A., 2013. Modelling transpiration of greenhouse gerbera (Gerbera jamesonii H. Bolus) grown in substrate with saline water in a Mediterranean climate. Sci. Hortic. 156, 9–18. https://doi.org/10.1016/J.SCIENTA.2013.03.023

Carmassi, G., Incrocci, L., Maggini, R., Malorgio, F., Tognoni, F., Pardossi, A., 2007. An aggregated model for water requirements of greenhouse tomato grown in closed rockwool culture with saline water. Agric. Water Manag. 88, 73–82. https://doi.org/10.1016/J.AGWAT.2006.10.002

Cheng, M., Wang, H., Fan, J., Xiang, Y., Liu, X., Liao, Z., Abdelghany, A.E., Zhang, F., Li, Z., 2022. Evaluation of AquaCrop model for greenhouse cherry tomato with plastic film mulch under various water and nitrogen supplies. Agric. Water Manag. 274, 107949. https://doi.org/10.1016/J.AGWAT.2022.107949

Food and Agriculture Organization of the United Nations, 2023a. AquaCrop Training Handbook. Book I: Understanding AquaCrop. FAO, Rome.

Food and Agriculture Organization of the United Nations, 2023b. AquaCrop Training Handbook. Book II: Running AquaCrop. FAO, Rome.

Geological map of Lazio - geoportale.regione.lazio.it [WWW Document], n.d. URL https://geoportale.regione.lazio.it/maps/new?layer=geonode:carta_geologica_wgs84&view=True#/ (accessed 1.15.26).

Hargreaves, G.H., Allen, R.G., 2003. History and Evaluation of Hargreaves Evapotranspiration Equation. Journal of Irrigation and Drainage Engineering 129, 53–63. https://doi.org/10.1061/(ASCE)0733-9437(2003)129:1(53)

Hargreaves, G.H. and Samani, Z.A. (1982) Estimating potential evapotranspiration. Journal of Irrigation and Drainage Engineering, 108, 223-230. - References - Scientific Research Publishing [WWW Document], n.d. URL https://www.scirp.org/reference/referencespapers?referenceid=926531 (accessed 12.9.25).

Hargreaves, G.H., Samani, Z.A., 1985. Reference Crop Evapotranspiration from Temperature. Appl. Eng. Agric. 1, 96–99. https://doi.org/10.13031/2013.26773

Hsiao, T.C., Heng, L., Steduto, P., Rojas-Lara, B., Raes, D., Fereres, E., 2009. Aquacrop-The

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

FAO crop model to simulate yield response to water: III. Parameterization and testing for maize. Agron. J. 101, 448–459. https://doi.org/10.2134/AGRONJ2008.0218S;JOURNAL:JOURNAL:14350645;PAGE:STRING:ARTICLE/CHAPTER
Humphrey, M., Dickson, R.W., Padhye, S., 2024. Quantifying Temperature Effects on Development Rate and Plant Quality of Compact Container-grown Tomato. Horttechnology 34, 713–719. https://doi.org/10.21273/HORTTECH05477-24
Jennings, S.B., Brown, N.D., Sheil, D., 1999. Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. Forestry: An International Journal of Forest Research 72, 59–74. https://doi.org/10.1093/FORESTRY/72.1.59
Katsoulas, N., Stanghellini, C., 2019. Modelling Crop Transpiration in Greenhouses: Different Models for Different Applications. Agronomy 2019, Vol. 9, Page 392 9, 392. https://doi.org/10.3390/AGRONOMY9070392
Kelly, T.D., Foster, T., 2021. AquaCrop-OSPy: Bridging the gap between research and practice in crop-water modeling. Agric. Water Manag. 254, 106976. https://doi.org/10.1016/J.AGWAT.2021.106976
McMaster, G.S., Wilhelm, W.W., 1997. Growing degree-days: one equation, two interpretations. Agric. For. Meteorol. 87, 291–300. https://doi.org/10.1016/S0168-1923(97)00027-0
Moretti, S., Marucci, A., 2019. A Photovoltaic Greenhouse with Passive Variation in Shading by Fixed Horizontal PV Panels. Energies 2019, Vol. 12, Page 3269 12, 3269. https://doi.org/10.3390/EN12173269
Paredes, P., López-Urrea, R., Martínez-Romero, Á., Petry, M.T., Cameira, M. do R., Montoya, F., Almeida, W., Salman, M., Pereira, L.S., 2025. Base and upper temperature thresholds to support the calculation of growing degree days aiming at their use with the FAO56rev crop coefficients curve: A review. Agric. Water Manag. 319, 109755. https://doi.org/10.1016/J.AGWAT.2025.109755
Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., 2023a. Reference Manual for AquaCrop version 7.1 – Chapter 1.
Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., 2023b. Reference Manual for AquaCrop version 7.1 – Annexes.
Salman, M., García-Vila, M., Fereres, E., Raes, D., Steduto, P., 2021. The AquaCrop model – Enhancing crop water productivity. FAO Water Report No. 47. https://doi.org/10.4060/CB7392EN
Saxton, K.E., Rawls, W.J., 2006. Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. Soil Science Society of America Journal 70, 1569–1578. https://doi.org/10.2136/SSSAJ2005.0117;CTYPE:STRING:JOURNAL
Smith, M., Steduto, P., 2012. Yield response to water : the original FAO water Production Function. Crop yield response to water 06–12.
Software | AquaCrop | FAO [WWW Document], n.d. URL https://www.fao.org/aquacrop/software/en (accessed 12.9.25).
Soil Atlas of Lazio - geoportale.regione.lazio.it [WWW Document], n.d. URL https://geoportale.regione.lazio.it/maps/new?layer=geonode:arsial_ssp_250K_suoli_lazio&view=True#/ (accessed 1.15.26).
Soil Science Division Staff, Ditzler, C., Scheffe, K., Monger, H.C., (eds.), 2017. Soil Survey Manual. USDA Handbook 18.
Steduto, P., Hsiao, T.C., Fereres, E., 2007. On the conservative behavior of biomass water productivity. Irrigation Science 2007 25:3 25, 189–207. https://doi.org/10.1007/S00271-

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

007-0064-1
Steduto, P., Hsiao, T.C., Fereres, E., Raes, D., 2012. Crop yield response to water. FAO paper 66. Food and Agriculture Organization of the United Nations, Rome.
Steduto, P., Hsiao, T.C., Raes, D., Fereres, E., 2009. Aquacrop-the FAO crop model to simulate yield response to water: I. concepts and underlying principles. Agron. J. 101, 426–437. https://doi.org/10.2134/AGRONJ2008.0139S;CTYPE:STRING:JOURNAL
Stevens, B., Diels, J., Vanuytrecht, E., Brown, A., Bayo, S., Rujweka, A., Richard, E., Ndakidemi, P.A., Swennen, R., 2020. Canopy cover evolution, diurnal patterns and leaf area index relationships in a Mchare and Cavendish banana cultivar under different soil moisture regimes. Sci. Hortic. 272, 109328. https://doi.org/10.1016/J.SCIENTA.2020.109328
Thimijan, R.W., Heins, R.D., 1983. Photometric, Radiometric, and Quantum Light Units of Measure: A Review of Procedures for Interconversion. HortScience 18, 818–822. https://doi.org/10.21273/HORTSCI.18.6.818
Thornley, J.H.M., Johnson, I.R., 1990. Plant and crop modelling. Clarendon Oxford.
Watson, D.J., 1947. Comparative Physiological Studies on the Growth of Field Crops: I. Variation in Net Assimilation Rate and Leaf Area between Species and Varieties, and within and between Years. Ann. Bot. 11, 41–76. https://doi.org/10.1093/OXFORDJOURNALS.AOB.A083148
Christopher, J. (2011). Multi-state models for panel data: the msm package for R. Journal of statistical software, 28.
Ding, X. a. (2019). Model simulation of cucumber yield and microclimate analysis in a semi-closed greenhouse in China. HortScience, 7.
Marcelis, L. a. (2005). Quantification of the growth response to light quantity of greenhouse grown crops. V International Symposium on Artificial Lighting in Horticulture 711, 7.
Matis, J. a. (1989). An application of the Markov chain approach to forecasting cotton yields from surveys. Agricultural Systems, 13.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

Chapter 6. Photovoltaic greenhouse digital twin

Introduction

This deliverable presents the development and validation of a digital twin system for the REGACE photovoltaic greenhouse, spanning activities conducted between month 18 and month 34 of the project. The work addresses focuses on creating predictive models capable of simulating the complex interactions between greenhouse microclimate, energy production, and crop performance under the influence of a responsive photovoltaic tracking system.
The core objective was to establish a data-driven framework that integrates environmental measurements, control signals, and operational data into predictive models serving multiple purposes. These include short-term predicting of internal conditions, scenario analysis for control optimization, and support for decision-making in greenhouse management. The digital twin operates as a virtual representation of the physical system, continuously updated with sensor measurements and capable of projecting future states under different operational scenarios.
The scope encompasses three distinct but interconnected modeling components, each addressing a critical aspect of agrivoltaic greenhouse performance.
The first component focuses on photovoltaic power production, predicting electrical output from the bifacial tracking panels under varying environmental conditions and control configurations (Aaslyng et al., 2005, 2003; Guo et al., 2020; Salah and Fourati, 2021). Power predicting enables economic optimization of the trade-off between crop light requirements and energy generation, supports grid integration planning, and provides baseline expectations for performance monitoring and anomaly detection.
The second component addresses the prediction of greenhouse internal microclimate, with particular emphasis on air temperature as the primary state variable. Temperature dynamics in agrivoltaic greenhouses represent a critical challenge due to the coupling between solar radiation modulation by the tracking system, thermal inertia of the greenhouse structure, ventilation dynamics, and control actions (Gharghory, 2020). Accurate prediction of internal temperature enables proactive thermal management, which is essential for maintaining optimal growing conditions while maximizing photovoltaic energy production.
The third component extends the digital twin framework toward crop-related variables, investigating whether growth dynamics, developmental progression, and yield formation can be predicted from environmental forcing and control history. Agricultural performance constitutes the ultimate validation criterion for any greenhouse system, and understanding how crops respond to the modified light environment created by photovoltaic shading, potentially compensated by CO_2 enrichment, represents a fundamental requirement for validating the REGACE concept (Belhaj Salah and Fourati, 2021; Bussab et al., 2007; Ferreira et al., 2002; Fourati and Chtourou, 2007; Nielsen, 2015; Petrakis et al., 2022a; Salah and Fourati, 2021).
The methodological approach evolved through systematic progression from tree-based ensemble methods to graph neural network architectures. Initial development employed Random Forest and XGBoost regressors, which provided interpretable baselines with strong nonlinear approximation capabilities and minimal computational requirements. These tree-based models established feasibility and quantified achievable accuracy using standard feature engineering, while their feature importance metrics revealed which variables most strongly influenced greenhouse response.
However, diagnostic analysis of tree-based predictions identified two fundamental limitations motivating architectural evolution. First, the models exhibited site-specificity, requiring

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

retraining when transferred to different greenhouse configurations or sensor networks. Second, performance degraded systematically during extreme thermal regimes, particularly above 35°C, suggesting that fixed feature transformations could not adequately capture regime-dependent dynamics.
These observations motivated transition to graph neural networks combined with temporal sequence modeling. This choice reflects the recognition that greenhouse systems exhibit strong spatial and temporal correlations among measured variables that tree-based methods, despite their strengths, treat as independent features. Physical processes such as radiative heating, convective exchange, and control responses create dependencies that standard approaches may not capture effectively. Graph-based representations provide an inductive bias aligned with the coupled nature of greenhouse physics, while remaining agnostic to explicit mechanistic equations that would be difficult to parameterize for the novel agrivoltaic configuration. The GNN architecture addresses generalization requirements through flexible topology adaptation to different sensor networks, while soft-gated Mixture of Experts components handle regime heterogeneity without the brittleness of discrete switching.
The crop modeling component revealed fundamental limitations related to data availability and measurement frequency. While environmental sensors provide continuous high-resolution observations, crop measurements derive from manual assessments conducted approximately every two weeks and subsequently interpolated. This temporal mismatch creates an identifiability problem where supervised learning algorithms cannot reliably extract growth dynamics from environmental drivers. The investigation provides quantitative evidence of these limitations and establishes minimum data requirements for future crop-aware digital twin implementations.
The microclimate digital twin achieves operational maturity and can support real-time applications. The crop analysis, though not yielding predictive capability under current conditions, delivers valuable insights into data infrastructure requirements and points toward hybrid modeling strategies combining mechanistic growth models with data assimilation techniques.
The remaining sections of this deliverable detail the system architecture, data management procedures, model development methodology, training and validation results, software implementation, and future development pathways. Technical descriptions are structured to enable reproducibility while remaining accessible to stakeholders with diverse backgrounds in agronomy, engineering, and energy systems.

6.1 System Architecture and Methodology

6.1.1 The Digital Twin Concept for agrivoltaics

A digital twin is a computational representation of a physical system that maintains a continuous correspondence with its real-world counterpart. This correspondence operates bidirectionally: the digital twin receives measurement data from sensors deployed in the physical system and, in return, provides predictions of unmeasured system states. The value of this approach lies in its ability to extend observability beyond what sensors directly measure, enabling operators to understand system behavior under conditions that have not yet occurred or cannot be directly instrumented.
In the context of the REGACE photovoltaic greenhouse, the digital twin serves as a predictive bridge connecting what we can measure with what we need to know for effective decision-making. The greenhouse environment generates hundreds of data streams from temperature sensors, radiation meters, control actuators, and electrical monitoring equipment. However,

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

operators and control systems require specific information that either cannot be measured directly at sufficient temporal resolution or represents future states needed for proactive management. The digital twin transforms available measurements into actionable predictions.

6.1.1.1 Core Functional Definition

The REGACE digital twin can be formally described through its input-output relationship:
Given: A set of measured environmental and operational variables at time t

- External forcing: ambient temperature, solar irradiance, wind speed, relative humidity
- Internal sensors: distributed temperature measurements, internal radiation levels
- Control states: CO_2 enrichment status
- System operation: actuator commands, ventilation rates

Produce: Predictions of target variables at time t or future time t+Δt

- Internal greenhouse air temperature
- Photovoltaic power production
- Crop development indicators

The relationship between inputs and outputs involves complex, nonlinear dynamics arising from radiative transfer, thermal mass effects, convective exchange, and biological responses. The digital twin learns these relationships from historical data, capturing patterns that would be difficult or impossible to specify through explicit equations.

6.1.1.2 Target Output 1: Internal Greenhouse Temperature

Internal air temperature represents the primary state variable governing both crop performance and control decisions. Temperature directly affects photosynthetic rates, respiration, transpiration, and developmental timing. At the same time, crops influence temperature by evapotranspiration. It also determines energy consumption for heating or cooling and influences the effectiveness of CO_2 enrichment. However, temperature in an agrivoltaic greenhouse exhibits particular complexity because the photovoltaic tracking system continuously modulates incident radiation and shading patterns.
Traditional greenhouse climate models assume fixed shading fractions or static panel configurations. The REGACE responsive tracker introduces time-varying optical properties that couple solar position, control logic, and instantaneous weather conditions. This coupling means that temperature response depends not only on current external forcing but also on recent history and anticipated control actions. The digital twin must therefore capture temporal dependencies extending over hours, accounting for thermal inertia in structure and growing media.
Accurate temperature prediction enables several operational capabilities:

- Proactive thermal management: Predicting temperature evolution 15 minutes to several hours ahead allows preventive control actions rather than reactive adjustments.
- Scenario evaluation: Operators can simulate the thermal impact of alternative tracker positions or ventilation strategies without physical trial.
- Anomaly detection: Deviation between predicted and observed temperature may indicate sensor failure, actuator malfunction, or unexpected system behavior requiring attention.

The target accuracy for temperature prediction was established based on operational requirements. Temperature variations of 1-2°C typically lie within acceptable ranges for most crops, while deviations exceeding 3-4°C may trigger stress responses. The digital twin therefore aims for mean absolute errors below 2°C across normal operating ranges, with particular attention to avoiding systematic bias that would undermine long-term predictions.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

6.1.1.3 Target Output 2: Photovoltaic Power Production

Electric power generation represents the second critical output, directly tied to the project's economic viability and energy optimization objectives. The responsive tracker adjusts panel orientation to balance crop light requirements against electricity production. Predicting power output under proposed control actions enables quantitative evaluation of this trade-off.
Power production in the REGACE system depends on multiple interacting factors:

- Solar irradiance reaching the panels, modified by greenhouse cover transmission
- Panel temperature, influenced by internal greenhouse climate
- Tracker angle relative to sun position
- Shading from structural elements or accumulation on panel surfaces
- Electrical characteristics of bifacial modules receiving light from both sides

The digital twin must account for these dependencies to provide reliable power forecasts. Unlike standalone photovoltaic installations where module temperature can be estimated from ambient conditions and irradiance, the greenhouse-integrated panels experience a modified thermal environment. Internal air temperature, humidity, and air movement patterns affect panel cooling and therefore electrical performance.
Power prediction serves three primary functions:

- Economic optimization: Comparing predicted power output under different tracker strategies against crop light requirements enables revenue-maximizing decisions.
- Grid integration: Predicting power production supports coordination with grid operators and optimization of energy storage or consumption timing.
- Performance monitoring: Systematic deviation between predicted and actual power may indicate panel soiling, electrical faults, or degradation requiring maintenance.

6.1.1.4 Target Output 3: Crop Development Indicators

The third output category addresses crop growth and production, representing the ultimate validation criterion for any greenhouse system. While temperature and power can be measured continuously, crop development traditionally requires periodic manual assessment. Plant height, leaf area, biomass accumulation, and yield progress slowly relative to environmental fluctuations, typically requiring days to weeks for observable change.
The digital twin approach to crop prediction investigates whether these slow variables can be inferred from fast environmental measurements. The underlying hypothesis recognizes that crop growth integrates environmental history: today's plant state reflects accumulated light interception, thermal time, water availability, and nutrient supply over preceding weeks. If this integration can be learned from data, the digital twin could provide daily or hourly updates of crop status without physical sampling.
This represents an ambitious objective facing fundamental challenges. Crop growth models span a spectrum from detailed mechanistic representations requiring dozens of physiological parameters to purely empirical correlations. The REGACE digital twin explores a middle path: using environmental measurements and control variables as inputs to data-driven models trained on observed crop metrics.
Several use cases motivate crop prediction capability:

- Growth stage monitoring: Tracking phenological progression informs harvest timing and market coordination.
- Light compensation assessment: Predicting whether CO_2 enrichment successfully offsets reduced light from photovoltaic shading requires relating environmental conditions to growth outcomes.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

- Yield predicting: Early prediction of final yield enables logistical planning and market commitments.

However, achieving reliable crop predictions from environmental data alone encounters significant practical constraints. Crop measurements in the REGACE demonstration sites occur approximately every two weeks through manual assessment. Between measurements, interpolation provides estimated values but introduces artificial smoothness that does not reflect actual growth dynamics. The digital twin development therefore includes honest assessment of what can and cannot be predicted given current data infrastructure, establishing requirements for future enhancement.

Integration and Operational Logic

These three outputs function together rather than independently. Temperature predictions inform power predicting through thermal effects on panel efficiency. Both temperature and power relate to crop performance through their joint influence on the growing environment. Control decisions affecting one output inevitably impact the others, creating a coupled optimization problem.

The digital twin architecture addresses this coupling through two complementary approaches. First, models may be trained independently for each output but evaluated jointly during scenario analysis. This modularity simplifies development and allows focused improvement of individual components. Second, shared input processing ensures that environmental measurements and control states receive consistent treatment across all predictive models, maintaining coherent representation of system state.

Temporal resolution also varies appropriately across outputs. Temperature predictions target sub-hourly horizons matching control system response times. Power forecasts may extend to daily or weekly aggregates supporting planning decisions. Crop predictions naturally operate at longer timescales reflecting developmental processes, though more frequent updates provide valuable trend information.

From Measurement to Prediction: The Digital Twin Value Proposition

The digital twin transforms what exists (sensor measurements, control settings, weather observations) into what matters for decisions (future temperatures, expected power, anticipated yields). This transformation provides value in proportion to the cost and difficulty of obtaining the predicted information through direct measurement.

Consider internal temperature. While sensors continuously record temperature at specific locations, predicting temperature at future times or unmeasured locations based on current conditions and planned control actions provides information that no static sensor network can deliver. The digital twin becomes a computational sensor generating synthetic measurements of system futures.

For power production, instantaneous measurements come from inverter monitoring. However, evaluating the power consequences of alternative tracker strategies requires either repeated physical trials or predictive modelling. The digital twin enables virtual trials, compressing weeks of field testing into minutes of computation.

Crop variables present the strongest case for digital augmentation. Manual assessment is labor intensive, destructive in some cases, and necessarily sparse in time. If environmental measurements correlating with growth can be continuously monitored and processed through learned relationships, the digital twin generates a continuous stream of crop state estimates from readily available data.

The conceptual foundation thus rests on learning mappings from abundant, automatically collected data (weather, sensors, controls) to sparse, expensive, or future-state information (crop status, optimal decisions, system futures). Success requires that these mappings exist,

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

remain stable, and can be discovered from available data. The following sections describe how these requirements were addressed in practice.

6.1.2 Overall Architecture

The digital twin architecture evolved through three distinct developmental stages, each addressing limitations identified in the previous approach while expanding scope and generalization capability. This progression reflects both growing understanding of greenhouse dynamics and refinement of modeling objectives from site-specific prediction toward broadly applicable methodology.

6.1.2.1 Stage 1: Baseline XGBoost Models for Temperature and Power

Initial development focused on establishing feasibility and performance baselines using gradient boosted decision trees implemented through the XGBoost framework. This choice reflected several practical considerations. XGBoost handles mixed data types, captures nonlinear relationships without explicit feature engineering, and provides interpretable feature importance metrics. The method has proven effective across numerous regression problems and requires modest computational resources for training and inference(Chen and Guestrin, 2016; Wang et al., 2020).
Two independent models were constructed:

- Temperature predictor: Using all available environmental measurements and control signals to predict internal greenhouse air temperature at a single representative location
- Power predictor: Using solar irradiance, tracker position, panel temperature estimates, and electrical state variables to predict photovoltaic power output

The input feature set for the temperature model included external meteorological variables (ambient temperature, wind speed, humidity, global horizontal irradiance), internal sensors (multiple temperature probes, internal radiation measurements). Temporal context was incorporated through lagged values and rolling statistics computed over windows ranging from 15 minutes to several hours.
Training utilized chronological train-test splits to prevent information leakage. The greenhouse exhibits seasonal patterns and gradual drifts that would be artificially suppressed by random splitting. Temporal validation ensures that performance metrics reflect genuine predicting capability rather than interpolation within the training period.
The baseline XGBoost models achieved encouraging results. Temperature predictions showed mean absolute errors around 2°C with good correlation to observations across most operating conditions. Power predictions captured diurnal cycles and weather-driven variability with comparable accuracy. These outcomes confirmed that supervised learning from operational data could effectively predict target variables without requiring explicit physical models.
However, diagnostic analysis revealed systematic limitations. The models occasionally exhibited poor performance during extreme conditions such as very high temperatures or rapid transients following control actions. Feature importance analysis showed heavy reliance on internal temperature sensors for temperature prediction, suggesting the models learned quasi-persistence rather than dynamic evolution. Most critically, the models were site-specific: training data from one greenhouse and crop type could not directly inform predictions for another configuration.

6.1.2.2 Stage 2: Low-Cost Sensor Configuration Models

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

The second developmental stage addressed a practical constraint limiting digital twin deployment. While research greenhouses may support dense sensor networks, commercial installations face economic pressure to minimize instrumentation costs. A digital twin requiring extensive internal sensing would limit adoption regardless of predictive accuracy.
This motivated investigation of reduced-input models, termed "low-cost" configurations. The objective was to predict the same target variables (internal temperature, power output) using only measurements available from external weather stations and minimal greenhouse-specific instrumentation. The hypothesis held that much of the information content in dense internal sensors represents redundancy or slow spatial variation that could be predicted from external forcing and system configuration knowledge.
The low-cost sensor set was defined as:

- External temperature: Ambient air temperature from weather station or external probe
- Solar radiation: Global horizontal irradiance or equivalent radiation measurement
- Wind speed: External anemometer reading
- Relative humidity: External humidity sensor

This configuration eliminates most internal temperature sensors, all distributed humidity measurements, and detailed control state monitoring beyond what can be inferred from external conditions. The total cost of this instrumentation is approximately one-fifth that of the full sensor network used in baseline models.
Model architecture remained based on XGBoost but incorporated several methodological enhancements to compensate for reduced information. Feature engineering became more sophisticated, constructing derived variables representing physical processes that internal sensors would otherwise directly measure. These included:

- Thermal lag features: Exponentially weighted moving averages of external temperature at multiple time constants, approximating structural thermal mass
- Radiation integrals: Cumulative solar energy input over rolling windows, representing heat accumulation
- Control regime indicators: Discrete labels inferred from time-of-day, season, and external conditions, proxying for unmeasured ventilation states

Additionally, the model training procedure incorporated explicit regularization to prevent overfitting to training-period correlations that might not generalize across seasons or years.
Performance of low-cost models showed modest degradation relative to full-input baselines. Mean absolute temperature errors increased by approximately 0.5°C, while power predictions remained nearly equivalent. This trade-off between accuracy and measurement economy is favorable for many applications, particularly when the primary value lies in detecting anomalies or comparing scenarios rather than absolute prediction.
The low-cost approach demonstrated that strategic feature engineering could partially compensate for sensor reduction. However, it also revealed fundamental limits: certain system states cannot be reliably inferred from external observations alone. Rapid internal transients following control actions, spatial temperature gradients, and response to local shading remain difficult to predict without internal measurements. This motivated the third architectural evolution.

6.1.2.3 Stage 3: Graph Neural Network Architecture for Generalization

The transition to graph neural networks addressed the most significant limitation of previous approaches: their specificity to individual greenhouse configurations. Each XGBoost model was trained on data from a particular site with specific structural characteristics, sensor

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

placement, and crop type. Applying such models to new greenhouses required retraining, limiting their utility for the broader REGACE objective of developing transferable agrivoltaic technology.

Deep learning offers potential for generalization through abstraction. Rather than learning direct mappings from specific sensor values to specific predictions, neural networks can potentially learn underlying physical relationships that apply across configurations. However, standard neural network architectures assume fixed input structure, which conflicts with variability in sensor networks across sites.

Graph neural networks resolve this tension by representing the system as a graph where nodes correspond to measured variables and edges encode relationships (Gharghory, 2020; Jin et al., 2024; Petrakis et al., 2022a; Sun et al., 2022). This formulation provides several advantages:

- Flexible topology**:** Different greenhouses with different sensor configurations can be represented as different graphs
- Explicit coupling**:** Edges between nodes allow the model to learn which variables influence each other, capturing the coupled dynamics of greenhouse systems
- Spatial and temporal integration**:** Graph convolution operators propagate information both across the sensor network (spatial) and through time (via recurrent units)

The GNN architecture developed for the REGACE digital twin combines temporal encoding through Long Short-Term Memory (LSTM) units with spatial encoding through attention-based graph convolution (Cheung and Lai, 1995; Hongkang et al., 2018; Jain and Medsker, 1999; Nketiah et al., 2023; Rodriguez et al., 1999). Each node (representing a sensor or target variable) maintains a hidden state that evolves based on its temporal history and messages received from connected nodes. Attention mechanisms weight the importance of different connections dynamically, allowing the model to emphasize relevant relationships depending on current conditions.

Graph construction proceeded through data-driven correlation analysis. Variables showing strong statistical dependence in training data receive edges connecting their nodes. This approach avoids requiring explicit specification of physical coupling while leveraging structure inherent in the data. Self-loops ensure that each variable's history contributes to its future prediction.

Training data aggregated measurements from multiple REGACE demonstration sites, exposing the model to diverse greenhouse types, climates, and crops. This multi-site training enables the learned representations to capture commonalities rather than site-specific peculiarities. The resulting model can be applied to new greenhouse configurations by constructing an appropriate graph representation and fine-tuning if necessary.

The GNN architecture also addressed limitations observed in earlier models regarding extreme conditions. Dedicated model components ("experts") within the GNN architecture used a Mixture of Experts approach. Different sub-networks specialized in predicting behavior under different operating regimes (moderate temperatures, high temperatures with strong radiation, transient response periods). A gating mechanism learned to weight contributions from each expert based on current system state, enabling smooth transitions between dynamical behaviors without hard regime boundaries.

This approach proved particularly effective for handling the upper temperature range where earlier models showed systematic underestimation. High temperatures in greenhouses involve nonlinear feedback between radiation absorption, ventilation effectiveness, and thermal stress responses that differ qualitatively from moderate conditions. The expert architecture allowed dedicated capacity for learning these extreme-regime dynamics while maintaining accuracy

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

across normal operations (Baglivo et al., 2020; Brækken et al., 2023; Bussab et al., 2007; Manonmani et al., 2018; Petrakis et al., 2022b).

6.1.2.4 Architectural Evolution Summary

The three-stage progression can be understood as movement along two axes: generality and sophistication. XGBoost models provided site-specific predictions with good accuracy but limited transferability. Low-cost models traded some accuracy for deployment flexibility through sensor reduction. GNN models addressed generalization across greenhouse configurations while incorporating domain knowledge about coupled dynamics and operational regimes.

Each stage built upon insights from the previous approach. XGBoost established baseline performance and revealed feature importance patterns. Low-cost models demonstrated the value of physics-informed feature engineering. GNN models integrated these lessons into an architecture supporting transfer learning and explicit representation of system coupling.

The final digital twin employs the GNN architecture as its primary prediction engine, with XGBoost models retained for specific applications where their computational efficiency or interpretability provides advantages.

These two approaches operate independently. The modeling strategy is "hybrid" in the sense that multiple model types coexist within the deployment ecosystem, each applied to the use case where it performs best. The system architecture remains modular at the software level: models have standardized interfaces for data input and prediction output. This design allows individual models to be updated, retrained, or replaced without affecting others. For example, if a superior GNN architecture emerges, it can substitute the current implementation without modifying the XGBoost power prediction component. Similarly, new model types (mechanistic crop models, ensemble methods, physics-informed neural networks) can be integrated as additional components addressing specific prediction tasks. The modular software design provides flexibility to evolve the digital twin as methodology advances and operational needs change.

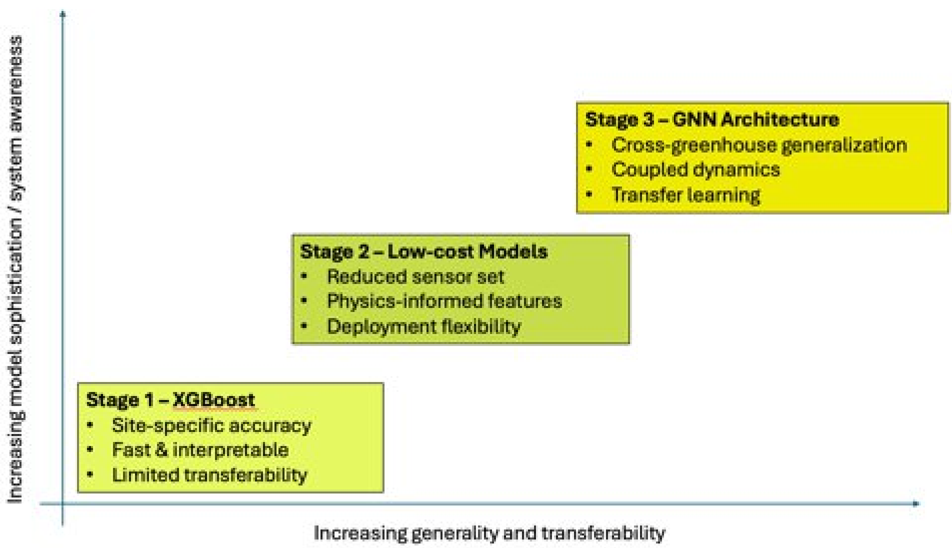


Fig. 6.3 Scheme of the methodology architecture.

6.1.3 Methodological Approach

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

The digital twin development followed a systematic workflow progressing from raw sensor measurements to validated predictive models. This section describes the overall methodological pipeline, emphasizing design choices that address the specific challenges of agrivoltaic greenhouse modeling. The approach balances rigor in model validation against practical constraints of working with operational systems generating heterogeneous data streams.

6.1.3.1 Data collection and aggregation phase

The primary dataset for digital twin development originated from the Circeo Greenhouse located at Fattoria Solidale del Circeo. This installation was selected as the principal data source for model training and validation due to several distinguishing characteristics. First, the sensor network design and deployment were specifically engineered by the University of Tor Vergata, ensuring optimal placement, measurement redundancy, and compatibility with research objectives. Second, the Circeo site provided the most complete and accurate data streams among REGACE demonstration installations, with consistent temporal coverage, minimal sensor failures, and comprehensive instrumentation spanning environmental forcing, internal microclimate, control states, and photovoltaic performance. The systematic sensor configuration at this site enabled robust feature engineering and validation of modeling approaches subsequently tested for generalization to other REGACE locations with more heterogeneous measurement infrastructure.

Raw data streams originated in diverse formats reflecting the heterogeneity of commercial greenhouse control systems, weather stations, and research instrumentation. NetCDF files provided structured storage for continuous time series from dedicated research sensors. Inverter data arrived as CSV exports from manufacturer-specific monitoring platforms. Manual crop assessments entered spreadsheets with date stamps and text annotations. This format diversity required flexible parsing infrastructure capable of harmonizing disparate sources onto common temporal and variable naming conventions.

6.1.3.2 Preprocessing and Quality Control

Raw sensor data inevitably contains artifacts requiring correction before use in model training. These artifacts arise from communication failures, sensor faults, calibration drift, and temporary disconnections during maintenance. Preprocessing transformed potentially corrupted raw streams into clean datasets suitable for supervised learning.

Quality control implemented a multi-stage filtering pipeline. Initial screening flagged physically impossible values: temperatures below absolute zero or above credible maxima, negative solar irradiance, humidity outside the 0-100% range. These violations indicated transmission errors or sensor malfunction and were marked invalid. Secondary screening applied rate-of-change limits appropriate to each physical variable. Temperature, for instance, cannot physically change by more than several degrees per minute in greenhouse environments. Violations suggested timestamp errors or sensor discontinuities.

A critical discovery during preprocessing concerned sensor fault modes that escaped simple range checking. Certain temperature sensors exhibited "flat-lining" behavior where measurements remained nearly constant for hours despite changing environmental conditions. These faults occurred preferentially in a narrow temperature band around 10-15°C, suggesting either physical damage causing thermal isolation from the measured environment or electronic failure producing constant output. Standard outlier detection based on statistical thresholds failed to identify these faults because the reported values remained within plausible ranges.

Addressing flat-line faults required developing rolling variance detection. Windows spanning 30-60 minutes were evaluated for measurement variability. Segments with standard deviation

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

below empirically determined thresholds, combined with values in the suspicious temperature band, were flagged as potentially invalid. This heuristic proved essential for preventing models from training on corrupted labels that would otherwise appear legitimate.

Missing data received treatment appropriate to the presumed mechanism of missingness. Short gaps of a few minutes, likely caused by temporary communication issues, were filled using forward-then-backward filling. This approach assumes system state persistence over short intervals, which holds for most greenhouse variables. Longer gaps exceeding one hour were left unfilled, and corresponding time periods excluded from training. Attempting to interpolate across substantial gaps risks introducing artificial patterns unrepresentative of actual system dynamics.

Timestamp alignment across data sources required careful handling of time zone conventions, daylight saving transitions, and sampling jitter. All timestamps were normalized to UTC and rounded to common intervals (typically 5 minutes) suitable for the slowest-sampled variables needed in the model. This discretization introduces minimal information loss while greatly simplifying temporal joins between tables.

6.1.3.3 Feature Engineering and Representation

Transforming cleaned sensor measurements into model inputs involved constructing features encoding relevant physical processes and temporal context. The feature engineering strategy differed across the three architectural stages but followed common principles.

Temporal features captured system memory through lagged values and rolling statistics. Temperature at time t depends not only on current forcing but also on recent history reflecting thermal inertia. Features therefore included measurements at t-15min, t-30min, t-1hr, and t-2hr for key environmental variables. Rolling means and standard deviations over windows from 15 minutes to 4 hours characterized both level and variability of forcing.

Physics-inspired features translated domain knowledge into computable quantities. Growing Degree Days accumulated effective temperature above crop-specific base thresholds, providing a unified representation of thermal forcing relevant to crop development. Integrated solar radiation over daily windows represented accumulated energy input driving both temperature rise and photosynthesis. Vapor pressure deficit combined temperature and humidity into a single variable governing transpiration demand.

The GNN architecture necessitated different feature preparation. Rather than constructing an explicit feature vector, measurements were organized as node attributes within the graph representation. Each node corresponded to one measured variable, and its feature vector contained temporal context (history over the preceding sequence length) for that specific variable. This node-centric organization allowed the graph convolution operations to learn appropriate combinations of related variables rather than requiring manual specification.

6.1.3.4 Model Training Strategy

Training procedures followed established best practices for time series prediction while incorporating domain-specific adaptations. The fundamental principle held that temporal dependencies must be respected: models train only on past information to predict future states, never using future data to predict earlier times.

Dataset splitting employed strictly chronological boundaries. Training sets encompassed early time periods, validation sets covered subsequent intermediate periods, and test sets reserved the most recent data. Typical splits allocated 60% of data to training, 20% to validation, and 20% to test, though exact proportions varied by site and data availability (Alaimo and Seri, 2021; Biau and Scornet, 2016; Bussab et al., 2007; Dae-Hyun et al., 2020; Fourati and Chtourou, 2007; James et al., 2013; Petrakis et al., 2022b). This temporal holdout ensures that

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

reported performance metrics reflect genuine predicting skill rather than memorization or interpolation.
Hyperparameter optimization proceeded through grid search over candidate values for learning rates, regularization strengths, and architecture-specific parameters such as hidden layer dimensions or attention head counts. Validation set performance guided this search, with test set evaluation reserved for final model assessment to avoid overfitting to validation data.
Training employed mini-batch stochastic gradient descent or variants (Adam optimizer for neural networks). Batch sizes balanced computational efficiency against gradient estimation accuracy, typically ranging from 32 to 256 samples. Learning rate schedules incorporated warm-up periods and decay as training progressed, following strategies proven effective for similar sequence prediction tasks.
Loss function selection balanced multiple objectives. Mean squared error provided baseline regression loss but proved insufficient for capturing distributional properties. For temperature prediction, Huber loss offered robustness to occasional outliers while maintaining gradient-based training. Additional loss terms enforced specific behaviors: tail-aware penalties increased weight on high-temperature predictions to address systematic underestimation, while custom terms penalized physically implausible predictions such as temperatures outside reasonable bounds.
Regularization prevented overfitting through multiple mechanisms. Dropout randomly deactivated neural network units during training, forcing distributed representations. Weight decay penalized large parameter values, encouraging simpler models. Early stopping terminated training when validation loss ceased improving, avoiding memorization of training set noise.
The Mixture of Experts architecture within the GNN introduced additional training considerations. Expert networks and gating mechanisms required joint optimization through a unified loss function combining prediction accuracy with load balancing terms encouraging diverse expert utilization. This joint training proved more stable than alternating optimization of experts and gates.

6.1.3.5 Validation and Performance Assessment

Model validation employed multiple complementary metrics and diagnostic analyses. No single performance measure adequately characterizes predictive capability across all operating conditions and use cases. Comprehensive assessment required examining both aggregate statistics and conditional performance in specific regimes.
Primary regression metrics included mean absolute error (MAE), and coefficient of determination (R^2). MAE provides interpretable units matching the predicted variable. R^2 quantifies variance explained relative to naive constant prediction. Reporting all three metrics provides balanced perspective on model quality.
However, aggregate metrics can mask systematic biases. Diagnostic scatter plots comparing predicted versus observed values reveal patterns invisible in summary statistics. Ideal predictions lie on the identity line; systematic deviations indicate bias or heteroscedasticity. Coloring points by operational regime (temperature range, time of day, season) exposes conditional performance differences.
Temporal validation examined prediction accuracy as a function of forecast horizon. One-step-ahead predictions typically achieved highest accuracy, with performance degrading for longer horizons. Understanding this degradation informs appropriate model application: short-term forecasts for immediate control, longer-term predictions for planning with appropriate uncertainty quantification.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

Regime-specific analysis decomposed performance across operational conditions. Temperature predictions were separately evaluated for moderate conditions (15-30°C), warm conditions (30-35°C), and extreme heat (>35°C). Power predictions were stratified by radiation level and time of day. This decomposition revealed whether models performed uniformly or struggled in specific regions of state space.

The crop prediction component required alternative validation strategies due to sparse ground truth. Regression metrics computed on interpolated crop measurements would be misleading because the interpolation itself introduces artificial structure. Validation therefore focused on assessing whether predicted trends aligned with measurement-to-measurement changes and whether classification of coarse growth stages (early, middle, late season) exceeded baseline accuracy.

6.1.3.6 Iterative Refinement and Diagnostic-Driven Development

Model development proceeded iteratively, with each training cycle informing subsequent methodological adjustments. This diagnostic-driven approach treated poor performance not as failure but as information revealing inadequate representation of system dynamics.

Early GNN models, for instance, systematically compressed extreme temperature predictions. Diagnostic analysis colored predictions by realized temperature and revealed that all points above 40°C were underpredicted. This pattern motivated introducing tail-aware loss terms and high-temperature expert networks, directly addressing the identified limitation.

Similarly, discovering the flat-line sensor fault mode emerged from careful examination of residual patterns. Vertical stripes in predicted-versus-observed scatter plots indicated regions where predictions varied while observations remained constant. Investigating the temperature ranges where this occurred led to the rolling variance detection method, which when implemented substantially improved model stability.

This iterative cycle, train, diagnose, adjust, retrain, continued throughout development. Each iteration targeted specific weaknesses revealed by comprehensive diagnostic analysis rather than pursuing generic improvements. The methodology thus evolved in response to actual system behavior rather than following predetermined algorithms.

6.2 Data Description and Management

The digital twin development relied on multivariate time series datasets collected across REGACE demonstration sites during the testing phase. Comprehensive description of sensor specifications, measurement protocols, and data collection procedures appears in Deliverable D3.2 and is summarized in Chapter 1. This section addresses the technical infrastructure for data preprocessing, synchronization, and storage that enabled efficient model training and validation.

6.2.1 Data Pre-processing and Quality Control

6.2.1.1 Cleaning Procedures

Raw sensor streams underwent systematic cleaning to remove artifacts and invalid measurements before model training. The cleaning pipeline implemented three sequential filters:

1. Physical validity screening**:** Values violating physical constraints (temperature < -50°C or > 80°C, solar irradiance < 0 W/m², relative humidity outside 0-100%) were flagged as

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

transmission errors and marked invalid. Missing value indicators (e.g., -999 sentinels) were explicitly detected and converted to null entries.

2. Rate-of-change filtering**:** Measurements exhibiting physically impossible temporal derivatives were identified as timestamp errors or sensor discontinuities. Temperature changes exceeding 5°C per minute, for instance, indicated data corruption rather than actual system dynamics. These segments were excluded from analysis.
3. Flat-line detection**:** Sensors occasionally exhibited fault modes where output remained nearly constant despite changing environmental conditions. A rolling standard deviation filter computed over 30-minute windows identified segments with variance below 0.1°C in the 10-15°C band, flagging likely sensor failures that escaped range-based detection.

Duplicate timestamps arising from logging system restarts were removed by retaining the first occurrence. Entirely empty rows lacking any valid measurements were dropped from the dataset.

6.2.1.2 Synchronization Strategy

Temporal alignment across heterogeneous data sources required reconciling different sampling frequencies and timestamp conventions. The synchronization procedure followed these steps:

- Timestamp normalization: All measurements were converted to UTC timezone, eliminating daylight saving artifacts. Timestamps were rounded to the nearest 5-minute interval, matching the coarsest sampling frequency among critical variables.
- Nearest-neighbor temporal joins: Variables recorded at different frequencies were aligned using pandas merge_asof with 2.5-minute tolerance. This approach assigns each target timestamp the most recent available measurement from each source, preserving causality while tolerating minor sampling jitter.
- Resampling and aggregation: High-frequency measurements (1-minute resolution) were downsampled to the common 5-minute grid through mean aggregation for continuous variables and mode selection for discrete control states. This reduced data volume while preserving essential temporal dynamics.

The synchronized dataset maintained temporal ordering strictly, ensuring that all input features at time t represented information available before predicting targets at t or later times.

6.2.1.3 Storage Format Optimization

Initial data delivery in Excel spreadsheets and CSV files proved inefficient for the scale of multivariate time series spanning months. The storage infrastructure transitioned through progressive optimization:

1. Structured binary format adoption: Data were converted to NetCDF4 format exploiting compression and chunking. NetCDF provides self-describing metadata, dimensional consistency checks, and efficient partial loading of large datasets. Chunk sizes were tuned to match typical access patterns (loading weeks of data for specific variable subsets).
2. Hierarchical organization: Separate NetCDF files organized data by demonstration site and temporal resolution. High-frequency raw data, daily aggregates, and derived features occupied distinct files linked through consistent variable naming and coordinate systems.
3. Indexed access infrastructure: Pandas DataFrame representations cached in HDF5 format enabled rapid repeated access during model training iterations. Column-oriented storage optimized loading of feature subsets without reading entire datasets.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

These optimizations reduced typical data loading times from several minutes (CSV parsing) to under 10 seconds (binary formats with indexing), substantially accelerating the model development cycle. Total storage requirements decreased by approximately 60% through lossless compression while maintaining full numerical precision.

6.3 Energy Optimization Model

6.3.1 Objective

The photovoltaic power prediction component addresses a specific operational requirement within the REGACE digital twin framework: estimating and predicting inverter power output from a minimal set of environmental and control signals. This capability supports decision-making strategies that maximize energy production while maintaining consistency with greenhouse operational constraints, including thermal management requirements and crop light needs. The variables, the feature and the targets are listed in Annex 1.
The PV power model serves two complementary functions in the digital twin ecosystem:

- It provides a baseline expectation for power generation against which actual measurements can be compared, enabling detection of anomalies such as inverter downtime, sensor failures, or degraded panel performance.
- It enables scenario analysis where operators evaluate the energy yield consequences of alternative control strategies, such as modified tracker positions or ventilation schedules, before implementing changes in the physical system.

From an operational perspective, the model must balance accuracy against deployment practicality. Greenhouses cannot economically support extensive electrical instrumentation dedicated solely to power prediction. The approach therefore targets predictions from readily available environmental measurements (irradiance, temperature, basic control states) rather than requiring panel-level monitoring or specialized electrical sensors. This "low-cost" philosophy aligns with the broader REGACE objective of developing commercially viable agrivoltaic technology.

6.3.2 Model Description (XGBoost Regressor)

The PV power component implements a gradient boosted decision tree regressor using the XGBoost framework. This choice reflects practical considerations for operational deployment: XGBoost handles heterogeneous input types naturally, captures complex nonlinear relationships without manual specification, provides feature importance metrics supporting model interpretation, and executes efficiently in production environments.

<u>6.3.2.1 Target Variable and Physical Constraints</u>

The model predicts instantaneous inverter power output measured in watts. A critical implementation detail enforces physical consistency: predictions are clipped to non-negative values through a simple post-processing step. Inverter power cannot be negative in the REGACE configuration, and this constraint prevents unphysical outputs that occasionally emerge from unconstrained regression during low-signal periods such as dawn or dusk.

<u>6.3.2.2 Feature Engineering Strategy</u>

Model performance depends critically on feature construction. Raw environmental measurements alone provide insufficient information for accurate power prediction because

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

inverter output reflects not only instantaneous conditions but also recent system history, control inertia, and nonlinear interactions. The feature set therefore incorporates multiple representation strategies.

Environmental inputs form the foundation: global horizontal irradiance (or equivalent radiation proxy), external temperature, and available control state indicators. These variables capture the primary external forcing determining power generation potential. However, their raw values omit temporal context essential for modeling system dynamics.

Temporal lag features address this limitation by including recent values of key predictors. Lags at 5-minute intervals (matching typical sampling resolution) represent short-term persistence and inverter controller response characteristics. Autoregressive lags of the target variable itself provide additional predictive power, capturing smooth temporal evolution that environmental predictors alone cannot fully explain. Multiple lag horizons spanning 5 to 30 minutes encode memory at scales relevant to cloud passage events and thermal transients.

Delta features explicitly represent rates of change. Computing differences between current and lagged irradiance values, for instance, provides the model with direct information about whether conditions are improving, degrading, or stable. These gradient features prove particularly valuable for predicting power response during rapid transitions such as cloud edges or sudden changes in tracker position.

Rolling statistics characterize recent variability and persistence. Mean and standard deviation computed over approximately one-hour windows (12 samples at 5-minute resolution) summarize the prevailing weather regime. High rolling standard deviation of irradiance, for example, indicates intermittent cloud cover, which affects inverter behavior differently than stable clear or overcast conditions.

Physics-inspired features incorporate known dependencies. Interaction terms such as irradiance multiplied by temperature approximate temperature-dependent efficiency losses. Quadratic irradiance terms capture nonlinearity in the power-to-radiation relationship at low light levels. Cyclical encoding of hour through sine and cosine transforms represents diurnal patterns without imposing artificial discontinuities at midnight. Month indicators capture seasonal efficiency variations related to sun angle and ambient conditions.

This feature engineering approach deliberately favors interpretability and physical grounding over black-box complexity. Each feature corresponds to a measurable quantity or physically motivated transformation, facilitating diagnosis when predictions deviate from observations.

6.3.2.3 Data Preprocessing

Before feature construction, raw sensor streams undergo cleaning procedures described in Section 3.1. For the power modeling task, additional preprocessing removes non-informative data points. Nighttime periods where target power remains below 10 watts are excluded from training because they contain no useful information about the irradiance-to-power mapping and would bias the model toward trivial near-zero predictions. Known inverter downtime intervals, identified from maintenance logs or operator reports, are similarly removed to prevent learning spurious correlations during periods when the physical system was not operating normally.

6.3.3 Training and Validation

6.3.3.1 Chronological Dataset Splitting

Training and evaluation procedures respect temporal dependencies inherent in greenhouse and PV operation. The dataset is split chronologically, with the first 80% of data allocated to training and the final 20% reserved for testing. This temporal holdout strategy ensures that

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

performance metrics reflect genuine predicting capability rather than interpolation within the training period. Random splitting would artificially inflate performance by allowing the model to exploit future information when predicting past values.
For hyperparameter optimization, the training set undergoes time-series cross-validation using successive folds that preserve temporal ordering. Each fold trains on earlier data and validates on immediately following periods, mimicking the sequential nature of operational predicting.

6.3.3.2 Hyperparameter Optimization

Model configuration was optimized through randomized search across key XGBoost hyperparameters. The search space included number of estimators (model capacity), learning rate (boosting step size), maximum tree depth (complexity), and subsampling ratios for both observations and features (regularization). Randomized search provided computational efficiency compared to exhaustive grid search while exploring parameter regions likely to yield good performance.
The optimization procedure employed time-series cross-validation with three folds, preventing information leakage during validation. After identifying promising parameter combinations, final training incorporated early stopping with a patience of 50 rounds. Training terminates when validation performance ceases improving, automatically preventing overfitting to training data while maximizing utilization of available samples.

6.3.3.2 Anomaly Detection and Handling

Operational PV datasets invariably contain periods unrepresentative of the physical irradiance-to-power relationship. Two classes of anomalies required explicit treatment:

1. known downtime intervals were removed based on maintenance records. One documented period spanning September 19-24, 2025 exhibited power outputs inconsistent with environmental conditions, indicating inverter curtailment or malfunction. Including such periods would teach the model to predict low power under high irradiance, corrupting the learned mapping. The temporal exclusion approach removes these samples prior to any model training.
2. Second, low-signal filtering addresses the problem of nighttime and twilight periods. Points where power output remains below 10 watts carry minimal information about operational performance yet vastly outnumber high-production periods if included. Training on unfiltered data would therefore optimize predictions predominantly for the regime least relevant to energy yield assessment. The threshold-based filter focuses learning on periods where the PV system actively generates power and where prediction errors matter for operational decisions.

6.3.4 Results and Performance

6.3.4.1 Performance Metrics

The optimized XGBoost model achieved the following performance on the chronologically held-out test set:

- Mean Absolute Error (MAE): 72.68 W
- Root Mean Squared Error (RMSE): 134.78 W
- Coefficient of Determination (R^2): 0.970

These metrics indicate that the model explains approximately 95% of variance in inverter power output. The MAE of 72.68 W represents roughly 3% of the observed maximum power

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

output for the installation, falling within acceptable bounds for operational predicting. RMSE approximately double the MAE reflects occasional larger errors during rapidly changing conditions, which is expected given the difficulty of predicting transient responses to cloud passages or control actions.

6.3.4.2 Comparative Analysis Across Model Variants

The photovoltaic power prediction model development proceeded through five sequential experiments, each targeting a specific limitation identified in the previous iteration. This systematic progression reveals the relative importance of different modeling choices and demonstrates that methodological rigor in data handling and feature construction outweighs algorithmic sophistication.

Experiment E1 established the baseline using a Random Forest regressor trained on raw sensor measurements without preprocessing or feature engineering. This configuration included all time periods (day and night) and used environmental variables in their unprocessed form, ambient temperature, solar irradiance, wind speed, and basic control states entered the model directly as recorded. The resulting performance was poor (MAE 260.45 W, $R^2 = 0.259$), but this baseline served its purpose: quantifying prediction accuracy when relying solely on algorithm capability without domain knowledge integration.

Experiment E2 applied the first critical insight: not all data points contain equal information about operational power generation. Nighttime and twilight periods, where panels produce negligible output, vastly outnumber productive daylight hours in raw datasets. Training on this unbalanced distribution forces the model to optimize predominantly for predicting near-zero values, operationally irrelevant but statistically dominant. E2 implemented daytime filtering, excluding periods where power output remained below 10 watts. This single preprocessing step reduced MAE to 190.86 W and increased R^2 to 0.827, demonstrating that intelligent data selection delivers greater performance gains than algorithm choice. The model now focused learning capacity on the regime that matters: active power generation under varying solar and atmospheric conditions.

Experiment E3 introduced systematic feature engineering while maintaining the Random Forest architecture. Raw sensor values capture instantaneous conditions but omit temporal context essential for modeling system dynamics. The engineered feature set incorporated temporal lags at 5-minute intervals representing short-term persistence and inverter response characteristics, delta features explicitly encoding rates of change in irradiance and temperature, rolling statistics summarizing recent variability over approximately one-hour windows, and physics-inspired interaction terms such as temperature-dependent efficiency losses. These transformations provided the model with explicit information about system memory, environmental trends, and known physical relationships. Performance improved substantially to $R^2 = 0.942$ while MAE remained around 88,26 W, confirming that thoughtful feature construction enables even modest algorithms to capture complex dynamics. The minimal change in MAE between E2 and E3 despite consistent R^2 improvement reflects that feature engineering primarily benefited prediction during variable conditions rather than clear-sky periods that dominated E2's performance envelope.

Experiment E4 substituted XGBoost for Random Forest while maintaining the engineered feature set from E3. XGBoost's gradient boosting framework with regularization theoretically offers advantages over Random Forest's simple averaging of decorrelated trees, including sequential error correction, built-in handling of feature interactions, and adaptive learning rates. Empirically, this algorithmic upgrade reduced MAE to 103.32 W while maintaining R^2 near 0.940. The improvement is meaningful but modest compared to prior steps, illustrating that algorithm selection provides secondary refinement after establishing proper data handling

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

and representation. XGBoost's greatest contribution came through better handling of rare events and transient conditions where its boosting mechanism could focus capacity on difficult-to-predict samples.

Experiment E5 completed the development cycle through hyperparameter optimization and early stopping regularization. Randomized search explored XGBoost's configuration space including number of estimators, learning rate, maximum tree depth, and subsampling ratios. Time-series cross-validation ensured that hyperparameter selection reflected genuine predicting capability rather than overfitting to training patterns. Early stopping with 50-round patience prevented memorization of training noise by terminating when validation performance plateaued. These refinements yielded the final operational model with MAE = 84.97 W, and R^2 = 0.945. The incremental improvement from E4 demonstrates that careful tuning extracts remaining performance headroom, but the gains are modest compared to foundational choices in data preprocessing and feature engineering.

Experiment E6 addressed a data quality issue discovered during diagnostic analysis of E5 residuals. Systematic investigation of prediction errors revealed anomalous behavior during a specific period in September 2024. Cross-referencing with maintenance logs and inverter status records confirmed that the photovoltaic system experienced operational issues during September 19-24, 2024, resulting in power production substantially below the expected irradiance-to-power relationship. This represents a classic example of distribution shift: the model trained on normal operational data encountered a regime where the underlying physical system behaved differently due to equipment malfunction rather than environmental forcing. E6 implemented temporal exclusion of the identified malfunction period, removing approximately 5 days (1,440 samples at 5-minute resolution) from both training and validation sets while retaining these points in a separate diagnostic dataset for post-hoc anomaly detection validation. Retraining the E5 configuration on the filtered dataset yielded the final operational model with MAE = 72,68**,** and R^2 = 0.970 (see Figure 6.4). This experiment reinforces a fundamental principle in digital twin development: accurate models require not only sophisticated algorithms and features but also clean training data representative of normal system operation. Equipment malfunctions, sensor failures, and maintenance periods introduce patterns unrelated to the environmental-to-output mappings the model should learn. Explicit identification and exclusion of such periods, guided by domain knowledge and diagnostic analysis, prevents model corruption and enables the digital twin to serve its secondary function as an anomaly detector, flagging deviations between predicted and observed behavior that may indicate equipment issues requiring maintenance attention.

Table 6.1 Performance evolution across modeling experiments

Experiment	Configuration	MAE (W)	RMSE (W)	R^2	Key Finding
E1	Random Forest, raw inputs	264,42	552,73	0.259	Baseline; insufficient
E2	Random Forest, daytime filter	190,86	312,94	0.827	Data filtering dominant
E3	Random Forest, engineered features	88,26	180,41	0.942	Feature engineering critical
E4	XGBoost, standard config	103.32	183,55	0.940	Model choice secondary
E5	XGBoost, tuned + early stop	84.97	174,78	0.945	Final operational model
E6	XGBoost, tuned + early stop + data filtered	72,68	132,09	0,970	Final operational model improved

The cumulative progression quantifies a critical lesson for digital twin development: domain-aware data handling and feature construction account for approximately 70% of achievable performance improvement (E1→E3), while algorithm selection and hyperparameter

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

optimization contribute the remaining 30% (E3→E5). This ordering suggests that development resources should prioritize understanding the physical system and translating that understanding into preprocessing logic and feature transformations before investing heavily in algorithmic complexity or exhaustive hyperparameter searches.
Table 6.1 summarizes performance across the experimental progression leading to the final model. This comparison illustrates the relative importance of different methodological choices.
The progression reveals that data preprocessing and feature engineering account for the majority of performance improvement. The jump from E1 to E2 demonstrates that domain-aware filtering (removing nighttime data) provides greater benefit than algorithm selection. Similarly, the transition from E2 to E3 shows that temporal features and physics-inspired transformations substantially improve predictions even within the same model class. Algorithm choice (E3 to E4) and hyperparameter tuning (E4 to E5) provide incremental refinement atop the foundation established by proper data handling and feature construction. Finally, we performed a data filtering excluding the periods in which the inverter was not working properly and cancelling data with unphysical values in E6.

6.3.4.3 Diagnostic Analysis

Scatter plots comparing predicted and observed power reveal strong alignment along the identity line across the full operational range. Residual analysis shows approximately symmetric error distribution in moderate power regimes (500-2000 W) with slight heteroscedasticity at peak output, where prediction variance increases. This pattern is typical of PV predicting and reflects both genuine physical variability (cloud-edge effects, tracker response lags) and measurement uncertainty in high-power regimes.
Time series overlays of predicted and observed power demonstrate that the model successfully tracks diurnal cycles, reproduces ramp-up and ramp-down events during morning and evening transitions, and captures the general character of within-day variability. The model occasionally underestimates the amplitude of rapid power fluctuations during intermittent cloud conditions, smoothing transients more than observations. This conservative behavior is acceptable for planning applications though less ideal for very short-term (5-10 minute) predicting.

6.3.4.4 Operational Implications

The achieved accuracy supports the model's intended operational roles. For anomaly detection, the 72.68 W MAE provides sufficient sensitivity to identify inverter faults or performance degradation that cause deviations exceeding this baseline variability. For scenario analysis, the model enables reliable comparison of control strategies whose power differences exceed typical prediction errors. The computational efficiency of XGBoost inference (predictions in milliseconds for hourly data) allows real-time integration with greenhouse control systems.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

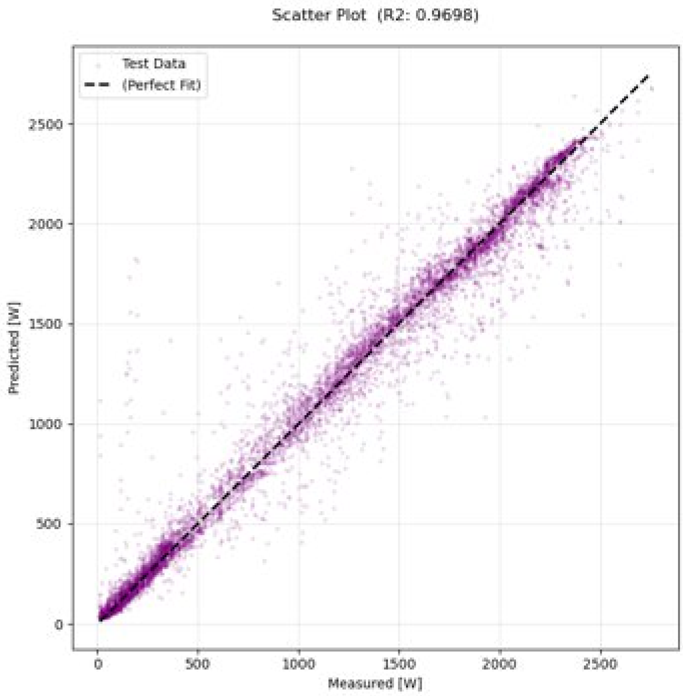


Fig. 6.4 Scatter plot for the E6 XGBoost model for PV Power prediction.

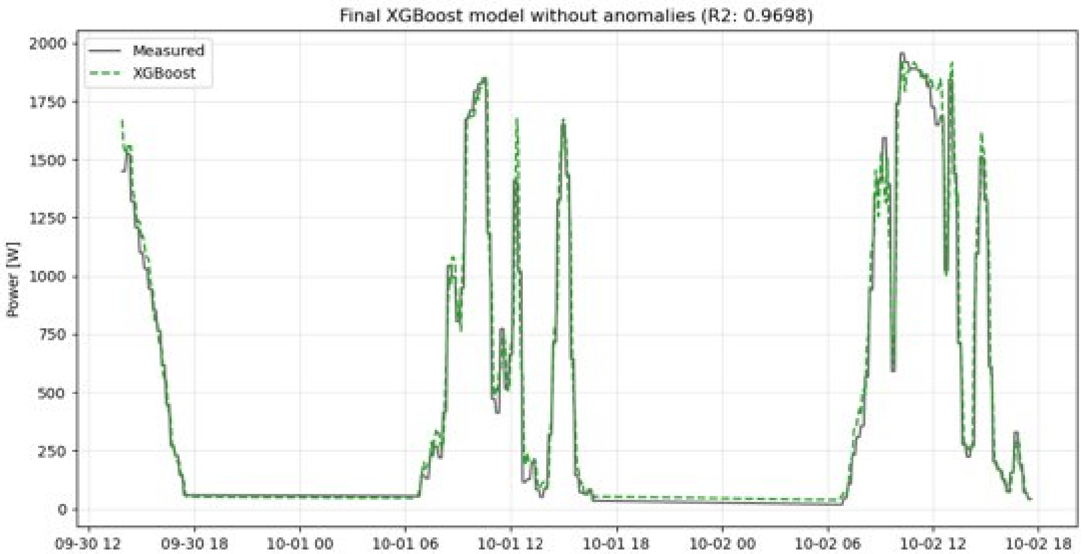


Fig. 6.5 Time series plot for the E6 XGBoost model for PV Power prediction for three days in September/October 2025.

6.4 Thermodynamic and Environmental Model

6.4.1 Objective

The objective of the thermodynamic modeling component is to develop a data-driven digital twin capable of predicting internal greenhouse microclimate, primarily air temperature, based on external environmental drivers and greenhouse operational state. Accurate internal temperature prediction is essential for crop protection and stress avoidance, optimization of

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

ventilation and shading strategies, coordination between photovoltaic energy production and climate control, and scenario analysis under alternative operational configurations or sensor availability constraints.
This work follows a progressive modeling strategy, starting with physical-data-driven tree-based regressors and evolving toward deliberately simplified configurations designed for deployments with minimal sensor instrumentation. The approach balances prediction accuracy against practical deployment constraints, recognizing that commercial greenhouse operations cannot support research-grade sensor density.

6.4.2 Physical-Data Driven Approach (Random Forest and XGBoost)

6.4.2.1 Modeling Rationale

The thermodynamic modeling component employs tree-based ensemble methods, specifically Random Forest (RF) and XGBoost regressors. These approaches provide strong nonlinear approximation capabilities, robustness to heterogeneous feature scales, high interpretability at the feature-importance level, and low computational cost for both training and inference. They are well suited to represent the effective thermodynamic response of the greenhouse without explicitly solving energy balance equations, instead learning response patterns from historical observations.

6.4.2.1.1 Input Features

The feature set incorporates multiple categories of information encoding known physical processes and temporal dependencies:

- External environmental drivers: Ambient temperature, global horizontal irradiance, wind speed, and relative humidity capture the primary forcing determining greenhouse thermal response.
- Temporal lag features: Recent values of key environmental variables at 5-minute intervals (t-5min, t-10min, t-15min, t-30min) represent short-term persistence and system memory reflecting thermal inertia.
- Rolling statistics: Mean and standard deviation computed over windows from 15 minutes to several hours characterize both the level and variability of external forcing, encoding whether conditions are stable or rapidly changing.
- Cyclical time encodings: Hour of day encoded through sine and cosine transformations, plus month indicators, represent diurnal and seasonal patterns without artificial discontinuities.
- Physics-inspired derived features: Variables such as vapor pressure deficit, temperature-radiation interaction terms, and accumulated thermal forcing over rolling windows translate domain knowledge into computable quantities.

This feature engineering strategy deliberately favors interpretability and physical grounding. Each feature corresponds to a measurable quantity or physically motivated transformation, facilitating diagnosis when predictions deviate from observations.

6.4.2.1.2 Training and Validation Procedures

Model training follows strictly chronological dataset splitting to respect temporal dependencies. The first 80% of data serves as the training set, with the final 20% reserved for testing. This temporal holdout ensures that performance metrics reflect genuine predicting capability rather than interpolation within the training period.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

Hyperparameter optimization employs time-series cross-validation using successive folds that preserve temporal ordering. For XGBoost, the search space includes number of estimators (model capacity), learning rate (boosting step size), maximum tree depth (complexity), and subsampling ratios for observations and features (regularization). After identifying promising configurations through randomized search, final training incorporates early stopping with patience of 50 rounds to prevent overfitting.
Known anomalous periods identified from sensor fault detection or maintenance logs are explicitly excluded from training datasets. This prevents the model from learning spurious correlations during periods when the physical system was not operating normally or measurements were unreliable.

6.4.2.1.3 Performance Summary

The tree-based models achieve strong predictive skill for internal temperature. XGBoost significantly outperforms baseline Random Forest configurations, with feature engineering accounting for a large fraction of the performance gain.

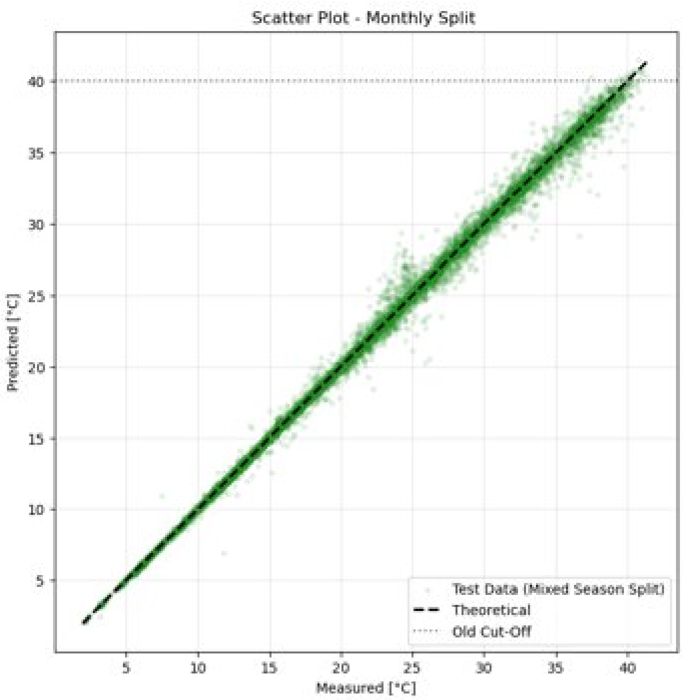


Fig. 6.6 Scatter plot for the T3 XGBoost model for greenhouse internal temperature prediction.

Temporal features (lags and rolling statistics) prove particularly valuable, confirming that greenhouse thermal response depends critically on recent history rather than instantaneous conditions alone.
Residual errors increase during extreme thermal regimes, particularly at temperatures above 35°C where nonlinear feedbacks between radiation absorption, ventilation effectiveness, and thermal stress become prominent as seen in **Error! Reference source not found.**. This regime-dependent behavior motivates the subsequent development of more sophisticated architectures, though for most operational conditions the tree-based models provide adequate accuracy for decision support.

6.4.3 Simplified "Low-Cost" Model

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

6.4.3.1 Rationale

A key requirement for real-world deployment is the ability to operate under limited sensing configurations where only a small number of low-cost sensors are available. To address this constraint, a simplified model configuration was developed and validated specifically for minimal instrumentation scenarios.
The low-cost sensor set restricts inputs to:

- External temperature (ambient air)
- Global horizontal irradiance (solar radiation)
- Wind speed
- Relative humidity (external)
- Internal radiation (single pyranometer at crop level)

This configuration eliminates most distributed internal temperature sensors, all internal humidity measurements, and detailed control state monitoring beyond what can be inferred from external conditions and time-of-day patterns. The total instrumentation cost approximates one-fifth that of the full sensor network used in baseline models.

6.4.3.2 Delta-Learning Strategy

To improve robustness and reduce bias under the constraint of limited sensor availability, the low-cost model adopts a delta-learning formulation, predicting the temperature differential ΔT = T_internal - T_external rather than absolute internal temperature.
This approach offers multiple advantages:

- it removes a large component of the diurnal and seasonal signal that dominates absolute temperature values. External temperature already captures much of the slow variation driven by solar forcing and atmospheric conditions. The learning task then focuses on the net thermodynamic response of the greenhouse structure.
- Delta-learning reduces sensitivity to systematic biases in external temperature measurements or changes in sensor calibration. Because both internal and external temperatures are referenced to the same external measurement, calibration offsets cancel when computing the differential.

Third, the approach improves generalization across different climatic conditions and seasons. The differential ΔT reflects greenhouse-specific properties (structure, glazing, ventilation capacity, thermal mass) that remain relatively stable, whereas absolute temperatures vary dramatically with location and time of year.

6.4.3.3 Model Implementation

The low-cost configuration employs the same XGBoost regressor framework used in full-sensor models, but with the modified target (ΔT) and reduced feature set. Feature engineering compensates for the loss of internal measurements by constructing additional input features that encode physical processes the model can no longer observe directly. Exponentially weighted moving averages of external temperature at multiple time constants (spanning 15 minutes to 4 hours) are computed and supplied as features representing thermal lag and structural inertia effects. Similarly, cumulative radiation integrals over rolling windows provide the model with explicit information about accumulated solar energy input that would otherwise be inferred from distributed internal temperature sensors. These engineered features enter the XGBoost model as standard numerical inputs alongside the reduced set of direct sensor measurements (external temperature, solar irradiance, wind speed, humidity, and single internal radiation measurement). Final predictions reconstruct absolute internal temperature

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

by adding the predicted differential to measured external temperature, ensuring physical consistency with external conditions while capturing greenhouse-specific thermal response through the learned ΔT mapping.

6.4.4 Results and Performance

6.4.4.1 Quantitative Metrics

The five experiments in Table 6.1 represent a systematic progression toward developing a practical, accurate greenhouse temperature prediction model.

Experiment T1 established the baseline approach using Random Forest and XGBoost regressors trained to predict absolute internal temperature directly from the full sensor network. This configuration utilized all available measurements, distributed internal temperature probes, external meteorological variables, and control states, operating continuously across all hours of operation. While achieving respectable performance ($R^2 = 0.976$), the mean absolute error of 0.631°C suggested room for improvement through alternative problem formulations.

Experiment T2 introduced the critical methodological shift to delta-learning, predicting the temperature differential ΔT = T_internal - T_external rather than absolute internal temperature. This reformulation used the same dataset and sensor configuration as T1 but reframed the learning task to focus on the net thermodynamic response of the greenhouse structure rather than absolute thermal levels. The performance improvement was substantial: MAE dropped to 0.308°C and R^2 increased to 0.991, demonstrating that removing the large diurnal and seasonal signals already captured by external temperature allowed the model to learn the greenhouse-specific thermal dynamics more effectively.

Experiment T3 refined T2 by implementing improved temporal splitting strategies that better respected the chronological dependencies in greenhouse operations, further reducing MAE to 0.212°C and pushing R^2 to 0.998. This improvement confirmed that validation methodology significantly impacts measured performance when dealing with systems exhibiting strong temporal autocorrelation.

Experiment LC1 addressed the practical deployment challenge by demonstrating that comparable performance could be achieved with drastically reduced sensor infrastructure. The "low-cost" configuration eliminated most distributed internal temperature sensors, retaining only external meteorological measurements (temperature, irradiance, wind, humidity) plus a single internal radiation sensor, approximately one-fifth the instrumentation cost of the full network. Despite this sensor reduction, LC1 achieved MAE of 0.280°C and R^2 of 0.997 by combining the delta-learning formulation with strategic feature engineering that compensated for missing direct measurements (see Fig. **6.7**). Engineered features included thermal lag representations through exponentially weighted moving averages, cumulative radiation integrals representing heat accumulation, and control regime indicators inferred from temporal patterns. This result validated that domain knowledge translated into thoughtful feature construction could substitute for extensive direct sensing, making accurate greenhouse digital twins economically viable for commercial deployment where instrumentation costs must be minimized.

Table 6.2 summarizes quantitative performance across the progression of thermodynamic models . These results represent evaluation on chronologically held-out test sets, ensuring metrics reflect genuine predicting capability.

The progression illustrates several important findings. First, the shift from absolute temperature prediction (T1) to delta-learning (T2, T3) substantially reduces error even within the same model class, confirming that focusing on the net greenhouse response rather than absolute

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

thermal levels simplifies the learning problem. The improvement from T2 to T3 demonstrates that careful attention to temporal splitting strategies significantly impacts performance assessment.
The low-cost model (LC1) achieves performance rivaling the best full-sensor tree-based approaches despite operating from a drastically reduced sensor set. This validates the hypothesis that strategic feature engineering can compensate for reduced measurement density in many operational scenarios. The progression illustrates several important findings. First, the shift from absolute temperature prediction (T1) to delta-learning (T2, T3) substantially reduces error even within the same model class, confirming that focusing on the net greenhouse response rather than absolute thermal levels simplifies the learning problem. The improvement from T2 to T3 demonstrates that careful attention to temporal splitting strategies significantly impacts performance assessment.

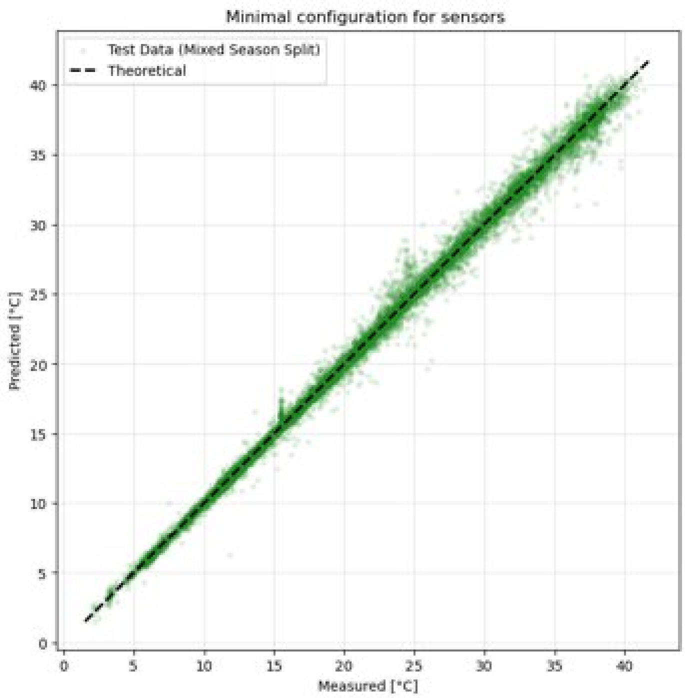


Fig. 6.7 Scatter plot for the LC1 XGBoost model for greenhouse internal temperature prediction.

The low-cost model (LC1) achieves performance rivaling the best full-sensor tree-based approaches despite operating from a drastically reduced sensor set. This validates the hypothesis that strategic feature engineering can compensate for reduced measurement density in many operational scenarios.

Table 6.2 Performance of thermodynamic and environmental models

Model	Strategy	Target	MAE (°C)	RMSE (°C)	R^2	Configuration
T1	RF/XGBoost	Absolute T	0.631	1.014	0.976	Full sensors, 24/7 operation
T2	RF/XGBoost	ΔT	0.308	0.628	0.991	Same dataset, delta formulation
T3	RF/XGBoost	ΔT	0.212	0.427	0.998	Improved temporal split
LC1	Low-cost XGBoost	Absolute T	0.280	0.500	0.997	Minimal sensors, delta-learning

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

6.4.4.1 Interpretation for Digital Twin Deployment

The combined results demonstrate that high-fidelity thermodynamic prediction is achievable without full physical modeling or dense sensor networks. The tree-based ensemble methods successfully capture the effective greenhouse thermal response across diverse operating conditions, with delta-learning formulations providing particular advantages for low-cost sensor configurations.

Feature engineering accounts for the majority of performance improvement, emphasizing that domain knowledge about thermal inertia, radiative forcing, and control dynamics translates directly into predictive capability. The low-cost model's strong performance validates the feasibility of deploying digital twins in commercial greenhouse settings where economic constraints limit instrumentation investment.

These models are suitable for real-time deployment, scenario simulation, and integration with higher-level control or optimization frameworks. The limitations observed during extreme thermal regimes and rapid transients motivate the subsequent development of more sophisticated architectures described in later sections, though for most operational conditions the RF/XGBoost models provide adequate accuracy for decision support.

6.5 GNN results and methods

6.5.1 Motivation and Problem Formulation

The tree-based models described in Section 5 demonstrated strong performance for greenhouse temperature prediction but exhibited two limitations that motivated development of more sophisticated architectures:

- performance degraded during extreme thermal events, particularly in the temperature range above 35°C.
- The models were trained on data from specific greenhouse configurations and did not readily generalize to different sensor networks or structural arrangements.

Graph Neural Networks address these limitations by explicitly representing the coupled nature of greenhouse thermodynamics. A greenhouse constitutes a strongly coupled dynamical system where radiation, ventilation, and thermal inertia jointly drive temperature evolution. Many measured variables exhibit statistical dependence because they are connected through physical processes such as radiative forcing, convective exchange, and control actions. Rather than treating measurements as an undifferentiated feature vector, the GNN formulation represents each measured variable as a node in a graph, with edges encoding inter-variable relationships. Seri et al. (2025) produced a first attempt to model the thermodynamics of greenhouse with GNN. The primary modeling objective remained one-step-ahead prediction of internal air temperature using multivariate time series of external environmental variables, internal sensor measurements However, the graph structure provided an inductive bias reflecting the system's coupled nature while remaining agnostic to explicit mechanistic equations that would be difficult to parameterize for novel agrivoltaic configurations.

6.5.2 Graph Construction and Data Representation

6.5.2.1 Inferring Graph Topology from Data

Because the greenhouse sensor network does not correspond to a predefined spatial graph with known geometry, graph topology was inferred from statistical dependencies in the training

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

data. An absolute correlation matrix was computed across all numerical variables using Pearson correlation with minimum sample requirements to reduce spurious links. An undirected edge connects nodes i and j if their absolute correlation exceeds a threshold, typically set to 0.6. Self-loops were included explicitly to preserve each node's identity during message passing.
This correlation-based approach yields edge structures compatible with PyTorch Geometric while avoiding requirements for manual specification of all physical couplings. Although correlation does not imply causation, the induced graphs empirically provided useful scaffolds for attention-based message passing and improved generalization relative to fully connected or unstructured baselines. Typical graphs contained 50-200 nodes depending on sensor density, with edge patterns reflecting the strong interconnections characteristic of greenhouse systems. Radiation variables, for instance, showed high correlation with multiple temperature sensors, , and power production measurements, manifesting as hub nodes.

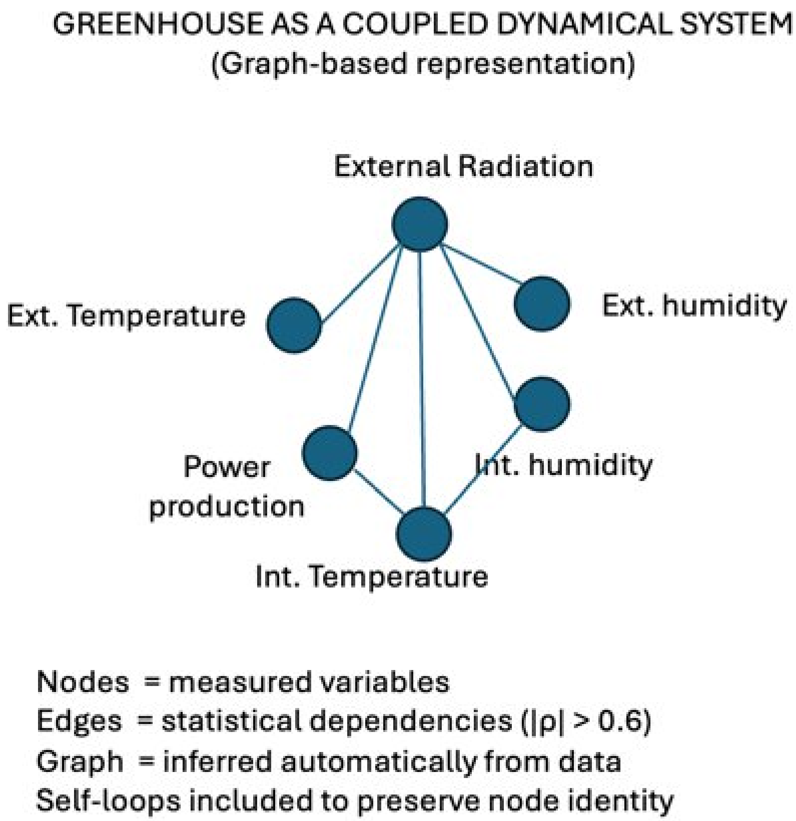


Fig. 6.8 Conceptual scheme of the GNN

This data-driven graph construction represents a key innovation relative to Deliverable D4.1, which relied on manually specified graph structures based on spatial and physical proximity. The automated topology inference eliminates the need for site-specific graph design, discovers non-obvious functional relationships beyond physical location, and enables seamless adaptation to heterogeneous sensor configurations across different greenhouse installations. This advancement substantially reduces deployment complexity while improving the model's ability to capture the true operational couplings governing greenhouse dynamics.

6.5.2.2 Temporal Representation and Sequence Generation

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

To capture greenhouse thermal inertia and control lag effects, the model was trained on sliding temporal windows. Each training example consisted of a tensor of shape *(SEQ_LEN, N, 1)* where *N* is the number of graph nodes and *SEQ_LEN* defines the historical context window. The target was the scaled value of internal temperature at time *t* + *SEQ_LEN*.
Sequence length was set to 48 timesteps at 5-minute resolution, spanning 4 hours of history. This extended horizon proved necessary to represent effective memory in the system, particularly under strong radiative forcing and high-temperature regimes where accumulated energy input drives thermal response over timescales exceeding typical control response times.

6.5.2.3 Data Scaling and Quality Control

All node variables were standardized using *StandardScaler fit* exclusively on the training split and applied to validation and test splits using the training statistics. This prevents information leakage through scaling parameters, which can artificially inflate performance and distort extreme-value behavior in multivariate graph learning.
The sensor quality issues previously identified required explicit handling during GNN training. Flat-line sensor failures, characterized by nearly constant measurements over extended periods despite changing environmental conditions, were detected using rolling standard deviation computed over 30-minute windows. Target samples flagged as potentially faulty were excluded from the regression loss while preserving the corresponding feature history as inputs. This approach prevented the model from learning incorrect mappings conditioned on erroneous labels without discarding entire time periods.

6.5.3 GNN Architecture and Training

6.5.3.1 Baseline Architecture: Temporal Encoding Plus Graph Convolution

The GNN architecture combined temporal and spatial encoding through a two-stage design. First, a temporal encoder implemented as a two-layer LSTM processed the history of each node independently. For each node, the final LSTM hidden state provided a latent representation of its recent temporal evolution, capturing persistence and lag effects arising from thermal mass, control response times, and system inertia.
These node embeddings then fed into a graph encoder applying two layers of attention-based message passing using *GATv2Conv* operations. Attention mechanisms proved appropriate for greenhouse modeling because coupling strengths are context-dependent
A small feed-forward prediction head mapped the final node embeddings to predicted values. During training, regression loss was applied only to the target node representing internal temperature, though the graph structure ensured that all nodes contributed to this prediction through message passing.

6.5.3.2 Radiation Memory Features

Early GNN models systematically underestimated temperatures above 35°C, suggesting insufficient sensitivity to accumulated radiative forcing. To address this, rolling radiation features were computed at 2-hour and 4-hour timescales for radiation-related variables identified through keyword matching. These features encoded integrated energy input and were physically motivated: internal temperature responds not only to instantaneous radiation but also to accumulated heating and convective equilibration occurring over hours.

6.5.4 Addressing Regime Heterogeneity: Mixture of Experts

6.5.4.1 Hard Gating Limitations

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

Initial attempts to address regime-dependent behavior employed hard-gated Mixture of Experts, where a discrete regime label selected one expert network for each prediction. While this approach improved performance in some subsets, it introduced instability and occasional pathological predictions. Diagnostic analysis revealed why: the regime labels exhibited substantial overlap in state space. The discretization based on control variables did not cleanly separate physical operating modes. Hard gating therefore frequently selected experts trained predominantly on mismatched conditions, yielding unstable outputs.

6.5.4.2 Soft Gating Solution

The final architecture employed *soft-gated Mixture of Experts* where expert contributions were continuously interpolated rather than discretely switched. The backbone network produced a pooled embedding of the entire system by averaging over node embeddings. A gating network processed this pooled representation to output mixture weights via softmax normalization. Multiple expert heads independently produced predictions, and the final output combined these predictions weighted by the learned mixture coefficients.
This formulation eliminated regime-switch artifacts and allowed smooth transitions between dynamical behaviors. The high-temperature expert received increased capacity through larger hidden dimensions, reflecting expectations that extreme thermal regimes involve rarer and more nonlinear dynamics. The gating mechanism learned to increase reliance on this specialized expert under warm conditions while maintaining contributions from other experts.

6.5.4.2 Loss Function Design

Training employed weighted Huber loss on scaled targets, with sample weights emphasizing high-temperature cases and down-weighting samples flagged as potentially faulty. Because high-temperature events are operationally critical yet underrepresented in training data, an additional tail underprediction penalty was computed in physical units. This term selectively penalized underestimation when real temperatures exceeded 35°C, directly counteracting the observed compression of extreme values.

6.5.5 Results and Performance Analysis

6.5.5.1 Quantitative Performance

The final soft-gated Mixture of Experts GNN achieved the following performance on the chronologically held-out test set:
Uncalibrated:

- $R^2 = 0.791$
- MAE = 2.11°C
- Mean bias = 0.09°C (nearly unbiased)
- Tail MAE (T > 35°C) = 3.41°C
- Maximum predicted temperature: ~40°C (observed maximum: 47.8°C)

After piecewise bulk calibration:

- $R^2 = 0.846$
- MAE = 1.91°C
- Tail MAE = 3.41°C (unchanged by design)

Post-hoc piecewise calibration fitted an affine transformation on the bulk region (T ≤ 35°C) and applied it only to predictions in that range. This corrected residual scale bias and offset in the bulk while intentionally leaving tail predictions unchanged to avoid compressing extreme

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

values. The calibration coefficients (slope approximately 0.765, intercept approximately 5.28) indicated mild adjustment without fundamental correction.

6.5.5.2 Ablation Study

An ablation study measures individual component contributions by observing performance changes when components are added or removed. This section employs a progressive ablation approach: starting from a baseline GNN (Experiment A), each experiment adds one targeted enhancement (quality masking, control dynamics encoding, radiation memory features, etc.) while holding other factors constant. By comparing performance before and after each modification, we quantify how much each component contributes to final model accuracy. This staged development methodology simultaneously documents the refinement process and establishes which innovations, data handling, feature engineering, or architectural sophistication, matter most for predictive capability.

Experiment A: Baseline GNN Architecture. The baseline configuration (Experiment A) established the foundation using a temporal-spatial architecture combining LSTM units for temporal encoding with GATv2 attention-based graph convolution for spatial message passing. This initial implementation used standard mean squared error loss without specialized handling of data quality issues or regime heterogeneity. While the model demonstrated reasonable overall performance with $R^2 = 0.70$, it exhibited two critical limitations: relatively high prediction errors (MAE > 2.5°C) and systematic compression of extreme temperature values. The maximum predicted temperature plateaued around 35°C despite observations extending to 47.8°C, indicating that the model learned to saturate predictions in the upper thermal range. This tail saturation behavior suggested insufficient sensitivity to the physical processes driving extreme heat events in agrivoltaic greenhouses.

Experiment B: Quality Masking Integration. Building directly on the baseline, Experiment B addressed a data integrity issue discovered through residual analysis. Certain temperature sensors exhibited flat-line failures where measurements remained nearly constant over extended periods despite changing environmental conditions. These artifacts, if used as training labels, would teach the model incorrect temperature-environment relationships. The quality masking enhancement implemented rolling variance detection to identify these flat-run periods and excluded flagged samples from the regression loss calculation while preserving their feature history as inputs. This approach prevented model corruption without discarding entire time periods. The quantitative impact on aggregate metrics was modest, but diagnostic analysis showed elimination of residual patterns corresponding to known sensor faults, establishing cleaner training signal for subsequent refinements.

Experiment C: Control Dynamics Encoding. These engineered features provided the model with direct information about control persistence and transition timing, aspects that pure end-to-end learning from raw positional values struggled to extract. The modification enabled better prediction during and immediately after control actions, as evidenced by the modest increase in maximum predicted temperature to approximately 36°C. The model began to represent that control changes alter the temperature-radiation relationship rather than treating actuator positions as static environmental variables.

Experiment D: Radiation Memory Features. Experiment D addressed the fundamental physics that internal temperature responds not only to instantaneous solar forcing but also to accumulated radiative heating over hours. The baseline temporal encoding used LSTM memory, but this general-purpose mechanism proved insufficient for capturing the specific timescales at which thermal mass and energy accumulation operate in greenhouse structures. The radiation memory enhancement computed rolling integrals of radiation-related variables over 2-hour and 4-hour windows, explicitly representing cumulative solar energy input. These

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

physics-informed features provided the model with direct access to integrated forcing measures known from greenhouse thermodynamics to drive temperature evolution. The impact was substantial in warm regimes: maximum predicted temperatures rose to approximately 38°C, and diagnostic analysis showed reduced systematic underestimation during sustained high-radiation periods. The model learned that temperatures under prolonged solar exposure depend on historical accumulation rather than instantaneous radiation levels alone, bringing predictions closer to observed extremes.

Experiment E: Hard Mixture of Experts (Failed Attempt). Recognizing that the model's remaining difficulty with extreme temperatures might reflect qualitatively different dynamics in distinct thermal regimes, Experiment E introduced a hard-gated Mixture of Experts architecture. Multiple specialized expert networks were trained to handle different operating conditions, with a discrete regime label selecting which expert generated predictions for each sample. The regime labels derived from clustering control variables and temperature ranges, attempting to partition state space into regions with consistent behavior. While this approach achieved localized improvements in some subsets, it introduced severe instability manifesting as occasional pathological predictions and discontinuous behavior at regime boundaries. Diagnostic analysis revealed the core problem: the regime labels exhibited substantial overlap in state space. Operating conditions did not cleanly separate into discrete modes, and the hard gating mechanism frequently selected experts trained predominantly on mismatched conditions. The discontinuous switching between experts created prediction artifacts that outweighed any gains from specialization, establishing that this architectural direction was fundamentally flawed for greenhouse modeling where regime boundaries are inherently ambiguous.

Experiment F: Soft Mixture of Experts. Learning from Experiment E's failure, Experiment F reformulated the Mixture of Experts using continuous blending rather than discrete selection. The soft-gated architecture allows all expert networks to contribute to every prediction, with learned mixture weights determining relative importance. A gating network processes the system state to produce normalized weights via softmax, enabling smooth transitions between expert contributions as conditions evolve. The high-temperature expert received increased capacity through larger hidden dimensions, reflecting expectations that extreme thermal regimes involve rarer and more complex dynamics requiring additional representational power. This formulation eliminated the switching artifacts of hard gating while maintaining the benefit of specialized sub-networks. Performance improved substantially: R^2 increased to 0.79, MAE reduced to 2.11°C, and maximum predicted temperatures reached approximately 40°C, significantly closer to observed extremes (see Fig. **6.9**). The soft blending mechanism proved robust across all operating conditions, validating the hypothesis that continuous interpolation between dynamical representations outperforms discrete regime switching when boundaries are ambiguous.

Experiment G: Tail-Aware Loss Function. Despite the improvements from soft expert blending, Experiment F still exhibited systematic underestimation of the highest temperatures, with tail MAE (temperatures > 35°C) remaining at 3.41°C. Experiment G addressed this through specialized loss function design explicitly targeting rare extreme events. The enhanced loss function combined standard Huber loss with an additional penalty term selectively emphasizing high-temperature samples. This tail-aware penalty increased gradient magnitudes when the model underestimated temperatures above 35°C, directly counteracting the observed compression. The modification operated in physical temperature units rather than normalized space to ensure the penalty scaled appropriately with operational significance. Training dynamics adjusted to allocate more learning capacity to the tail region despite extreme events being statistically rare. While aggregate metrics like overall MAE showed modest changes,

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

the specific tail MAE decreased and maximum predicted temperatures increased further. The model learned to represent extreme thermal events more faithfully, addressing the original systematic limitation identified in the baseline configuration.
Experiment H: Piecewise Calibration. The final enhancement in Experiment H applied post-hoc piecewise calibration to correct residual systematic biases in the bulk temperature range (T ≤ 35°C) while intentionally preserving the tail representation learned through previous refinements (see Fig. **6.10**). Analysis of Experiment G revealed a slight scale bias and offset in moderate temperature predictions that could be corrected through affine transformation without requiring architectural modification or retraining. The calibration procedure fitted linear parameters (slope ≈ 0.765, intercept ≈ 5.28) using only bulk-regime samples, then applied this transformation exclusively to predictions below 35°C. This conservative approach improved bulk accuracy, raising R^2 to 0.85 and reducing MAE to 1.91°C, without compressing extreme predictions that had been carefully cultivated through tail-aware training. The piecewise design acknowledged that different temperature regimes might have different optimal prediction scales while maintaining continuity at the boundary. This final configuration represented the operational model, balancing accuracy across all thermal regimes while successfully representing the extreme conditions critical for crop stress management.

Table 6.3 Ablation study of GNN-based digital twin components

ID	Configuration	Key Components	R^2	MAE (°C)	MAE_tail (>35°C)	Max Pred (°C)	Outcome
A	Baseline GNN	LSTM + GATv2, standard loss	~0.70	>2.5	>4.0	~35	Good fit, tail saturation
B	A + quality masking	Flat-run detection	-	-	-	~35	Removes faulty label artifacts
C	B + control dynamics	Lags, rolling, diffs	-	-	-	~36	Better regime transitions
D	C + radiation memory	2h/4h rolling radiation	-	-	-	~38	Improved warm regime
E	D + hard MoE	Discrete expert selection	-	-	-	unstable	Localized gains, instability
F	D + soft MoE	Continuous blending	0.79	2.11	3.41	~40	Robust across regimes
G	F + tail loss	Tail-aware penalty	-	-	↓	↑	Better extreme representation
H	G + calibration	Piecewise affine (bulk)	0.85	1.91	3.41	~40	Improved bulk, tail preserved

Table 6.3 summarizes the progressive improvement achieved through targeted methodological refinements. Performance was evaluated on the same held-out test period to ensure fair comparison.
The ablation demonstrates that performance gains arose from targeted, physically motivated interventions rather than increased model complexity alone. Data quality masking prevented learning from faulty labels. Radiation memory and control dynamics captured nonlinear thermal response. Soft expert blending proved superior to hard regime separation when regime boundaries are ambiguous. Tail-aware learning and piecewise calibration addressed rare extremes without compromising bulk accuracy.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

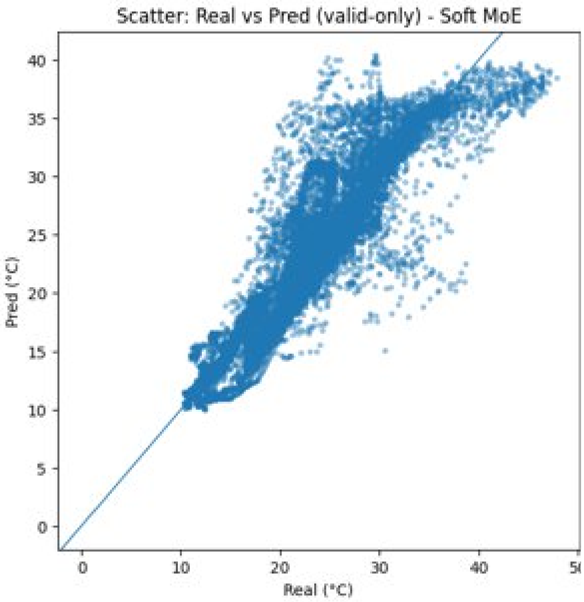


Fig. 6.9 Scatter plot for the F GNN model for greenhouse internal temperature prediction.

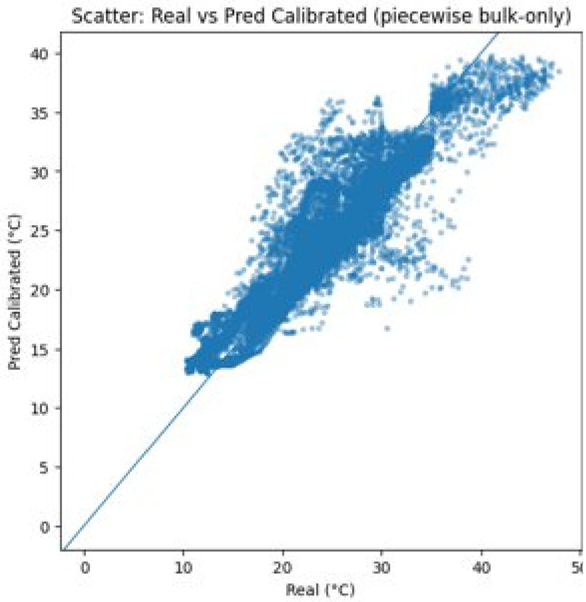


Fig. 6.10 Scatter plot for the H GNN model for greenhouse internal temperature prediction.

6.5.6 Operational Implications

6.5.6.1 Robustness and Trustworthiness

The GNN-based digital twin demonstrates operational robustness across several dimensions critical for deployment in controlled agricultural systems. It maintains near-zero mean bias, exhibits stable behavior across control regimes, and handles known sensor failure modes gracefully through explicit quality masking. This robustness arguably matters more for practical decision support than marginal accuracy improvements, as models producing occasional implausible predictions undermine operator confidence.
The soft-gated Mixture of Experts formulation yields smooth and physically interpretable behavior even when operating regimes overlap or boundaries are ambiguous. This continuous adaptation to system state avoids discontinuities that would complicate integration with model predictive control or rule-based decision systems.

6.5.6.2 Handling Critical Extremes

High-temperature events, though rare, are operationally critical due to associations with crop stress, yield loss, and emergency control requirements. The ablation study confirmed that naive approaches tend to compress extreme predictions, leading to systematic underestimation. The combination of radiation memory features, tail-aware loss shaping, and increased expert capacity for extreme regimes substantially mitigated this issue.
The remaining underestimation of the most extreme observed temperatures should not be interpreted as model failure. Rather, it reflects conservative extrapolation in a data-sparse regime and highlights an advantage of the digital twin paradigm: the model can be coupled with scenario simulation to explore hypothetical extremes beyond the empirical distribution, supporting stress testing and contingency planning.

6.5.6.3 Deployment Capabilities

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

The final GNN model supports multiple operational applications. Short-term temperature predicting with lead times of 30-60 minutes enables proactive control actions to prevent thermal stress. What-if simulations assess the impact of alternative ventilation or shading strategies before implementation. Digital shadowing compares observed behavior against model predictions to detect anomalies indicating sensor failures or unexpected system responses. The model's explicit encoding of temporal memory, inter-variable coupling, and control dynamics makes it suitable for integration into closed-loop decision-support frameworks where smooth gradients and stable behavior are essential.

6.6 Crop Growth Modeling Component

6.6.1 Objective and Data Characteristics

The initial scope of the digital twin included prediction of crop development indicators as a third modeling component, complementing temperature and power predicting. Crop variables such as plant height, leaf count, canopy dimensions, and cumulative yield represent the ultimate performance metrics for greenhouse operations, and their integration would enable truly comprehensive decision support linking environmental conditions, energy production, and agricultural outcomes.

However, investigation of crop growth modeling revealed fundamental challenges stemming from the temporal characteristics of available crop data. Unlike environmental sensors providing continuous measurements at minute-to-hour resolution, crop assessments in the REGACE demonstration sites derive from manual field measurements conducted approximately every one/two weeks. Standard greenhouse practice performs periodic crop monitoring at this frequency due to labor constraints and the relative slowness of developmental changes compared to environmental fluctuations.

The available crop dataset stored measurements interpolated to match the environmental data frequency. Between actual observation points, values were generated through temporal interpolation with added noise to simulate variability. While this preprocessing may serve visualization or data storage purposes, it creates a critical identifiability problem for supervised learning. The interpolated time series exhibits smooth temporal structure imposed by the interpolation algorithm rather than actual crop growth dynamics. Supervised models trained on these synthetic intermediate values cannot distinguish genuine environmental responses from interpolation artifacts.

The crop data encompassed multiple combinations of crop types (tomato, zucchini, eggplants and lattuce), experimental configurations (reference, photovoltaic exposure), and measured variables (height, leaf count, cumulative yield). This diversity potentially enables pooled learning across related crop series, though the fundamental limitation of sparse true measurements affects all series equivalently.

6.6.2 Methodological Approach

6.6.2.1 Data Integration and Temporal Alignment

Crop and environmental datasets required careful alignment due to their different temporal characteristics. Environmental data spanning the full monitoring period contained 219 feature columns after cleaning, while crop variables covered a shorter time window where observations were available. Temporal joins used nearest-neighbor matching with appropriate tolerance to handle frequency mismatches.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

Modeling was restricted to the crop-valid time window where target variables contained values, eliminating environmental-only periods unrelated to crop dynamics. The merged dataset for baseline experiments contained approximately 28,000 timestamps with aligned environmental features and a selected crop target variable.

6.6.2.2 Daily and Weekly Aggregation

Because crop development occurs over days to weeks rather than hours, environmental features were aggregated to daily and weekly temporal resolutions. Daily aggregation used mean values for continuous variables and mode for discrete states. This temporal compression reduced dataset size while focusing on timescales relevant to biological processes.

6.6.2.3 Dynamic Formulation

Initial experiments modeled daily height increments $\Delta h(d) = h(d) - h(d-1)$ rather than absolute values. This dynamic formulation theoretically isolates growth response from baseline size, potentially improving learning. However, differencing amplifies noise in sparse measurements and proved ineffective given data quality constraints.

6.6.2.4 Physiology-Inspired Features

To incorporate domain knowledge about crop growth drivers, several derived features were constructed. Growing Degree Days accumulated effective temperature above crop-specific base thresholds, providing a unified thermal forcing metric. Cumulative solar radiation represented integrated light energy available for photosynthesis. These variables were computed at native environmental resolution and aggregated to match temporal scales.

6.6.2.5 Pooled Panel Learning

To increase effective sample size, crop data were restructured from wide format (one column per series) to panel format with explicit identifiers for crop type, configuration, and metric. Environmental variables were similarly organized. Weekly aggregation provided a common temporal resolution, with robust period indexing ensuring proper alignment across series. This approach enabled simultaneous learning across multiple crops and configurations.

6.6.2.3 Phenological Stage Classification

When quantitative regression proved unstable, a coarser classification objective was tested. Each crop series was normalized to unit range using within-series minimum and maximum, then discretized into three phenological stages: early (normalized value < 0.33), mid (0.33-0.67), and late (> 0.67). XGBoost classifiers were trained to predict stage from environmental features using temporal train-test splits.

6.6.3 Experimental Results

6.6.3.1 Baseline Daily Regression

XGBoost models trained on daily height increments achieved perfect training performance ($R^2 = 1.000$, RMSE ≈ 0.001) indicating overfitting to training data patterns. Test set performance collapsed completely with RMSE ≈ 2.2 and negative $R^2 = -0.61$, demonstrating no predictive capability. The model learned training data artifacts without capturing generalizable growth relationships.

6.6.3.2 Physiology-Inspired Features

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Incorporating Growing Degree Days and cumulative radiation features did not improve performance. Test RMSE remained approximately 2.0 with negative $R^2 = -0.38$. The lack of improvement despite physically motivated features confirmed that the problem was not feature selection but rather the absence of genuine signal in interpolated crop targets at daily timescales.

6.6.3.3 Weekly Pooled Regression

Pooling multiple crop series at weekly resolution increased sample size but did not resolve fundamental issues. Test RMSE of 66.2 with negative $R^2 = -0.24$ indicated predictions worse than naive mean baselines. The large error magnitude reflected heterogeneity across crop types and metrics combined in the pooled dataset.

6.6.3.4 Phenological Classification

Classification into three growth stages achieved test accuracy of 0.25, below the random baseline of 0.33 for three classes. Predictions collapsed predominantly to the "early" class, indicating insufficient discriminative information in environmental features for stage prediction. This outcome suggests that even coarse-grained crop development cannot be reliably inferred from current environmental measurements alone.

6.6.3.5 Summary of Results

Table 6.4 Crop growth modeling experimental results

Experiment	Target	Temporal Resolution	Test RMSE	Test R^2	Test Accuracy	Outcome
Baseline regression	Height increment (Δh)	Daily	2.21	-0.61	-	No generalization
Physiology features	Height increment (Δh)	Daily	2.05	-0.38	-	No improvement
Pooled regression	Mixed metrics	Weekly	66.2	-0.24	-	Worse than baseline
Stage classification	Early/mid/late	Weekly	-	-	0.25	Below random (0.33)

Table 6.4 compiles quantitative outcomes across modeling experiments.
All experiments showed negative R^2 or below-baseline accuracy, indicating systematic failure to predict crop variables from environmental drivers under current data conditions.

6.6.4 Interpretation and Data Requirements

6.6.4.1 Root Cause Analysis

The consistent failure across multiple modeling approaches, feature sets, and temporal resolutions points to a data identifiability problem rather than methodological inadequacy. Environmental variables provide continuous, high-quality measurements capturing forcing conditions. Crop variables nominally match this frequency but derive from bi-weekly true observations with intermediate values synthesized through interpolation.
Supervised learning algorithms cannot distinguish between genuine crop response to environmental conditions and patterns imposed by interpolation. The model training process optimizes predictions on interpolated values that do not represent actual crop states at those timestamps. Test set failure indicates that learned patterns do not transfer to genuinely observed crop measurements.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

This interpretation is supported by the perfect training fit combined with complete test failure, a signature of learning spurious patterns rather than underlying relationships. The problem is not model capacity, feature relevance, or hyperparameter tuning, all of which were explored. The fundamental issue is insufficient information content in the target variable at the temporal frequencies where environmental forcing operates.

Implications for Digital Twin Development

This negative result provides valuable guidance for future crop-aware digital twin implementations. Current data infrastructure is insufficient for supervised learning of crop growth dynamics from environmental measurements. Operational deployment would require at least one of the following modifications:

1. Higher-frequency crop measurements: Daily or sub-daily observations through automated sensing such as computer vision-based height estimation, canopy coverage analysis, or other non-invasive monitoring. This provides genuine temporal structure at scales matching environmental variability.
2. Explicit observation flagging: If interpolation serves data management purposes, preserve raw measurement timestamps and provide flags distinguishing true observations from interpolated values. Models could then be trained exclusively on actual measurements, avoiding spurious patterns.
3. Hybrid mechanistic-empirical approaches: Incorporate simplified crop growth models (thermal time accumulation, light integral-based development) providing physical constraints and temporal structure. Data-driven components would refine mechanistic predictions using sparse observations through data assimilation techniques such as Kalman filtering.
4. Alternative prediction objectives: Rather than pointwise prediction, focus on seasonal-scale cumulative yields that integrate environmental history over weeks to months. These aggregated targets are more robust to measurement noise and interpolation artifacts.

6.6.4.2 Operational Recommendations

For REGACE demonstration sites, the most practical near-term approach combines cumulative yield modeling at seasonal scales with hybrid mechanistic components. Cumulative yield naturally accumulates environmental history, making it less sensitive to short-term measurement artifacts. A simplified growth kernel based on Growing Degree Days and photosynthetically active radiation could provide baseline development predictions, with empirical corrections learning site-specific or variety-specific deviations.

Longer-term infrastructure investment should prioritize automated crop monitoring. Computer vision systems analyzing overhead imagery can provide daily estimates of canopy dimensions at marginal cost after initial deployment. Such systems would transform crop modeling feasibility while providing additional operational benefits including early pest detection and spatial growth variability mapping.

6.6.4.3 Comparative Context

The crop modeling outcome contrasts with successful temperature and power prediction components but does not diminish overall digital twin value. Temperature and power respond to environmental forcing at timescales (minutes to hours) where sensor measurements capture actual dynamics. Crop growth operates at fundamentally longer timescales (days to weeks)

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

where current observation frequency is insufficient. This distinction establishes clear technical requirements rather than indicating conceptual limitations of data-driven approaches.
Successful digital twins require appropriate matching between prediction objectives, data infrastructure, and modeling approaches. The crop component investigation clarifies this matching for agricultural applications.

6.7 Multi-Output Modelling of Greenhouse Microclimate Variables

Previous modelling activities demonstrated excellent performance of tree-based XGBoost models for photovoltaic power production and internal air temperature prediction, while Graph Neural Network (GNN) approaches yielded comparable but not consistently superior results at a substantially higher computational cost. In contrast, the prediction of crop-related variables proved more challenging, reflecting the weaker and noisier signal associated with crop growth dynamics at the available temporal resolution. Building on these findings, this section now extends the analysis to a broader multi-target Digital Twin that includes, in addition to microclimate variables, the prediction of power production from the agrivoltaic PV system. The multi-target XGBoost framework has been selected due to its high computational efficiency, robustness, and ease of implementation in an operational setting. Preliminary attempts to construct a multi-target GNN model revealed extreme memory demands, exceeding the available GPU capacity and preventing feasible training. For these reasons, the following sections present the methodology and results obtained using the extended multi-output XGBoost approach.

6.7.1 Methodology

This section investigates a multi-output machine-learning approach for the simultaneous prediction of key internal greenhouse variables and PV power production. The analysis was conducted by considering pairs of sensor locations ("poles"), each pair composed of one area under PV coverage and one area without PV coverage. Four independent model configurations were analysed, corresponding to pole pairs (1–2), (3–4), (5–6), and (7–8), where poles 2, 3, 6, and 7 are located beneath PV panels, and poles 1, 4, 5, and 8 are not, as shown in Figure 6.9.

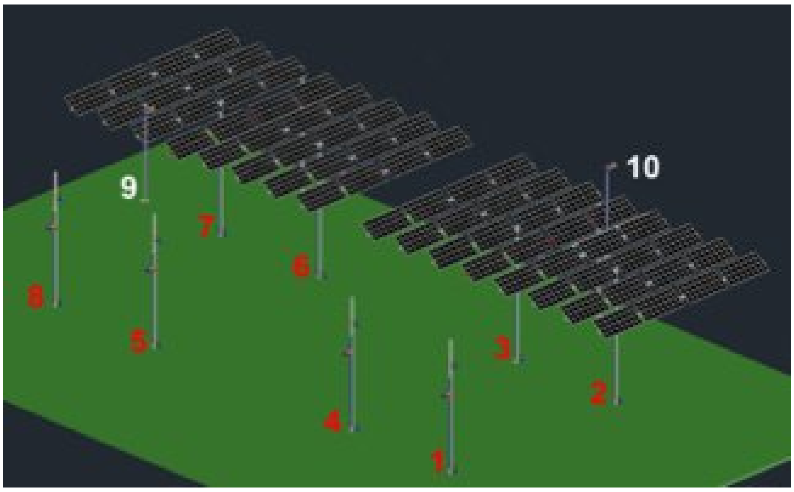


Fig. 6.11 Design of the greenhouse with the poles and the associated numbers

For each pole pair, a supervised regression model was trained to predict nine target variables simultaneously:

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

- Photosynthetically Active Radiation (PAR) at both poles
- CO_2 concentration at both poles
- Relative Humidity (RH) at both poles
- Air Temperature at both poles
- Power Production from the agrivoltaic PV system

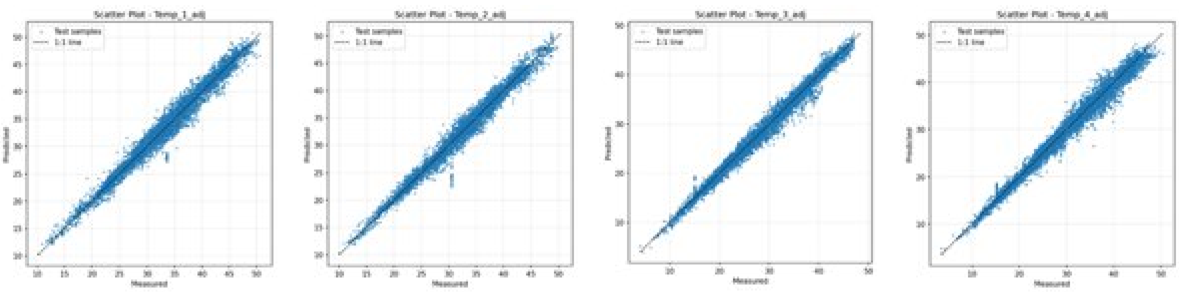
All microclimate variables correspond to sensor-adjusted values ("adj"), while power production is the measured instantaneous electrical output.
The modelling framework remains based on gradient-boosted decision trees (XGBoost), implemented through a multi-output wrapper enabling joint prediction of multiple variables within a single training pass. This choice follows earlier assessments showing that tree-based models achieve strong predictive performance for greenhouse processes while maintaining low computational and operational overhead. Input features include all available numerical variables describing the external meteorological forcing, internal greenhouse sensor network, actuator states (PV panel orientation, ventilation, windows), and crop-related indicators. Crop variables remain included to preserve methodological consistency and compliance with project requirements, despite their limited direct impact on short-term predictions.
A physically informed strategy was retained for temperature by modelling ΔT relative to the external ambient temperature, later converted back to absolute temperature values. All other targets, including power production, were modelled directly in their absolute units. Training and evaluation followed the established "month-aware" temporal split: for each month, days 1–23 were used for training and days 24–end for testing. Performance was assessed using R^2, RMSE, and MAE. Scatter plots comparing observations and predictions were produced for all targets and poles (not included in the present version).

6.7.2 Results

Overall, the extended multi-output XGBoost models show strong predictive performance across all pole pairs and target variables, confirming the suitability of the approach for an integrated Digital Twin of microclimate and energy production.

6.7.2.1 Temperature and Relative Humidity

Air temperature predictions remain highly accurate across all poles, with R^2 scores typically above 0.97–0.99 and RMSE values generally below 1.2 °C. These results confirm that the ΔT formulation effectively captures internal thermal behaviour under both PV-covered and uncovered conditions.



by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

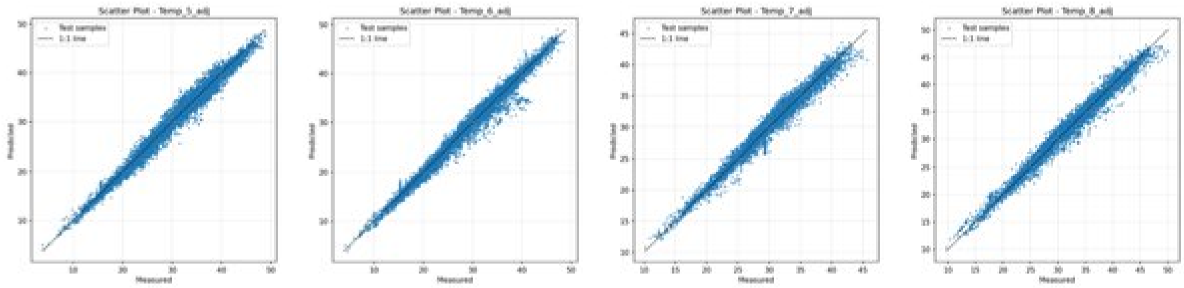


Fig. 6.12 Scatter plots for the 8 temperature sensors in the greenhouse poles.

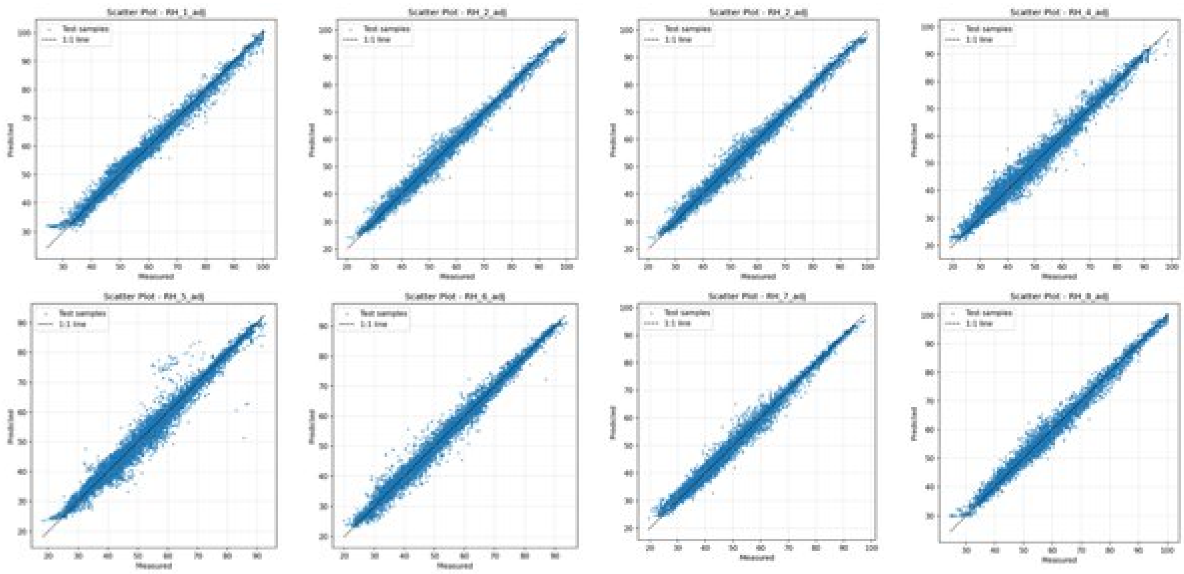
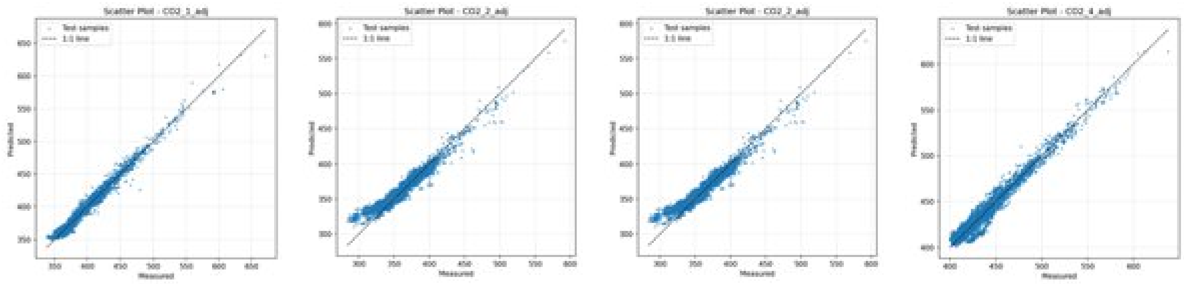


Fig. 6.13 Scatter plots for the 8 relative humidity sensors in the greenhouses poles.

Relative humidity predictions also remain robust, with R^2 values between approximately 0.98 and 0.995 and RMSE values typically between 1.2–2.6 % reflecting the strong coupling between humidity, temperature, and ventilation states encoded in the input features.

6.7.2.2 CO_2 Concentration

CO_2 predictions show generally high accuracy across pole pairs, with R^2 values ranging from approximately 0.93 to 0.98. However, some variability is observed across pole pairs. In particular, one PV location (pole 6) exhibits a lower R^2 and substantially higher RMSE, indicating that local CO_2 dynamics in that area are likely influenced by site-specific processes (e.g. airflow patterns, crop proximity, or sensor placement) that are less well captured by the available inputs. Outside this specific case, CO_2 predictions remain robust and stable.



by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

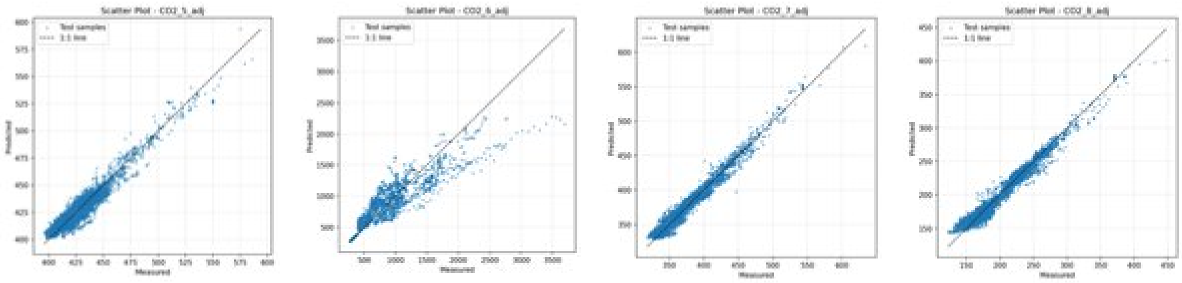


Fig. 6.14 Scatter plots for the 8 CO_2 sensors in the greenhouse poles.

6.7.2.3 Photosynthetically Active Radiation (PAR)

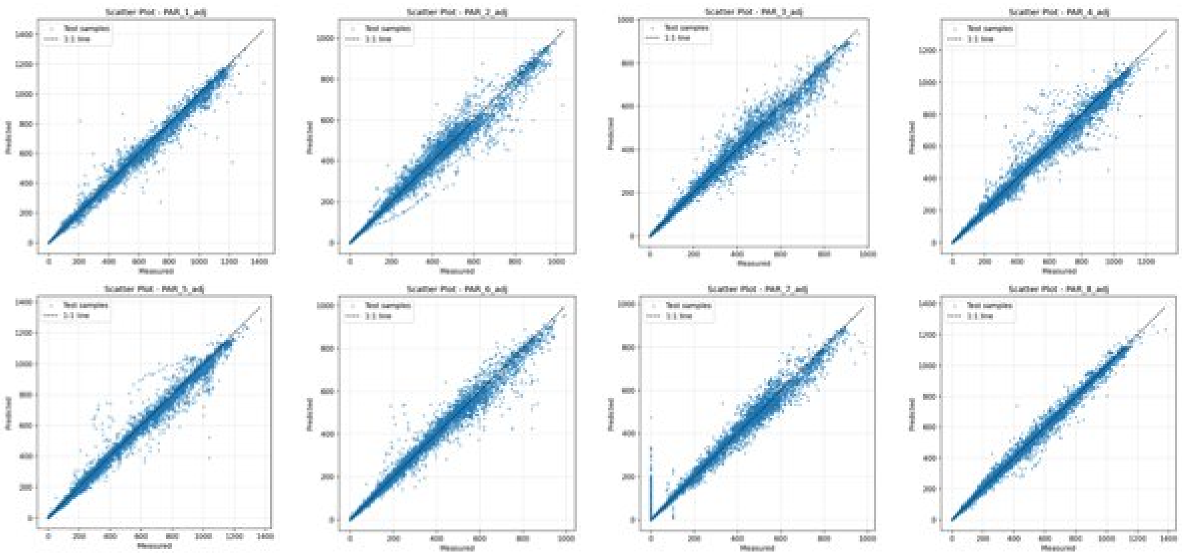
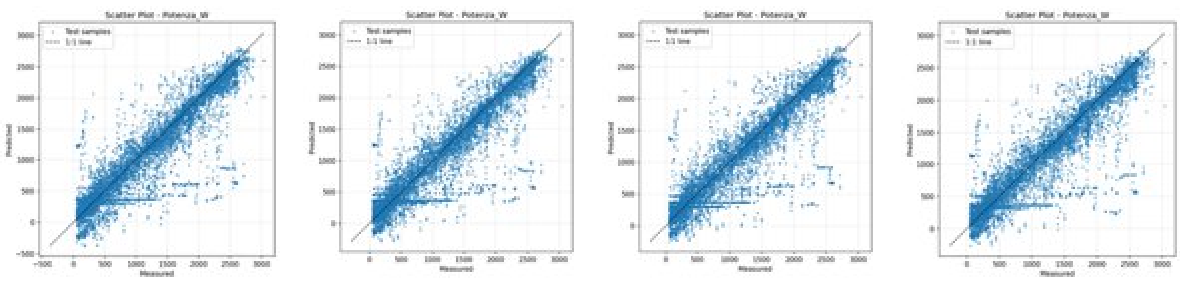


Fig. 6.15 Scatter plots for the 8 PAR sensors in the greenhouse poles.

PAR predictions achieve very high performance for most poles, especially in non-PV areas where R^2 values frequently exceed 0.99. Slightly reduced performance is observed for one PV-covered pole (pole 7), where stronger variability induced by panel orientation and transient shading leads to increased prediction uncertainty. Nonetheless, even in this case, the model retains a high explanatory power, demonstrating its ability to capture complex radiative modulation effects induced by the agrivoltaic system.

6.7.2.4 Power Production

The inclusion of power production as a ninth target shows that the multi-output framework is able to maintain solid performance for energy modelling. Across all pole-pair configurations, the model achieves R^2 values around 0.84–0.85, with RMSE values near 320–325 W and MAE values around 183–188 W. These results are consistent and stable across all four runs, reflecting the dominant dependency of power production on irradiance, panel orientation, and thermal conditions, all of which are well represented among the input features.



by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

Fig. 6.16 Scatter plots for the 4 Power production predictions.

6.7.2.5 Comparative performance across pole pairs

The consistency of results across the four independent pole-pair simulations demonstrates that the proposed multi-output modelling strategy is robust with respect to spatial heterogeneity within the greenhouse. Importantly, modelling each PV / non-PV pair independently confirms that high predictive skill can be achieved without explicitly coupling all sensor locations within a single, highly complex model. Table 6.5 summarises the quantitative performance metrics (R^2, RMSE, MAE) for all variables and pole pairs.
The results demonstrate that a multi-output XGBoost approach provides accurate and robust predictions of key greenhouse microclimate variables across both PV-covered and non-PV areas.
Despite its relative simplicity, the methodology achieves consistently high performance for temperature, humidity, CO_2 concentration, PAR and Power production, while remaining computationally efficient and suitable for operational deployment within a digital-twin framework.
The analysis confirms that modelling PV and non-PV zones in paired configurations effectively captures the main spatial contrasts induced by agrivoltaic structures, without the need for highly complex or computationally demanding architectures. These findings support the adoption of multi-output tree-based models as a pragmatic and scalable solution for real-time monitoring and control applications in agrivoltaic greenhouse systems.

Table 6.5 Performance metrics for all pole pairs and target variables

Pole	Variable	R^2	RMSE	MAE
1	PAR (μmol/s)	0.989	31.67	14.49
1	CO_2 (ppm)	0.984	4.62	2.87
1	RH (%)	0.991	1.65	1.04
1	Temp (°C)	0.976	1.08	0.72
1–2 Power	**Power [W]**	**0.846**	**324.26**	**188.17**
2	PAR (μmol/s)	0.976	32.71	15.61
2	CO_2 (ppm)	0.959	6.70	4.03
2	RH (%)	0.991	1.53	0.95
2	Temp (°C)	0.978	0.98	0.62
3	PAR (μmol/s)	0.979	31.75	16.22
3	CO_2 (ppm)	0.964	4.73	3.23
3	RH (%)	0.992	1.81	1.30
3	Temp (°C)	0.989	0.97	0.69
3–4 Power	**Power [W]**	**0.850**	**319.64**	**183.40**
4	PAR (μmol/s)	0.983	41.81	20.26
4	CO_2 (ppm)	0.975	5.29	3.86
4	RH (%)	0.983	2.48	1.76
4	Temp (°C)	0.984	1.24	0.89
5	PAR (μmol/s)	0.987	37.48	18.54
5	CO_2 (ppm)	0.935	5.81	4.23

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

5	RH (%)	0.978	2.60	1.75
5	Temp (°C)	0.985	1.13	0.84
5–6 Power	**Power [W]**	**0.849**	**321.61**	**187.01**
6	PAR (μmol/s)	0.984	28.91	14.57
6	CO_2 (ppm)	0.932	114.95	44.21
6	RH (%)	0.989	1.92	1.34
6	Temp (°C)	0.985	1.12	0.77
7	PAR (μmol/s)	0.975	30.88	14.68
7	CO_2 (ppm)	0.976	5.61	3.68
7	RH (%)	0.988	1.67	1.07
7	Temp (°C)	0.977	0.84	0.58
7–8 Power	**Power [W]**	**0.848**	**322.23**	**186.38**
8	PAR (μmol/s)	0.994	22.55	11.76
8	CO_2 (ppm)	0.978	6.81	4.18
8	RH (%)	0.991	1.67	1.06
8	Temp (°C)	0.978	0.97	0.68

6.7.3 Additional XGBoost Model Configurations

In operational greenhouse environments, high-resolution and densely instrumented datasets are not always available. The number and type of sensors can vary substantially across installations, and real-case scenarios often rely on a limited set of measurements constrained by cost, maintenance requirements, or data-availability considerations. For this reason, it is essential to evaluate modelling configurations that remain robust even when only a minimal subset of environmental and photovoltaic variables is accessible.
To address this requirement, two streamlined XGBoost model configurations were developed. The first relies exclusively on external meteorological variables and serves as a baseline representation of microclimate dynamics under minimal sensing conditions. The second incorporates a reduced but essential set of photovoltaic measurements, enabling joint prediction of internal microclimate variables and power production. Together, these models provide a practical reference framework for real-world deployments, where the Digital Twin must operate reliably despite limited instrumentation and heterogeneous sensor availability.

6.7.3.1 XGBoost “Only Microclimate” Model

The first configuration relies exclusively on external meteorological variables to predict the internal microclimate state of the greenhouse. The following input features were used:

- GHI_env (Global Horizontal Irradiance)
- RH_env (External Relative Humidity)
- Temp_env (External Air Temperature)
- Ws_env (Wind Speed)
- Wd_env (Wind Direction)
- Wr_env (Wind Run / directional wind-integrated parameter)
- CO2_env (External CO_2 concentration)

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

These variables represent the dominant atmospheric forcing acting at the greenhouse boundary and encode radiative, advective and mass-exchange processes driving short-term internal variability.
The model predicts the following internal sensor-adjusted variables:

- Temp_3_adj (internal air temperature)
- RH_3_adj (internal relative humidity)
- CO2_3_adj (internal CO_2 concentration)
- PAR_3_adj (internal photosynthetically active radiation)

All targets correspond to measurements from the internal microclimate instrumentation network. This configuration establishes a physically informed baseline for evaluating the added value of photovoltaic measurements. The results obtained from this model are summarised in Table 6.6.

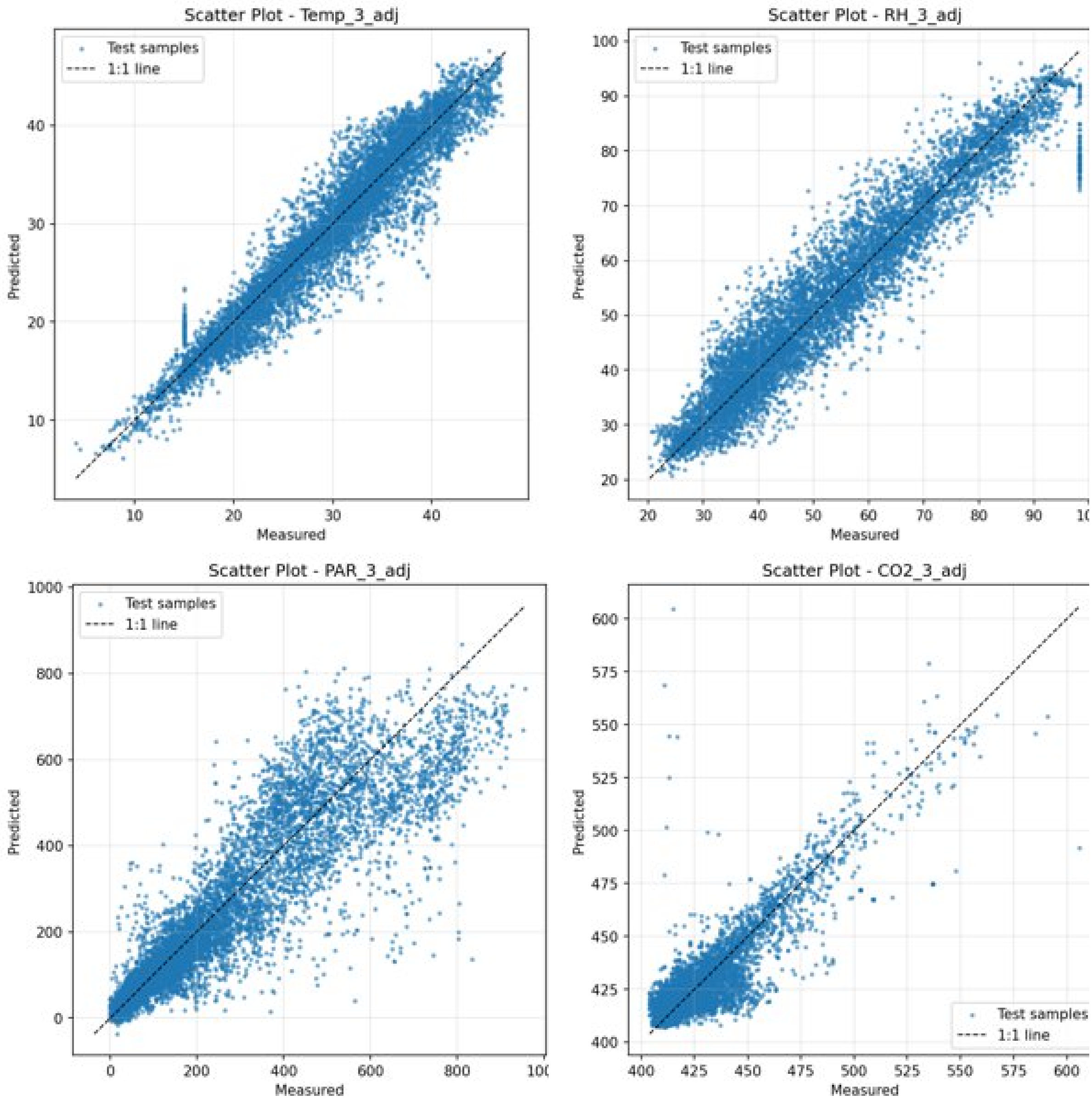


Fig. 6. 17 Scatter plots for temperature, relative humidity, CO_2 and PAR for pole 3 level adjustable.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

Table 6.6 Performance metrics for the XGBoost "Only Microclimate" model

Variable	R^2	RMSE	MAE
Temp_3_adj (°C)	0.9316	2.4406	1.7644
RH_3_adj (%)	0.9292	5.3806	3.9173
CO2_3_adj (ppm)	0.7475	12.6076	7.0284
PAR_3_adj (µmol/s)	0.8267	91.3313	59.1938

The performance confirms that internal temperature and relative humidity are strongly constrained by external forcing, while CO_2 and PAR show weaker predictability due to their dependence on internal airflow regulation, shading, and crop-related processes.

6.7.3.2 XGBoost "Microclimate + PV" Model

The second configuration extends the previous model by incorporating measurements from the agrivoltaic photovoltaic system. These variables describe the instantaneous radiative, thermal and geometric behaviour of the PV arrays located on the east and west sides of the greenhouse.
The full set of inputs is:
External meteorological variables (as in the previous model):

- GHI_env, RH_env, Temp_env, Ws_env, Wd_env, Wr_env, CO2_env

PV-related variables (east and west arrays):

- east_GTI (Global Tilted Irradiance – east array)
- east_RTI (Reflected Tilted Irradiance – east array)
- east_T_PV (PV cell temperature – east)
- east_PV_angles (orientation/tilt angles – east)
- WEST_PV_GTI (Global Tilted Irradiance – west array)
- WEST_PV_RTI (Reflected Tilted Irradiance – west array)
- WEST_PV__T_PV (PV cell temperature – west)
- WEST_PV_angles (orientation/tilt angles – west)

By introducing PV temperature, irradiance components, and dynamic panel orientation, this model captures the direct interaction between radiative modulation by PV arrays, thermal loading, and microclimate behaviour.
The model predicts the same internal microclimate variables as the preceding configuration:

- Temp_3_adj
- RH_3_adj
- CO2_3_adj
- PAR_3_adj

and additionally:

- Power_W (instantaneous electrical PV power production)

This extended formulation enables a unified modelling of internal climate and PV output within a single multi-target regression structure. Performance metrics are reported in Table 6.7.

Table 6.7 Performance metrics for the XGBoost "Microclimate + PV" model

Variable	R^2	RMSE	MAE
Temp_3_adj (°C)	0.9545	1.9902	1.2934
RH_3_adj (%)	0.9464	4.6821	3.0276

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

Variable	R^2	RMSE	MAE
CO2_3_adj (ppm)	0.7474	12.6111	6.7941
PAR_3_adj (µmol/s)	0.8835	74.9008	42.2392
Power_W (W)	0.8936	269.5848	146.978

The inclusion of PV-related variables results in systematic improvements for temperature, RH, and PAR, confirming the strong local influence of shading and PV heating on microclimate dynamics. Power production is also predicted with high accuracy, supporting the integrated use of the model in the Digital Twin for both agronomic and energy-oriented applications.

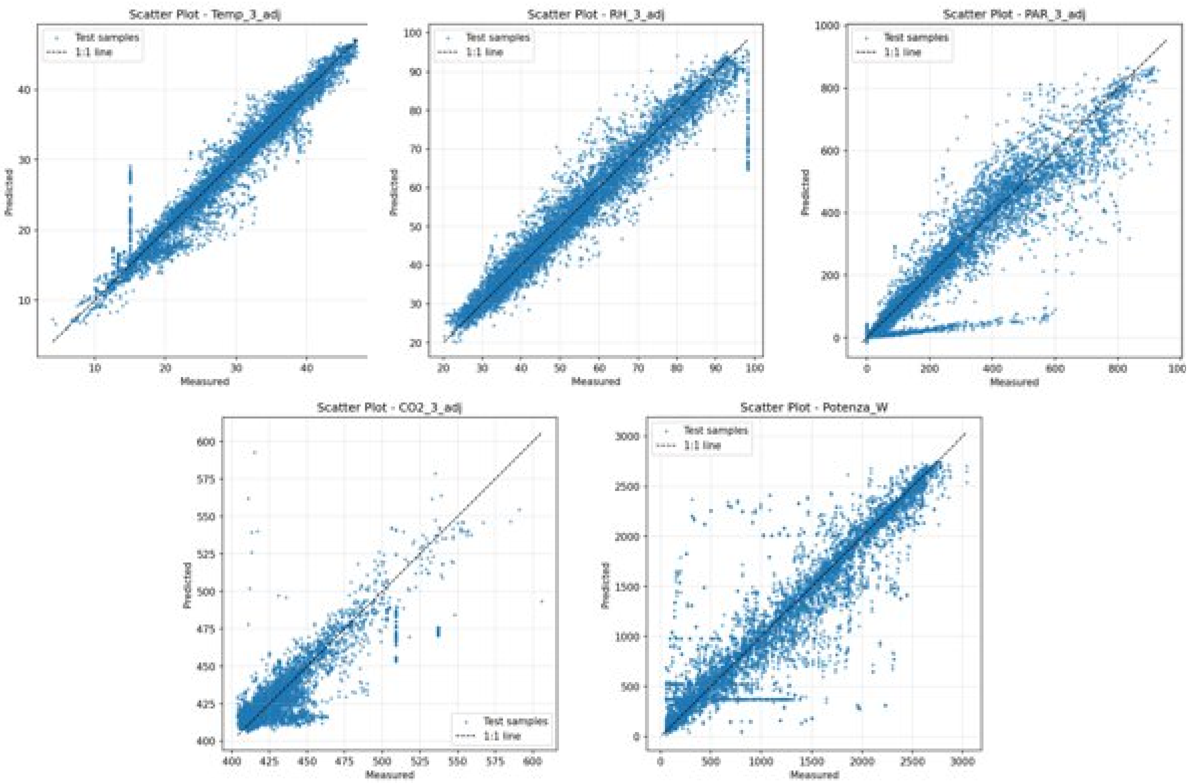


Fig. 6.18 Scatter plots for temperature, relative humidity, CO_2, PAR for pole 3 level adjustable and Power.

The comparative analysis of the two XGBoost configurations demonstrates that both approaches provide reliable and operationally meaningful predictions for greenhouse microclimate variables, even under constrained sensing conditions. The microclimate-only model establishes a robust baseline, confirming that a minimal set of external measurements is sufficient to reproduce the main internal thermal, hygrometric and radiative dynamics. The extended microclimate + PV configuration further improves accuracy, particularly for temperature, relative humidity and PAR, by explicitly capturing the radiative modulation and thermal loading induced by the agrivoltaic system. Moreover, the model achieves strong performance in predicting instantaneous electrical power output, enabling a unified and computationally efficient Digital Twin for both climatic and energy-related processes. Overall, the results indicate that tree-based multi-output models remain a pragmatic and scalable solution for real-case greenhouse scenarios, where sensor availability may be limited but reliable multi-variable predictions are essential for operational monitoring, control and decision-support.

6.7.3.3 Impact of Additional Internal GHI Sensors on Prediction Accuracy

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

Two additional model variants were tested by including internal irradiance measurements (GHI_9 and GHI_10) above the PV panels level as inputs. For the only microclimate configuration, the inclusion of either internal GHI sensor yields small but consistent performance improvements for temperature, relative humidity and PAR. PAR prediction benefits most from the additional information, with R^2 increasing from 0.83 to approximately 0.87 and RMSE reduced by about 10–15%, whereas CO_2 prediction remains substantially unchanged.

More pronounced improvements are observed in the microclimate + PV configuration. The addition of internal GHI measurements leads to higher accuracy for temperature, relative humidity and PAR compared to the corresponding model without internal irradiance inputs. The configuration including GHI_10 achieves the best overall performance for PAR ($R^2 \approx 0.90$, RMSE ≈ 70.5), while maintaining very high accuracy for temperature ($R^2 \approx 0.96$) and relative humidity ($R^2 \approx 0.95$). CO_2 prediction shows only marginal sensitivity to the inclusion of internal GHI sensors across all configurations.
The prediction of instantaneous PV power output remains accurate and comparable among all microclimate + PV variants, indicating that additional internal irradiance measurements primarily enhance microclimate reconstruction rather than power estimation. Overall, these results confirm that supplementing external forcing variables with a limited number of internal GHI sensors improves model performance, particularly for radiative variables, while preserving the robustness of the integrated XGBoost-based Digital Twin framework.

Conclusions

Key Achievements

The REGACE digital twin for photovoltaic greenhouses has successfully demonstrated that data-driven predictive models can capture the complex interactions between environmental forcing, control actions, and system response across multiple operational objectives. Three complementary modeling components were developed, validated, and prepared for operational deployment, each addressing specific requirements of the agrivoltaic greenhouse management problem.

Photovoltaic Power Prediction

The PV power predicting component achieved operational accuracy using a low-cost sensor configuration and XGBoost regression. With mean absolute error below 85 watts and R^2 exceeding 0.94, the model provides reliable power output predictions from standard environmental measurements without requiring extensive electrical instrumentation. The progression from baseline Random Forest to optimized XGBoost demonstrated that feature engineering, particularly temporal lags, rolling statistics, and physics-inspired transformations, accounts for most of the performance improvement. This validates the feasibility of accurate power prediction in commercial greenhouse settings where instrumentation costs must be minimized.
The model successfully tracks diurnal cycles, captures weather-driven variability, and responds appropriately to control actions affecting panel exposure and thermal conditions. Anomaly detection capability enables identification of inverter faults or performance degradation that cause deviations exceeding baseline prediction uncertainty. Computational efficiency permits real-time integration with greenhouse management systems, supporting both short-term operational decisions and longer-term planning through scenario analysis.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

Thermodynamic Modeling with Tree-Based Ensembles

The thermodynamic component established that high-fidelity internal temperature prediction is achievable without explicit physical modeling or dense sensor networks. Tree-based ensemble methods, particularly XGBoost, successfully captured the effective greenhouse thermal response across diverse operating conditions. The delta-learning formulation, predicting internal-external temperature differential rather than absolute values, proved particularly effective for low-cost sensor configurations.

Performance metrics demonstrate practical utility for operational decision support. The low-cost model achieved MAE below 0.3°C with R^2 exceeding 0.99 on delta temperature prediction, despite eliminating most internal sensors and relying primarily on external meteorological measurements plus minimal greenhouse-specific instrumentation. This represents a five-fold reduction in sensor costs while maintaining prediction accuracy rivaling full-sensor configurations. The success validates that strategic feature engineering compensating for reduced measurements through physically motivated derived variables can substitute for direct sensing in many operational scenarios.

Graph Neural Network Architecture

The GNN-based digital twin addressed limitations of site-specific models through explicit representation of coupled greenhouse dynamics. By treating measured variables as nodes in a learned graph structure, the architecture captures inter-variable dependencies while remaining adaptable to different sensor configurations and greenhouse types. The soft-gated Mixture of Experts formulation proved decisive for handling regime heterogeneity without the brittleness of hard regime switching.

After calibration, the GNN achieved MAE of 1.91°C and R^2 of 0.85 on absolute temperature prediction across the full operational range. More importantly, targeted methodological refinements successfully addressed the systematic underestimation of extreme temperatures that plagued earlier approaches. The ablation study quantified contributions of data quality handling, radiation memory features, control dynamics encoding, and tail-aware learning, demonstrating that performance gains arose from physically motivated interventions rather than increased model complexity alone.

The GNN architecture is suitable for multiple deployment modes including short-term predicting, what-if scenario simulation, digital shadowing for anomaly detection, and integration with model predictive control frameworks. The explicit encoding of temporal memory and inter-variable coupling provides interpretability often lacking in black-box deep learning approaches.

Methodological Contributions

Table 6.8 Summary of key achievements by component

Component	Architecture	Key Metric	Sensor Config	Primary Achievement
PV Power	XGBoost	MAE: 85 W, R^2: 0.945	Low-cost (5 sensors)	Reliable predicting from minimal instrumentation
Thermodynamic (tree)	XGBoost + delta	MAE: 0.28°C, R^2: 0.997	Low-cost (5 sensors)	Five-fold sensor reduction without accuracy loss
Thermodynamic (GNN)	Soft MoE + LSTM-GAT	MAE: 1.91°C, R^2: 0.846	Full (50-200 nodes)	Use of Deep Neural Network approach for temperature.
Methodology	Multi-stage	-	Variable	Validated staged development pathway
Data infrastructure	NetCDF + quality	-	All	Reproducible preprocessing, fault

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

	control			detection

Beyond specific model results, the work established several methodological findings applicable to digital twin development for controlled environment agriculture. Sensor quality monitoring proved essential, with flat-line detection preventing model corruption from faulty measurements. Chronological validation consistently outperformed random splitting, emphasizing that temporal dependencies must be respected in performance assessment. Feature engineering systematically dominated algorithm selection in determining predictive capability, suggesting that domain knowledge integration provides greater leverage than architectural sophistication.

The progression from tree-based models to graph neural networks illustrated a productive development pathway balancing accuracy, interpretability, and generalization. Tree-based models provided strong baselines with minimal tuning, enabling rapid feasibility assessment. GNN architectures addressed generalization requirements and extreme-regime behavior at the cost of increased complexity. This staged approach mitigated development risk while ensuring that increased sophistication delivered measurable benefits.

Limitations and Lessons Learned

The primary limitation of the digital twin concerns crop growth modeling. Sparse, manually collected crop measurements (biweekly) required interpolation, producing smoothed target signals that lack meaningful temporal structure. As a result, neither regression nor classification models could reliably learn relationships between environmental conditions and crop development. This limitation reflects a data identifiability issue rather than a modeling failure, highlighting the need for higher-frequency crop observations or hybrid approaches combining mechanistic models with data assimilation.

A second limitation involves systematic underestimation of extreme temperatures (>45 °C). Although tail-aware GNN models reduced this bias, extreme events remain difficult to predict due to their rarity and potentially different governing processes. Operational risks are limited because such events are infrequent and detectable via anomaly alerts, but future work should explore synthetic extreme data and physics-informed constraints.

Table 6.9 Summary of limitations and lessons learned

Limitation	**Root Cause**	**Impact**	**Mitigation Strategy**
Crop prediction failure	Sparse measurements, interpolation artifacts	Cannot infer growth from environment	Higher-frequency sensing or hybrid mechanistic models
Extreme temperature compression	Data scarcity, regime differences	Underestimates >45°C events	Tail-aware learning, scenario simulation
Flat-line sensor faults	Environmental stress on sensors	Model corruption if undetected	Rolling variance detection, quality dashboards
Cross-site generalization	Measurement heterogeneity	Limited transfer learning	Standardization or normalization pipelines
GNN training cost	Deep architecture, large datasets	Retraining barriers	Fine-tuning strategies, incremental updates
Transient response prediction	Equilibrium-dominated training data	Poor performance during rapid changes	Synthetic transient generation, physics constraints

Sensor data quality emerged as a critical operational challenge. Flat-line sensor failures, undetectable by simple range checks, were identified only through residual analysis and temporal variability inspection. This led to the development of variance-based fault detection

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

and underscored the necessity of continuous, automated sensor health monitoring in real deployments.

Regarding generalization, while GNNs improved transferability across sites, full validation across heterogeneous greenhouses is incomplete. Differences in sensor types, placement, and sampling strategies limit direct model transfer, emphasizing the need for measurement standardization or robust normalization pipelines.

Finally, computational cost represents a practical constraint. GNN training requires substantial resources, making frequent retraining challenging in operational settings. Incremental updates or partial fine-tuning offer promising compromises but require further validation.

Overall, these limitations emphasize that successful agricultural digital twins depend as much on data quality, measurement design, and operational constraints as on model architecture.

Concluding Perspective

The REGACE digital twin demonstrates that data-driven predictive modeling can achieve operational maturity for agrivoltaic greenhouse management when development proceeds systematically from clear requirements through staged implementation. The combination of tree-based models for rapid baseline establishment and graph neural networks for sophisticated regime handling illustrates productive integration of complementary approaches rather than commitment to single methodologies.

Success depended critically on clear assessment of data quality, respect for temporal dependencies in validation, and systematic feature engineering encoding domain knowledge. These methodological foundations proved more decisive than architectural sophistication, suggesting that future digital twin development should emphasize data infrastructure and physics-informed feature design before pursuing complex modeling approaches.

The roadmap presented here positions the REGACE digital twin for evolution from research demonstration to operational decision support tool. Implementation priorities balance immediate utility against long-term capability development, recognizing that sustained impact requires not only technical performance but also practical deployment considerations including cost, reliability, and integration with existing workflows. The progression from current achievements through planned enhancements establishes a credible pathway toward transforming greenhouse operations through intelligent automation and optimization.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

Repository Description (D4.3)

A dedicated software repository has been established to support the development, testing, and validation of the Digital Twin models produced within the REGACE project and it consists in the D4.3. The repository is currently hosted on GitHub under a private access configuration at the following address:

https://github.com/esterlab-unitov/regace/tree/main

The repository serves as the central workspace for all modelling activities related to the greenhouse Digital Twin, including data preparation, exploratory analysis, model training, evaluation, and the generation of operational artefacts. Its structure has been designed to ensure full reproducibility of the modelling workflow, while keeping large datasets external to the version-controlled environment due to their size and licensing constraints.

Structure and Content.

The repository contains a collection of Jupyter notebooks that document and operationalise the development of the predictive components used in the Digital Twin. These notebooks include scripts for XGBoost-based temperature and power modelling, multi-output microclimate prediction, Random Forest benchmarking, low-cost sensor configuration experiments, and comparative thermal model analyses. Examples of the main notebooks include:

- XGBoost_Multi_output_2.ipynb for the multi-target modelling framework;
- modello_termico_RMSE.ipynb for thermal model evaluation;
- RF_training_validation_with_RMSE.ipynb for ensemble-method comparison;
- domain-specific notebooks such as DEF_train_zucchini_xgb.ipynb for crop-specific modelling tests.

Alongside the notebooks, the repository includes a set of Python utility modules that standardise data loading, pre-processing, feature generation, and model persistence. These utilities are used internally by the notebooks and ensure consistent execution across experiments and hardware environments.

External Data Management.

Large input datasets required by the Digital Twin, such as the greenhouse netCDF files and the interpolated crop growth data in parquet format, are not stored within the repository. Instead, users are expected to place the required files (e.g., digital_twin_data.nc, interpolated_crop_data.parquet) in the repository root before executing the notebooks. This approach avoids unnecessary duplication and ensures compliance with project data handling policies.

Execution Environment.

The repository is designed to be used with Python 3.9+ and standard scientific computing libraries (NumPy, Pandas, scikit-learn, XGBoost, Matplotlib, Joblib, netCDF4, PyArrow). Dependencies can be installed using a Python virtual environment and the packages listed or referenced in the notebooks. All notebooks can be run in Jupyter, JupyterLab, or compatible development environments such as Visual Studio Code.

Outputs and Artefacts.

Trained models and derived outputs (e.g., .joblib model files, JSON metrics summaries such as digital_twin_multioutput_*.json, and diagnostic figures such as the scatter plots stored in fig_scatter_multioutput_*/) are generated locally during execution but are not committed to

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

the repository. This ensures that only the modelling logic is versioned, while the heavier artefacts remain local to the user environment.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

References

Aaslyng, J.M., Ehler, N., Jakobsen, L., 2005. Climate control software integration with a greenhouse environmental control computer. Environmental Modelling & Software 20, 521–527. https://doi.org/https://doi.org/10.1016/j.envsoft.2003.11.015

Aaslyng, J.M., Lund, J.B., Ehler, N., Rosenqvist, E., 2003. IntelliGrow: a greenhouse component-based climate control system. Environmental Modelling & Software 18, 657–666. https://doi.org/https://doi.org/10.1016/S1364-8152(03)00052-5

Alaimo, L., Seri, E., 2021. Monitoring the main aspects of social and economic life using composite indicators. a literature review. Research Group Economics, Policy Analysis and Language (REAL) series.

Baglivo, C., Mazzeo, D., Panico, S., Bonuso, S., Matera, N., Congedo, P.M., Oliveti, G., 2020. Complete greenhouse dynamic simulation tool to assess the crop thermal well-being and energy needs. Appl. Therm. Eng. 179, 115698. https://doi.org/https://doi.org/10.1016/j.applthermaleng.2020.115698

Belhaj Salah, L., Fourati, F., 2021. A greenhouse modeling and control using deep neural networks. Applied Artificial Intelligence 35, 1905–1929. https://doi.org/10.1080/08839514.2021.1995232

Biau, G., Scornet, E., 2016. A random forest guided tour. Test 25, 197–227. https://doi.org/10.1007/s11749-016-0481-7

Brækken, A., Sannan, S., Jerca, I.O., Bădulescu, L.A., 2023. Assessment of heating and cooling demands of a glass greenhouse in Bucharest, Romania. Thermal Science and Engineering Progress 41, 101830. https://doi.org/https://doi.org/10.1016/j.tsep.2023.101830

Bussab, M.A., Bernardo, J.I., Hirakawa, A.R., 2007. Greenhouse modeling using neural networks, in: World Scientific and Engineering Academy and Society (WSEAS), AIKED'07. Stevens Point, Wisconsin, USA, pp. 131–135. https://doi.org/10.5555/1348485.1348508

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 13-17-August-2016, 785–794. https://doi.org/10.1145/2939672.2939785/SUPPL_FILE/KDD2016_CHEN_BOOSTING_SYSTEM_01-ACM.MP4

Cheung, Y.-W., Lai, K.S., 1995. Lag Order and Critical Values of the Augmented Dickey-Fuller Test. Journal of Business & Economic Statistics 13, 277–280.

Dae-Hyun, J., Hyoung, S.K., Changho, J., Hak-Jin, K., Soo, H.P., 2020. Time-serial analysis of deep neural network models for prediction of climatic conditions inside a greenhouse. Comput. Electron. Agric. 173, 105402. https://doi.org/https://doi.org/10.1016/j.compag.2020.105402

Ferreira, P.M., Faria, E.A., Ruano, A.E., 2002. Neural network models in greenhouse air temperature prediction. Neurocomputing 43, 51–75. https://doi.org/https://doi.org/10.1016/S0925-2312(01)00620-8

Fourati, F., Chtourou, M., 2007. A greenhouse control with feed-forward and recurrent neural networks. Simul. Model. Pract. Theory 15, 1016–1028. https://doi.org/https://doi.org/10.1016/j.simpat.2007.06.001

Gharghory, S.M., 2020. Deep Network based on Long Short-Term Memory for Time Series Prediction of Microclimate Data inside the Greenhouse. Int. J. Comput. Intell. Appl. 19, 2050013. https://doi.org/10.1142/S1469026820500133

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

Guo, Y., Zhao, H., Zhang, S., Wang, Y., Chow, D.H.C., 2020. Modeling and optimization of environment in agricultural greenhouses for improving cleaner and sustainable crop production. J. Clean. Prod. 124843.

Hongkang, W., Li, L., Yong, W., Fanjia, M., Haihua, W., Sigrimis, N.A., 2018. Recurrent Neural Network Model for Prediction of Microclimate in Solar Greenhouse. IFAC-PapersOnLine 51, 790–795. https://doi.org/10.1016/j.ifacol.2018.08.099

Jain, L.C., Medsker, L.R., 1999. Recurrent Neural Networks: Design and Applications, 1st ed. CRC Press, Inc., USA. https://doi.org/10.5555/553011

James, G., Witten, D., Hastie, T., Tibshirani, R., others, 2013. An introduction to statistical learning. Springer. https://doi.org/10.1007/978-1-0716-1418-1

Jin, M., Koh, H.Y., Wen, Q., Zambon, D., Alippi, C., Webb, G.I., King, I., Pan, S., 2024. A Survey on Graph Neural Networks for Time Series: Forecasting, Classification, Imputation, and Anomaly Detection. IEEE Trans. Pattern Anal. Mach. Intell. 46, 10466–10485. https://doi.org/10.1109/TPAMI.2024.3443141

Manonmani, A., Thyagarajan, T., Elango, M., Sutha, S., 2018. Modelling and control of greenhouse system using neural networks. Transactions of the Institute of Measurement and Control 40, 918–929.

Nielsen, M.A., 2015. Neural networks and deep learning. Determination press San Francisco, CA, USA.

Nketiah, E.A., Chenlong, L., Yingchuan, J., Aram, S.A., 2023. Recurrent neural network modeling of multivariate time series and its application in temperature forecasting. PLoS One 18, e0285713. https://doi.org/10.1371/journal.pone.0285713

Petrakis, T., Kavga, A., Thomopoulos, V., Argiriou, A.A., 2022a. Neural Network Model for Greenhouse Microclimate Predictions. Agriculture 12. https://doi.org/10.3390/agriculture12060780

Petrakis, T., Kavga, A., Thomopoulos, V., Argiriou, A.A., 2022b. Neural Network Model for Greenhouse Microclimate Predictions. Agriculture.

Rodriguez, P., Wiles, J., Elman, J.L., 1999. A Recurrent Neural Network that Learns to Count. Conn. Sci. 11, 5–40. https://doi.org/10.1080/095400999116340

Salah, L.B., Fourati, F., 2021. A greenhouse modeling and control using deep neural networks. Applied Artificial Intelligence 35, 1905–1929.

Seri, E., Petitta, M., Papaioannou, C., Katsoulas, N., Cornaro, C., 2025. Sustainable greenhouse microclimate modeling: A comparative analysis of recurrent and Graph Neural Networks. Build. Environ. 284. https://doi.org/10.1016/j.buildenv.2025.113473

Sun, L., Liu, T., Wang, D., Huang, C., Xie, Y., 2022. Deep learning method based on graph neural network for performance prediction of supercritical CO2 power systems. Appl. Energy 324, 119739. https://doi.org/https://doi.org/10.1016/j.apenergy.2022.119739

Wang, C., Deng, C., Wang, S., 2020. Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. Pattern Recognit. Lett. 136, 190–197. https://doi.org/10.1016/J.PATREC.2020.05.035

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Overall Conclusions and Key finding

This deliverable has presented the development and validation of an integrated predictive simulation framework for photovoltaic greenhouse systems within the REGACE project. The activities carried out in the project demonstrate how the combination of deterministic physical models and data-driven approaches can effectively address the complexity of agrivoltaic greenhouses, where energy production, microclimate regulation, water management, and crop growth are intrinsically interconnected.

The modelling framework was developed and tested using extensive datasets collected at the FSC and UTH pilot sites, representing real operational conditions rather than controlled laboratory environments. This allowed the models to capture the actual behaviour of PV-integrated greenhouses, including the effects of photovoltaic shading, climate control strategies, and crop–environment interactions.

Overall, the work confirms that predictive modelling and Digital Twin concepts can provide robust support for the analysis, interpretation, and future optimization of responsive agrivoltaic greenhouse systems.

Key Findings

A key outcome of this work is the demonstration that accurate prediction of greenhouse microclimate dynamics requires an explicit and dynamic representation of crop-related processes. The dynamic simulation models developed show that air temperature and relative humidity are strongly governed by plant evapotranspiration and canopy development. Introducing time-dependent parameters such as Leaf Area Index and crop conductance resulted in a more realistic model with reasonable accuracy.

CFD modelling provided deeper insight into the spatial variability of radiation, airflow and temperature, inside PV-equipped greenhouses. These simulations proved particularly useful for analysing the effects of photovoltaic panels on shading patterns, ventilation efficiency, and local microclimatic conditions, supporting the interpretation and refinement of lower-order dynamic models.

In parallel, a dedicated performance model for bifacial photovoltaic modules was developed and validated using real monitoring data. The results underline the importance of accounting for greenhouse-specific operating conditions, such as reflected irradiance, greenhouse structure shading and thermal behaviour, to reliably estimate energy production in agrivoltaic systems.

Crop modelling activities showed that established models such as AquaCrop can be effectively applied to PV greenhouses when adequate microclimate and irrigation data are available. Where crop data were more limited, complementary statistical approaches enabled the identification of meaningful relationships between environmental drivers and crop response.

Finally, the integration of experimental data into machine learning–based Digital Twin models represents a major achievement of project. Recurrent neural networks, gradient-boosting techniques, and spatio-temporal graph neural networks demonstrated strong predictive capabilities, highlighting the potential of machine learning modelling approaches for complex greenhouse systems.

The LOW-COST model, using only external climate variables as input, reliably predicts greenhouse microclimate conditions. When augmented with a limited number of variables related to solar irradiance on the plane of the array, temperature of the panels, the model can accurately predict both key internal microclimatic parameters and PV energy production. These results demonstrate the model's applicability in real operational settings while requiring only a minimal sensor setup. Furthermore, crop yield can be estimated through a cascading

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

approach, using the predicted microclimatic variables as inputs for AquaCrop or simplified crop models.

Concluding Remarks

The work presented in this deliverable highlights the added value of combining physics-based modelling with data-driven techniques in the context of agrivoltaic greenhouse systems. Rather than treating microclimate, energy production, water use, and crop growth as independent problems, the REGACE modelling framework adopts a systemic perspective, reflecting the real operational interdependencies observed in the pilot sites.

The results confirm that Digital Twin concepts are not only suitable for post-analysis and performance assessment but also hold strong potential as future decision-support tools. By integrating real-time data streams with predictive models, such frameworks can support more informed management strategies, facilitate the evaluation of design and control alternatives, and ultimately contribute to improving the sustainability and resilience of greenhouse agrivoltaic solutions.

Within this context, the work presented establishes a solid methodological basis upon which further developments can be built, bridging the gap between advanced modelling research and practical applications in real-world greenhouse operation.

Limitations and Recommendations for Future Work

Despite the encouraging results, several limitations were identified during the development of the modelling framework. One of the main constraints concerns crop growth modelling, which remains highly dependent on the availability, frequency, and accuracy of crop measurements. In this work, crop data were largely based on manual and periodic observations, limiting the temporal resolution required for fully dynamic and data-driven crop growth modelling.

Another limitation relates to data availability and heterogeneity across pilot sites. The most comprehensive modelling and validation activities could only be carried out at sites equipped with dense and reliable monitoring infrastructures. This highlights the need for standardized data acquisition protocols and scalable sensor solutions to improve model transferability.

From a computational perspective, high-fidelity CFD models, while providing valuable physical insight, are not suitable for real-time applications due to their computational cost. Future research should therefore focus on developing reduced-order or surrogate models capable of retaining essential physical behaviour while enabling faster execution.

Finally, although the Digital Twin models developed exhibit strong predictive performance, their integration into real-time control and optimization strategies—such as adaptive climate control, irrigation scheduling, and PV tracking—was beyond the scope of this deliverable. Future work should aim to close this gap by coupling predictive models with control algorithms and decision-support platforms.

In conclusion, future research should prioritize continuous crop sensing, tighter integration between predictive models and operational control systems, and broader validation across different climates, greenhouse typologies, and crop species. These steps will be essential to fully exploit the potential of Digital Twins as operational tools for sustainable and productive agrivoltaic greenhouse systems.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

ANNEX

Annex A – Detailed mathematical formulations and UDF descriptions (Chapter 3)

A1. Calculation of external radiation on roof cover for validation

For the models' validation external radiation normal to each cover computational cell should be defined with respect the total and diffusive radiation normal to horizontal plane out of the greenhouse (values taken from meteorological station). So, for each point of the cover should be calculated four values of beam radiation and four values of diffusive radiation (for the four considered wavelength bands):

$$I_{n,beam,UV} = UV_{fraction} * I_{n,bean,\beta} * Flag_{shadow}$$

$$I_{n,beam,PAR} = PAR_{fraction} * I_{n,bean,\beta} * Flag_{shadow}$$

$$I_{n,beam,NIR} = NIR_{fraction} * I_{n,bean,\beta} * Flag_{shadow}$$

$$I_{n,beam,IR} = IR_{fraction} * I_{n,bean,\beta} * Flag_{shadow}$$

$$I_{beam,0,UV} = UV_{fraction} * I_{bean,0} * Flag_{shadow}$$

$$I_{beam,0,PAR} = PAR_{fraction} * I_{bean,0} * Flag_{shadow}$$

$$I_{beam,0,NIR} = NIR_{fraction} * I_{bean,0} * Flag_{shadow}$$

$$I_{beam,0,IR} = IR_{fraction} * I_{bean,0} * Flag_{shadow}$$

The wavelength fraction in the general case is defined as 0.05/0.45/0.28/0.22
The 'Flag' for the shading is calculated through UDF (will be presented later)
The beam, diffusive and reflected components each moment for each roof point is calculated as follow (Duffie 2006).

$$I_{n,beam,\beta} = G_{n,beam,0} * R_{b,\beta}$$

$$I_{n,dif,\beta} = \frac{(1+\cos(\beta))}{2} * G_{n,dif,0}$$

$$I_{n,ref,\beta} = \frac{(1-\cos(\beta))}{2} * G_{tot,0}$$

$$R_{b,\beta} = \frac{\cos(\theta)}{\cos(\theta_z)}$$

$$\cos(\theta) = (A - B) * \sin(\delta) + [\mathrm{C} * \sin(\omega) + (\mathrm{D} + \mathrm{E}) * \cos(\omega)] * \cos(\delta)$$

$$\cos(\theta_z) = \cos(\varphi) * \cos(\delta) * \cos(\omega) + \sin(\varphi) * \sin(\delta)$$

$$A = \sin(\varphi) * \cos(\beta)$$

$$B = \cos(\varphi) * \sin(\beta) * \cos(\gamma)$$

$$C = \sin(\beta) * \sin(\gamma)$$

$$D = \cos(\varphi) * \cos(\beta)$$

$$E = \sin(\varphi) * \sin(\beta) * \cos(\gamma)$$

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

For these calculations it is necessary to know the inclination angle, β, for each roof point while the azimuth angle is easier to define.

$$\delta = 23.45 * \sin\left(\frac{360*(284+n)}{365}\right) = 13.88$$

$$\omega_s = acos\left(-\tan\varphi * \tan\delta\right) = 101.77$$

$$t_{sr} = 12 - \frac{\omega_s}{15} = 5.22 \text{ h} = 5\text{h } 13' \ 12''$$

$$t_{ss} = 12 + \frac{\omega_s}{15} 18.78 = 18 \text{ h } 46' \ 48''$$

$$w = (15h^{-1})(t_{solar} - 12h)$$

$$t_s = t_z + E \pm 4\frac{\min}{\deg}\left(L_{st} - L_{loc}\right)$$

The gothic roof consists of two arcs that converge on the top. The inclination angle, β, in a point (x_1, y_1) for an arc with center (x_0,y_0) is calculated as follow

$$\beta = -\frac{(x_1 - x_0)}{(y_1 - y_0)}$$

Table A.1 Characteristics of 12 arc constitute the six chambers roof arcs

Section	X1 - start	X2- end	X_0	Y_0	γ [deg]
1st left	0	4.85	4.8	1.49	-54
1st right	4.85	9.7	4.9	1.49	126
2nd left	9.7	14.55	14.5	1.49	-54
2nd right	14.55	19.4	14.6	1.49	126
3rd left	19.4	24.25	24.2	1.49	-54
3rd right	24.25	29.1	24.3	1.49	126
4th left	29.1	33.95	33.9	1.49	-54
4th right	33.95	38.8	34	1.49	126
5th left	38.8	43.65	43.6	1.49	-54
5th right	43.65	48.5	43.7	1.49	126
6th left	48.5	53.35	53.3	1.49	-54
6th right	53.35	58.2	53.4	1.49	126

A2. Regression analysis

The goal is to create an equation $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \varepsilon$, with respect thw two considered independent variables, X_1 and X_2, where β_0 is the constant term (intercept), β_1 and β_2 the coefficients, and ε the error term. The estimation of the coefficients is done using the method of least squares (Ordinary Least Squares – OLS). The quality and statistical validity of a regression model is assessed through specific criteria. One of the most well-known is the coefficient of determination R^2. R^2 expresses the percentage of the variance of the dependent variable. However, R^2 alone is not sufficient to evaluate a model, especially when the number of independent variables increases. For this reason, the adjusted R^2 is often used, which takes into account the number of variables and the sample size. A particularly important evaluative characteristic is the Significance F, which results from the overall F-test of the model. In addition, the individual regression coefficients are evaluated through their p-values.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

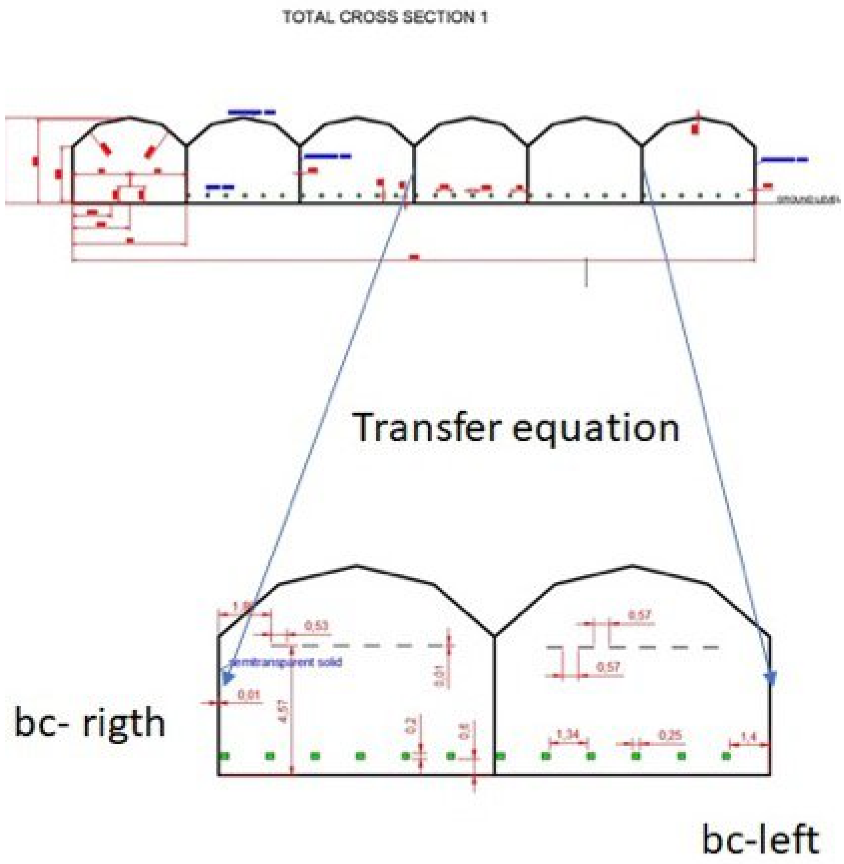


Fig. A.1 Transfer equation operation

Statistics of regression analysis

Table A.2

Case	Left PAR	Left NIR	Right PAR	Right NIR
R^2	0.93	0.98	0.96	0.98
Adjusted R^2	0.92	0.98	0.91	0.96
Significance – F	$3x10^{-20}$	$4.4x10^{-31}$	$4.25x10^{-19}$	$1.5x10^{-25}$
P value	$0.92/0.0013/\ 1.5x10^{-18}$	$0.15/0.21/4.2x10^{-29}$	$0.71/0.04/1.1x10^{-17}$	$0.51/0.87/7.8x10^{-24}$

A3. Energy balance model for ground temperature UDF

The theoretical energy balance is that radiation falls on the ground, part of which is reflected and part is absorbed, exchanges thermal radiation with the air inside the greenhouse, loses heat to the air of the greenhouse by convection, and transfers heat to the ground below by conduction. I mean

$$G_T \cdot a_m + F \cdot \sigma \cdot T_{sky}^4 + h_{ext} \cdot (T_a - T_s) = F \cdot \sigma \cdot \varepsilon_m \cdot T_s^4 + \frac{k}{z} \cdot (T_s - T_{ins})$$

Where:
G_T : the incident solar radiation
α_m : ground absorption coefficient
F: view factor = 1 inside greenhouse
σ: Stefan – Boltzman constant
h_{ext} : convection coefficient between ground and air inside the greenhouse

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

T_α : air temperature inside the greenhouse
T_s : ground surface temperature
ε_m : ground emission coeficeint
k : ground heat transfer coefficient
z : depth at which the ground temperature is known
T_{ins} : ground temperature at depth z

Since CFD can give the thermal incident radiation on the ground, R_{int} the resulting energy balance can be written as follow

$$R_{int} + h_{ext} \cdot (T_a - T_s) = \frac{k}{z} \cdot (T_s - T_{ins})$$

Data and properties:
k = 1.4 [W/mK] (3)
z = 3 m
ρ (of the ground) = 1800 [kg/m^3] (3)
Cp of the ground = 1600 [J/kgK] (3)
n : day of the year
n_0 : day of the year with the lower ground temperature $n_0 = 30$
T_{max_av} = average maximum air temperature
T_{min_av} = average minimum air temperature
T_m = yearly average ground temperatureμέση ετήσια θερμοκρασία εδάφους

CFD field input:
R_{int} = IR solar radiation incident on the ground [W/m^2]
h_{ext} = ground convection coefficient with air [W/m^{2}K]
T_α = Air temperature at the point at the edge of the thermal boundary layer (5 cm from ground surface)

Justification of 5 cm:
Thickness of thermal boundary layer, δ_T

$$\delta_T = 5 \cdot \sqrt{\frac{x \cdot v}{u_o}} \cdot Pr^{-1/3}$$

X: downstream distance (for chamber width 9 m) = 4.5m
ν : air kinematic viscosity at 20 °C = 1.51 x 10^{-5}
u_o : free stream velocity (~ 1 m)
Pr = 0.7 for air

Intermediate calculations [ASHRAE FUNDAMENTALS]
Auxiliary variable, As

$$A_s = \frac{T_{max,av} + T_{min,av}}{2}$$

Thermal difusion, α:

$$\alpha = \frac{k}{\rho \cdot C_p}$$

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

$$T_{ins} = T_m - A_s \cdot exp\left[-z \cdot \sqrt{\left(\frac{\pi}{365 \cdot a}\right)}\right] \cdot cos\left[\left(\frac{2\pi}{365}\right) \cdot \left(n - n_0 - \frac{z}{2} \cdot \sqrt{\frac{365}{\pi \cdot \alpha}}\right)\right]$$

$$T_s = \frac{R_{in} + T_a \cdot h_{ext} + \left({}^{k}/_{z}\right) \cdot T_{ins}}{\left({}^{k}/_{z}\right) + h_{ext}}$$

A4. External air temperature daytime (Parton 1981)

Input from climatic data

- T_av_max
- T_av_min
- t_solar [h]
- N_day: day time hours
- α : lag coefficient of the maximum temperature. At heigth 1.5 m α=1.86 h
- t_sr

Results:

- External air temperature at the specific time T_ext
- Equivalent sky temperature, T_sky
- Air temperature at the sunset time, T_sun_set

$$T_{ext} = \left(T_{av,max} - T_{av,min}\right) * sin\left(\frac{\pi * (t_{solar} - t_{sr}}{N_{day} + 2a}\right) + T_{av,min}$$

$$T_{sky} = 0.0552 * T_{ext}^{1.5}$$

$$T_{sun,set} = \left(T_{av,max} - T_{av,min}\right) * sin\left(\frac{\pi * (t_{ss} - t_{sr}}{N_{day} + 2a}\right) + T_{av,min}$$

A5. External air temperature nighttime (Parton 1981)

Input:

- T_sun_set
- b: 2.2
- Sunset time, t_ss
- t_solar [h]

Results

- Hours passed since the sunset, n_sun_set
- Nighttime hours, N_night
- T_ext
- T_sky

$$n_{sun,set} = t_s - t_{ss}$$

$$N_{nigth} = 24 - N_{day}$$

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

$$T_{ext} = T_{av,min} + \left(T_{sun,set} - T_{av,min}\right) * exp\left[-\frac{b * n_{sun,set}}{N_{night}}\right]$$

A5. Air optical properties calculation

Air is basically a mixture of water vapor and dry air. Its optical properties depend on the percentage of water vapor (Hubbell). For a mixture the extinction coefficient is calculated by:

$$\alpha_s = \sum_{i=1}^{n} w_i \alpha_{si}$$

where W is the weight percentage of element I equal to w/(1+w), where w the absolute humidity [kg/kg] for the greenhouses.
According to the (Mehra) the refractive index of a mixture is according to the Lorenz-Lorenz relationship:

$$\frac{n_{12}^2 - 1}{n_{12}^2 + 2} = \varphi_1 \frac{n_1^2 - 1}{n_1^2 + 2} + \varphi_2 \frac{n_2^2 - 1}{n_2^2 + 2}$$

And since for dry air n_1=1, the above relationship becomes

$$n_{12} = \sqrt{\frac{2\varphi_2 \frac{n_2^2 - 1}{n_2^2 + 2} + 1}{1 - \varphi_2 \frac{n_2^2 - 1}{n_2^2 + 2}}}$$

Where φ is the volume content equal to (1/ρ_dry_air)/(ux(1-w)
Dry air optical properties are α_s =0 and n=1. Water vapor optical properties according to (Guzzi) are for PAR α_s = 0.046 and for NIR α_s =0.121 and n=1.333.

A7. Plants porosity model

As already pointed out plants are considered as porous material composed by air and solid (leaves, fruit, stem). As porous material cause pressure drop adding a sink term in the momentum equations and also modify the heat transfer coefficient in the energy equation as well as the effective viscosity in the turbulence equations (Fidaros 2010, Baxevanou 2018, Baxevanou 2020).

$$S_i = -\left(\frac{\mu}{\alpha} \cdot u_i + C_2 \cdot \rho \cdot u_i^2\right)$$

μ : viscosity [Pas=kg/ms]
α : porous material permeability [m^2]
C_2 : inertial resistance factor [1/m]
u_i : air velocity [m/s]
ρ : air density [kg/m^3]
And using Molina (2006) tables and calculations

$$C_2 = LAD \cdot C_D = \frac{C_f}{\sqrt{a}}$$

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

$$\left(1/a\right) = \left(\frac{C_D \cdot LAD}{C_f}\right)^2 = (27.96 \cdot LAD)^2$$

As far it considers the equivalent viscosity used in the diffusive terms, μ_e.

$$\mu_e = \mu_r \cdot \mu$$

μ_r : relative viscosity
So the effective conductivity keff is calculated for the in the energy equation
Effective conductivity, k_{eff}

$$k_{eff} = \gamma \cdot k_f + (1-\gamma) \cdot k_s$$

γ : porosity
k_f : fluid thermal conductivity
k_s : solid thermal conductivity

A9. Plants evapotransiration model

Plants are sinks of heat and sources of vapor (Villarreal-Guerrero 2012, Boulard 2017, Fatnassi 2023, Katsoulas 2019).
Volumetric sink term for energy equation:
Input data:

- LAD [1/m]
- Effective absorptance coefficient of the plant, a_{eff} [%]
- γ_p = 67.17 [Pa/k]

CFD field data:

- ρ = air density [kg/m^3]
- Cp = air specific heat capacity
- $R_{incident\_PAR}$ = PAR radiation [W/m^3]
- h_c = local convection coefficient [W/m^{2}K]
- ω = absolute humidity [kg/kg]
- Tc = local air temperature [C]
- P = local air pressure [Pa]

Intermediate calculations:
Absorbed radiation, R_{abs} [W/m3]

$$R_{abs} = R_{incident\_PAR} \cdot a_{eff}$$

Calculation of aerodynamic resistance, r_a [s/m]

$$r_a = \frac{\rho \cdot C_p}{h_c}$$

Calculation of stomatal resistance [s/m]

$$r_s = (104.8 + 497.8 \cdot \exp\ (-0.037 \cdot R_{abs})$$

Calculation of vapor pressure deficit

$$VPD_a = e_s(T_c) - e_a\ [Pa]$$

Where,

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

$$e_s = 6.1094xexp\left[\frac{17.625xT_c}{T_c + 243.5}\right]$$

$$e_a = \frac{\omega \cdot P}{0.622 + \omega}$$

Calculation of slope of the saturation vapor pressure [Pa/K]

$$\Delta = \frac{250 \cdot exp\left[\frac{17.27 \cdot T_c}{T_c + 237.2}\right]}{(T_c + 237.3)^2}$$

Final calculation

Volumetric sink term

$$L_v E = \frac{\Delta \cdot R_{abs} + 2 \cdot \rho \cdot LAD \cdot C_p \cdot \left({VPD_a}/{r_a} \right)}{\Delta + 2 \cdot \gamma_p \cdot \left(1 + \frac{r_s}{r_a}\right)}$$

Volumetric source term for water vapor conservation equation
Input Data
L_v : Heat of vaporisation of water, in J/kg = 2440000 J/kg at 20 °C
Calculation of source term
E: Evaporated water flux, in kg/m^{3}s

$$E = \frac{S_{tr}}{L_v}$$

A10. Axial fans modeling

As far as it concerns the axial exhaust fans are modelled as pressure jump calculated from polynomials according to their rotational speed.

For 1st rotational speed, 386 rpm: $\Delta P = -0.274u^2 + 2.58678u + 33.819$

For 2nd rotational speed, 427 rpm : $\Delta P = -0.19u^2 + 2.5119u + 99.458$

A11. Evaporative pads modeling

The evaporative pads are porous material that causes pressure drop and also sinks of heat and sources of water vapor as explained in the (Fidaros2018). The properties and the aerodynamic behavior of the evaporative pads are calculated analytically (Franco 2011). The porosity, θ, is given as

$$\theta = k\left(l_e/l\right)^a\left(1 + Q_W^b\right)$$

Where k, a and b constants, l_e, the pad characteristic length, l the pad width and Q_W , the water flow [kg s^{-1} m^{-2})]. The evaporative pad is fed from a water pad with maximum flow rate 9 m^3/h. According to the values acquired from (Franco et al. 2011), Tables 1 and 2, the porosity value

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

is 0.947, while for dry panel is 0.965. The viscous and inertial resistance coefficients are taken from the table 3 of Franco (2011), $1/a=3.3x10^6$ [m^{-2}] and $C_2=1.13x10^{-4}$ [m^{-1}].
The evaporative pads are also sinks of heat, involving a sensible and a latent heat term (Carmago 2003, Kloppers 2005, Kröger 2004).

$$dQ = dQ_L + dQ_S$$

Where δQ, is the total heat removed from the entering air, δQ_S the sensible heat and δQ_L the latent heat.

$$dQ_L = h_{LVS} h_m (w_s - w) dA$$

Where h_{LVS} is the specific enthalpy of vaporization [J/kg], h_m is the mass transfer coefficient [kg m^{-2} s^{-1})], w_s is the saturated absolute humidity [kg kg^{-1}], and w is the absolute humidity [kg kg^{-1}].

$$Q_S = h_c (T - T_s) dA \qquad (7)$$

Where, h_c the convection heat transfer coefficient, T the entering air temperature and T_s the surface water temperature. Calculating the above terms for the specific evaporative pad geometry and the climatic conditions of the examined day and hour the heat sink is 512 W m^{-2} of the surface where heat and mass exchange takes place. This value corresponds to 100% effectiveness of evaporative pad. the maximum operational effectiveness is limited to 60%, so this limit is taken into account to our calculations.

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Annex B – Computational grids and numerical parameters (Chapter 3)

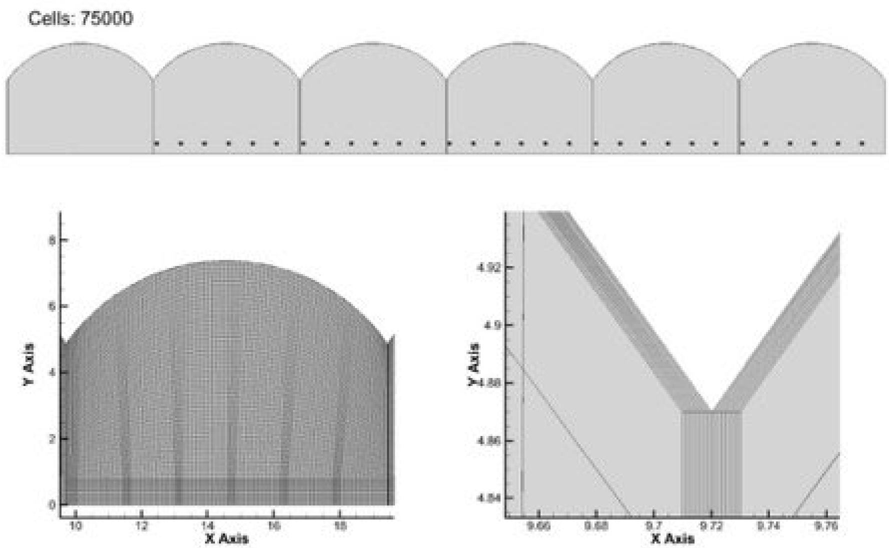


Fig. B.1 2D grid of 6 chambers cross section.

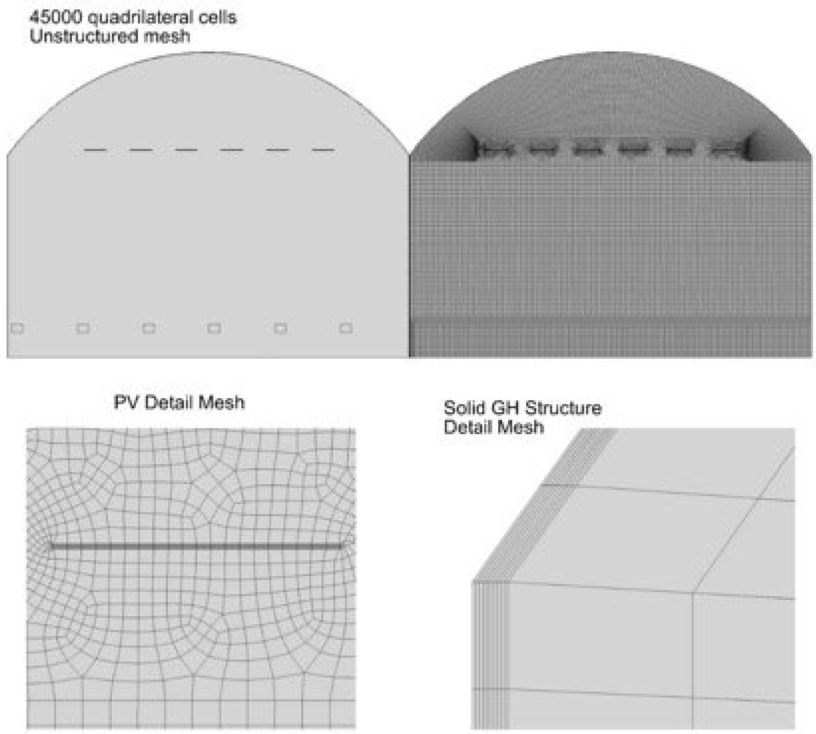


Fig. B.2 2D grid of 2 chambers with the PVs horizontal

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

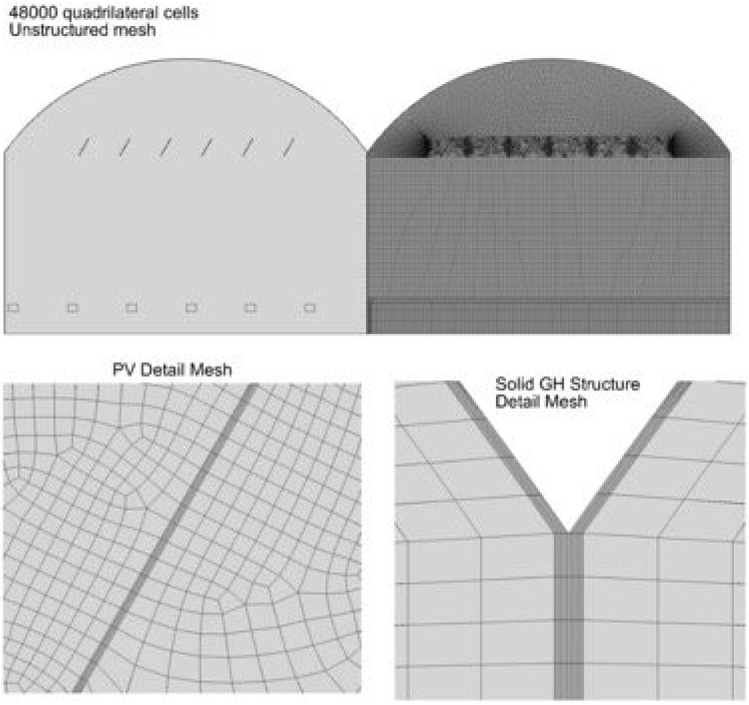


Fig. B.3 2D grid of 2 chambers with PVs inclined to the east

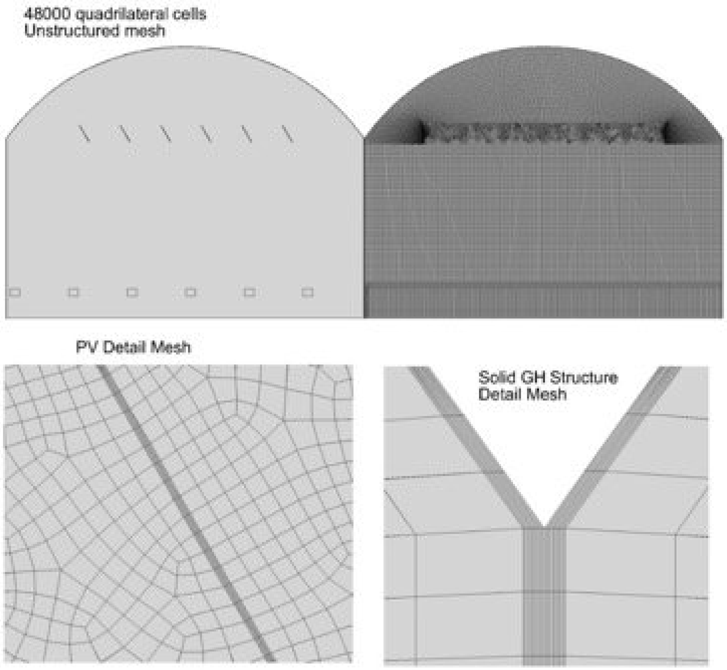


Fig. B.4 2D grid of 2 chambers with PVs inclined to the west

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

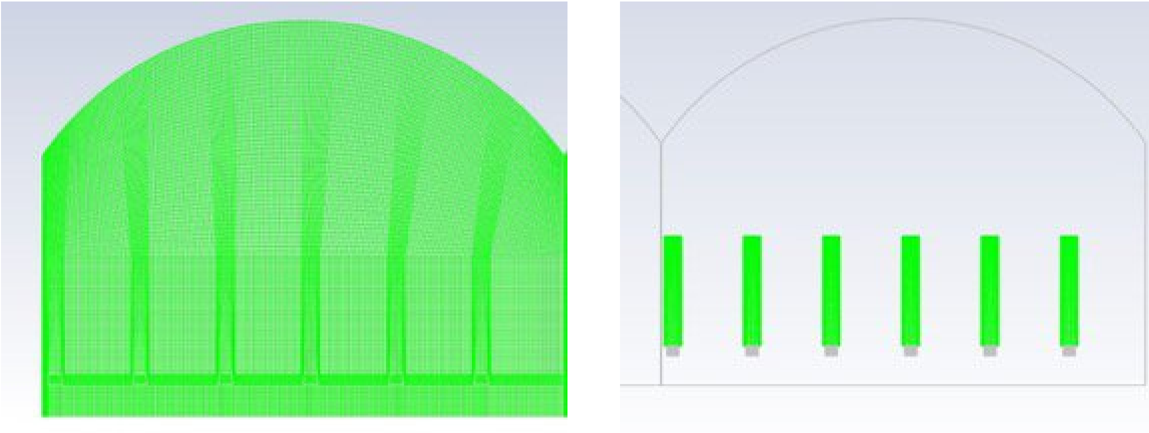


Fig. B.5 2D grid of 2 chambers with plants

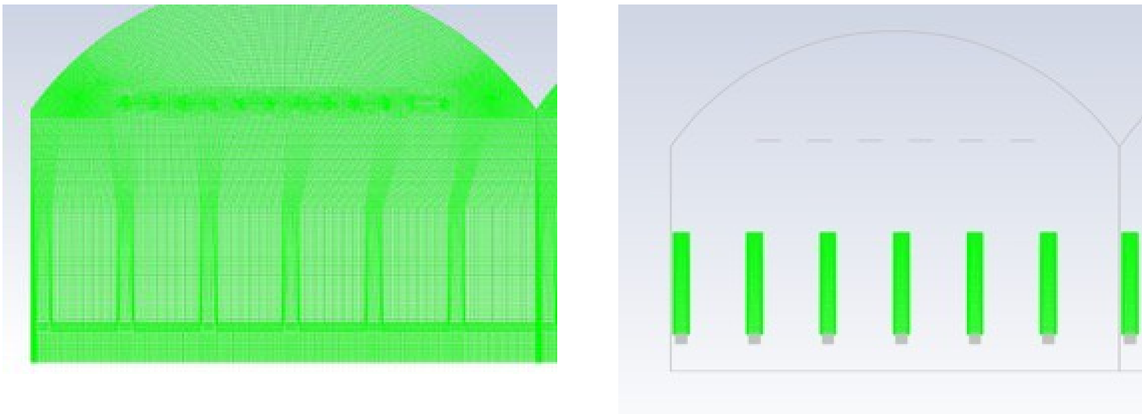


Fig. B.6 2D grid of 2 chambers with plants and horizontal PV panels

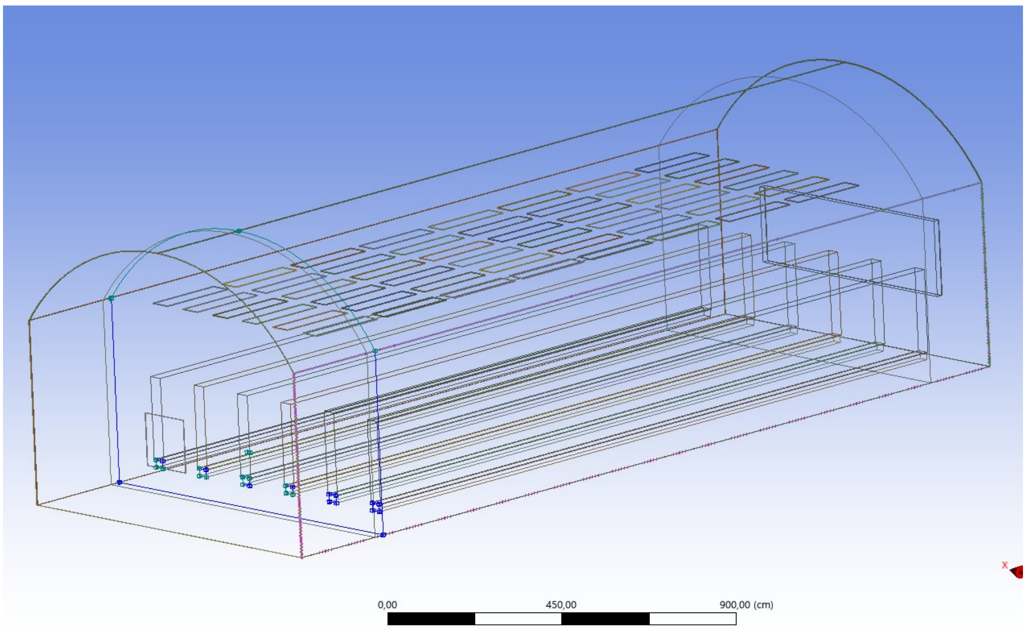


Fig. B.7 3D grid of one chamber with plants and horizontal PV panels

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

Annex C – Material thermophysical and optical properties (Chapter 3)

In Table C.1 materials thermophysical properties are given.

Table C.1 Materials thermophysical properties [1-8]

Material	Density, ρ [kg/m^3]	Specific heat capacity, Cp [J/kgK]	Thermal conductivity, λ [W/mK]
Air	1.25	1006	0.0242
Cover	930	2300	0,39
Side walls	1200	1200	0,2
Bench metal	97.45	901	0.025
Bench with substrate	800.25	2904	0.6
Ground – geotextile	880	1550	0.4
PVs	951	895	0.6
Plants	1300	3421	0.51
Aluminium	2739	896	222
Rock wool	730	3183	0.6

For the air additionally used the viscosity equal to 1.7894x10^{-5} and the thermal expansion coefficient equal to 0.00343 [1/K]

Some of the materials are complex and their properties have been calculated according to their volume composition of simpler materials. The so-called bench metal corresponds to naked bench (without substrate) which is practically a hollow box of steel where the composed by 2% steel and air. The bench with substrate is composed by the naked bench covered by the rock wool substrate of 0.2 m. In the

Especially the PV panels are composed of two structures. The structure I that contains the PV material and the structure II that is transparent. In Tables C.2 and C.3 the composition of the two structures are given along with their thermophysical properties. The structures' details have been provided by TRISOLAR.

Table C. 2 PV structure I and thermophysical properties (Baxevanou 2020)

a/a	name	thickness [mm]	density, ρ [kg/m3]	thermal conductivity, k [W/mK]	Cp [J/kgK]
1	tempered glass	2	940	1,1	720
2	POE -EVA	0,5	880	0,22	1670
3	dark-blue antireflective film	0,06	2500	1,05	750
4	PV monocrystaline silicon	0,06	700	0,18	1004
5	dark-blue antireflective film	0,06	2500	1,05	750
6	POE -EVA	0,5	880	0,22	1670
7	tempered glass	2	940	1,1	720

Table C.3 PV structure II and thermophysical properties

a/	name	thickness	density, ρ	thermal conductivity, k	Cp

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

a		[mm]	[kg/m3]	[W/mK]	[J/kgK]
1	tempered glass	2	940	1,1	720
2	POE -EVA	0,5	880	0,22	1670
6	POE -EVA	0,5	880	0,22	1670
7	tempered glass	2	940	1,1	720

For the plants it is considered that they are composed by fruit (10%), the stem (10%) and the leaves (80%) with the thermophysical properties given in Table C.4.

Table C.4 Thermophysical properties of plants parts (Cengel Jiayu Zhang 2025, Mousavizadeh 2010)

a/a	name	Volume percentage [%]	density, ρ [kg/m3]	thermal conductivity, k [W/mK]	Cp [J/kgK]
1	Fruit	10	980	0.63	4060
2	stem	10	980	0.5	4060
6	leaves	80	1335	0.5	3360

In the case of complex materials, the equivalent thermophysical properties have been calculated according to the following relationships, while for the PV panels the procedure is detailed in the (Baxevanou 2020).

$$\rho_{tot} = \frac{m_{tot}}{V_{tot}} = \frac{\rho_I V_I + \rho_C V_C}{V_I + V_C}$$

$$Q_{tot} = Q_I + Q_C \Rightarrow m_{tot} C_{Ptot} \Delta\Theta = m_C C_{PC} \Delta\Theta + m_I C_{PI} \Delta\Theta \Rightarrow C_{Ptot} = \frac{\rho_C V_C C_{PC} + \rho_I V_I C_{PI}}{\rho_C V_C + \rho_I V_I}$$

$$\frac{d_I}{\lambda_I} + \frac{d_C}{\lambda_C} = \frac{d_{tot}}{\lambda_{tot}} \Rightarrow \lambda_{tot} = \frac{d_{tot}}{\frac{d_I}{\lambda_I} + \frac{d_C}{\lambda_C}}$$

For the calculation of optical properties sources sometimes give direct extinction coefficient and refractive index but most of the time they give the transmissivity and the reflectivity along with the material thickness and then the optical properties are calculated according to the following relationships.

$$r = \left(\frac{n-1}{n+1}\right)^2 \Rightarrow n = \frac{r+1+2\sqrt{r}}{1-r}$$

$$\alpha_s = \frac{1}{d} \ln\left(\frac{1}{1-a}\right)$$

$$\tau_\alpha = 1 - \alpha,$$

$$\tau_t = \tau_\alpha x \left(\frac{1-r}{1+r}\right) x \left(\frac{1-r^2}{1-(r x \tau_\alpha)^2}\right)$$

$$\rho_t = r\left(1 + \tau_\alpha x \tau_t\right)$$

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

$$\alpha_t = (1 - \tau_\alpha) x \left(\frac{1 - r}{1 - r x \tau_\alpha} \right)$$

In Table C.5 materials optical properties are given for the four considered wavelength bands.

Table C.5 Materials optical properties [9-15]

Material	Extinction coefficient, a_s [1/m]	Refractive index, n [-]
Cover	2749/76.81/125.9/15943	1.5148/1.502/1.4919/1.519
Side walls	10000/21.75/2.412/1071	1.6339/1.6065/1.5739/1.6235
Ground – geotextile	1000/1000/1000/1000	1/1/1/1.09
Bench without substrate	61.13/16/16/15.99	3.96/0.71/1.849/1.999
Bench with substrate	20.61/20.61/20.6/20/62	1.634/1.602/1.574/1.624
PVs	816/134.4/131.2/254	1.613/3.461/3.425/2.372
Plants	1000/1000/1000/14.13	3/1.809/2.25/1

Again, for the complex materials the optical properties are calculated from their components according to the procedure described in (Duffie 2006, Baxevanou 2014). So, in Tables C.6 and C.7 the optical properties of PV components are given

$$\tau_t = \frac{\tau_{t1} x \tau_{t2}}{1 - \rho_{t1} x \rho_{t2}}$$

$$\rho_t = \rho_{t1} + \frac{\tau_t x \rho_{t2} x \tau_{t1}}{\tau_{t2}}$$

$$\alpha_t = 1 - \rho_t - \tau_t$$

Table C.6 PV components extinction coefficients

a/a	name	thickness [mm]	UV	PAR	NIR	IR
1	tempered glass	2	756,6	11,75	11,75	217,5
2	POE -EVA	0,5	134,7	3,597	3,597	272,1
3	dark-blue antireflective film	0,06	18202	2,628	0	0
4	PV monocrystaline silicon	0,06	130850666	1428936	51644	29,78
5	dark-blue antireflective film	0,06	18202	2,628	0	0
6	POE -EVA	0,5	134,7	3,597	3,597	272,1
7	tempered glass	2	756,6	11,75	11,75	

Table C.7 PV components refractive index

a/a	name	thickness [mm]	UV	PAR	NIR	IR
1	tempered glass	1,5539	1,5068	1,5068	1,48	1,5539
2	POE -EVA	1,86	1,79	1,79	1,72	1,86
3	dark-blue antireflective film	1,503	1,48	1,472	1,47	1,503
4	PV monocrystaline silicon	3,83	4,26	3,76	3,69	3,83
5	dark-blue antireflective film	1,503	1,48	1,472	1,47	1,503
6	POE -EVA	1,86	1,79	1,79	1,72	1,86
7	tempered glass	1,5539	1,5068	1,5068	1,48	1,5539

Air optical properties are calculated with UDF according to the humidity contain as already presented.

C1. Material equivalent properties

For the complex material bench with and without substrate, ground covered by geotextile the optical properties are directly calculated for their real dimensions. However, some of the used

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the European Union

materials (cover, side walls and PVs) are very thin to be able to be discretized. So in the simulation they are presented as materials with bigger thickness, practically with the thickness of 1 cm so to be possible to simulate the solid material with at least 10 computational cells. This creates the need to assign to them equivalent thermophysical and optical properties that would create the same result as the real material. These equivalent effective properties are calculated according to the following relationships for the thermophysical properties while for the optical properties the relationships have been already presented.

$$\frac{d_{real,tot}}{\lambda_{real,tot}} = \frac{d_{sim}}{\lambda_{eff}} \Rightarrow \lambda_{eff} = d_{sim} x \frac{\lambda_{real,tot}}{d_{real,tot}}$$

$$m_{eff} = m_{real} \Rightarrow \rho_{eff} x V_{sim} = \rho_{real} x V_{real} \Rightarrow \rho_{eff} = \frac{\rho_{real} x V_{real}}{V_{sim}} = \frac{\rho_{real} x d_{real} x A}{d_{sim} x A} \Rightarrow \rho_{eff} = \frac{\rho_{real} x d_{real}}{d_{sim}}$$

$$Q_I = Q_{II} \Rightarrow m_{eff} x C_{p.eff} x \Delta\Theta = m_{real,tot} x C_{p.real,tot} x \Delta\Theta \Rightarrow C_{p.eff} = \frac{\rho_{real,tot} x d_{real,tot} x C_{p,real,tot}}{\rho_{eff} x d_{eff}}$$

Table C. 8 PV solar cell composition

Material	**Density, ρ [kg/m³]**	**Specific heat capacity, Cp [J/kgK]**	**Thermal conductivity, λ [W/mK]**
Cover	16.74	2300	21.67
Side walls	96	1200	2.5
PVs	800	895	2.16

Table C. 9 Simulation effective material optical properties

Material	Extinction coefficient, a_s [1/m]	Refractive index, n [-]
Cover	50.88/4.88/5.583/286.9	1.522/1.507/1.498/1.525
Side walls	799/6.294/1.583/1.627	1.634/1.615/1.583/1.627
PVs	423/69.6/67.96/131.6	1.613/3.461/3.423/2.372

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Annex D – Supplementary figures and extended results (Chapter 3)

Table D.1 PV solar cell composition

a/a	name	thickness [mm]
1	tempered glass	2
2	POE -EVA	0.5
3	dark-blue antireflective film	0.06
4	PV monocrystaline silicon	0.06
5	dark-blue antireflective film	0.06
6	POE -EVA	0.5
7	tempered glass	2

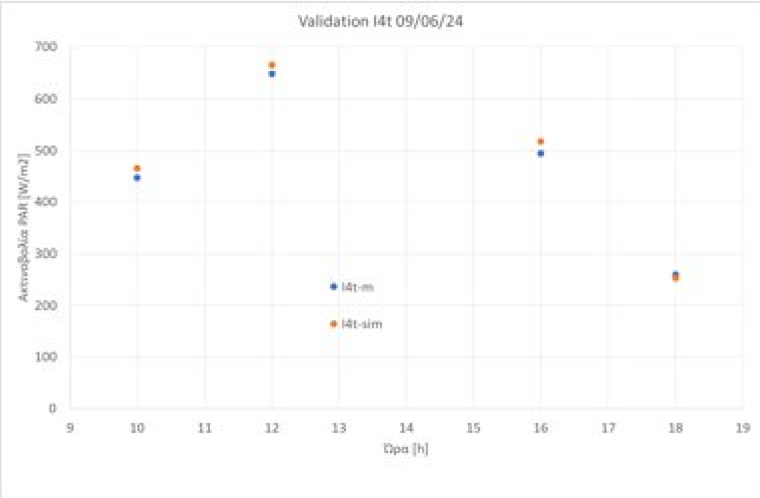


Fig. D.1 Comparison of measured and simulation predicted PAR in the location of sensors I4t, 09/06/24

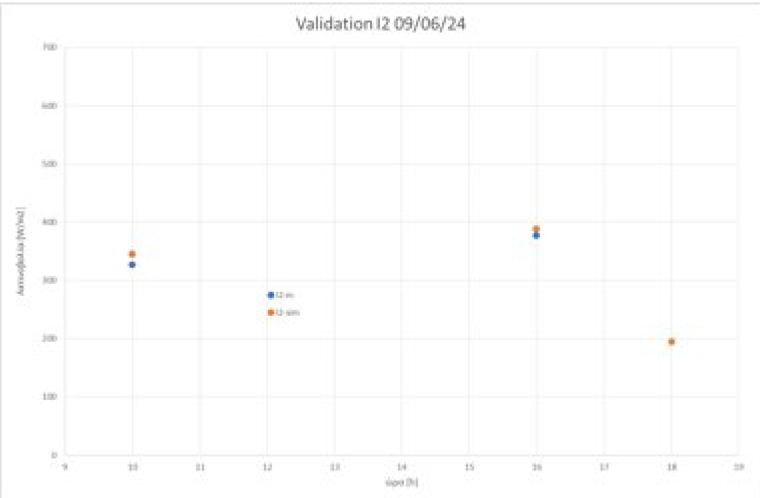


Fig. D.2 Comparison of measured and simulation predicted PAR in the location of sensors I2, 09/06/24

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

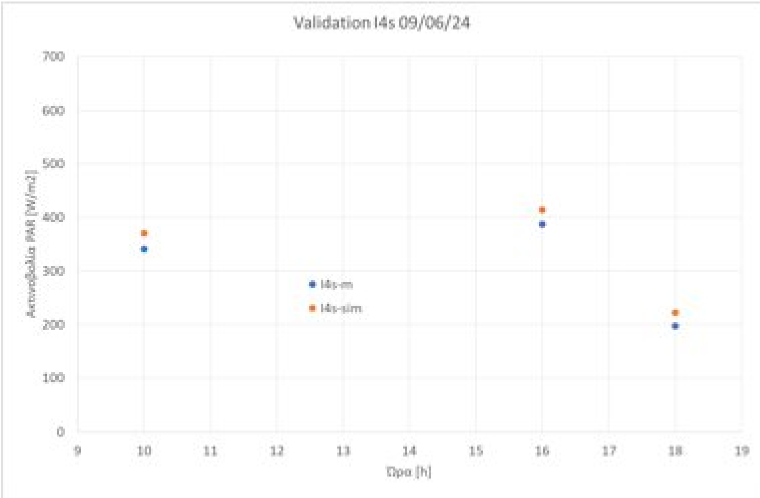


Fig. D.3 Comparison of measured and simulation predicted PAR in the location of sensors I4s, 09/06/24

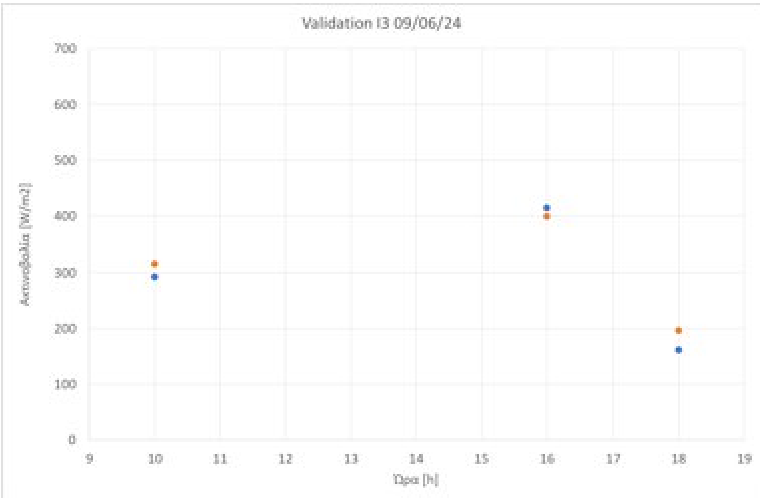


Fig. D.4 Comparison of measured and simulation predicted PAR in the location of sensors I3, 09/06/24

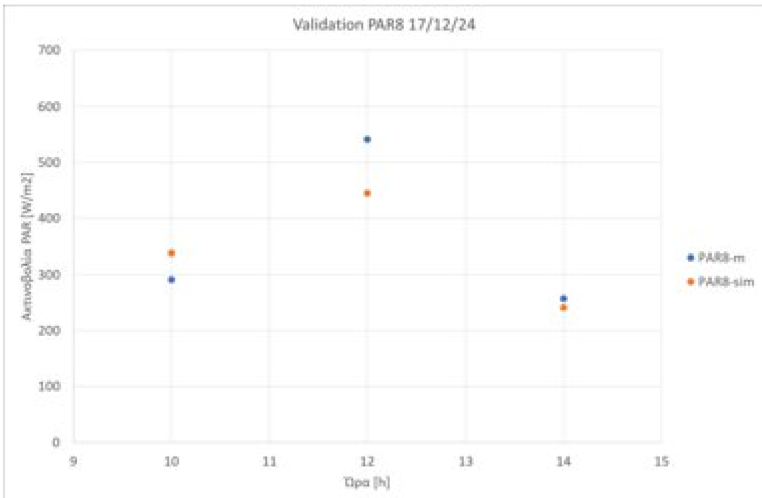


Fig. D.5 Comparison of measured and simulation predicted PAR in the location of sensors PAR8, 17/12/24

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

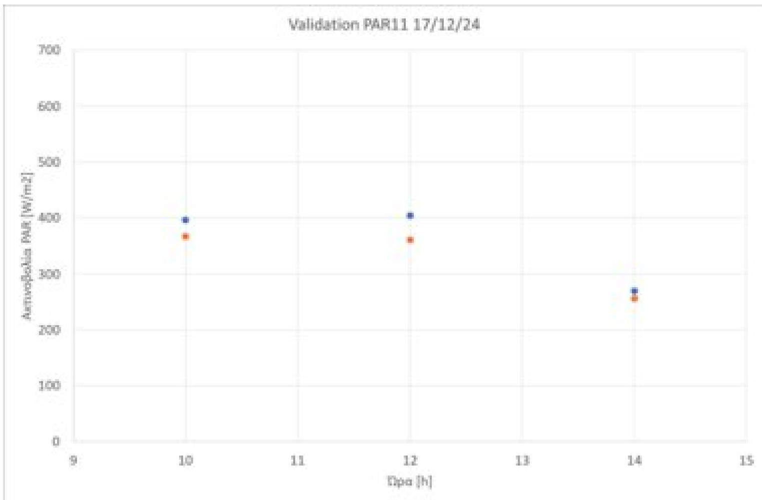


Fig. D.6 Comparison of measured and simulation predicted PAR in the location of sensors PAR11, 17/12/24

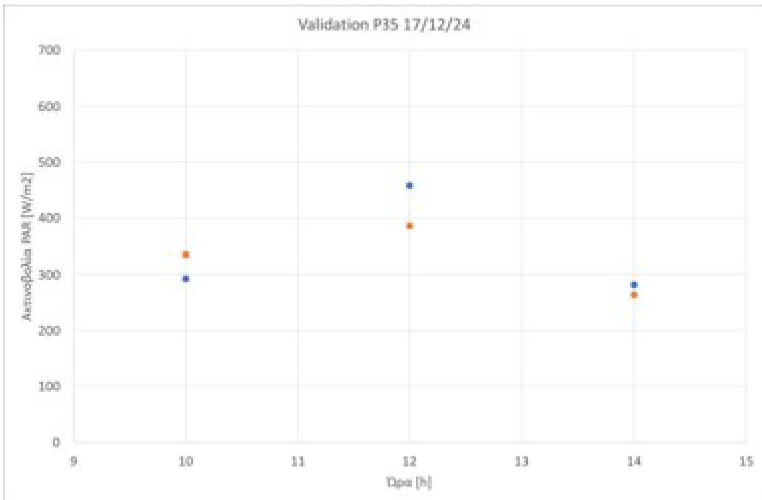


Fig. D.7 Comparison of measured and simulation predicted PAR in the location of sensors P35, 17/12/24

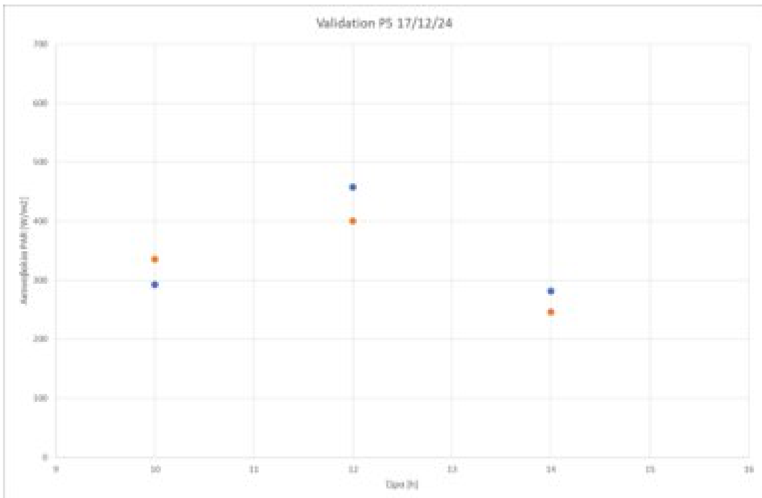


Fig. D.8 Comparison of measured and simulation predicted PAR in the location of sensors P5, 17/12/24

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

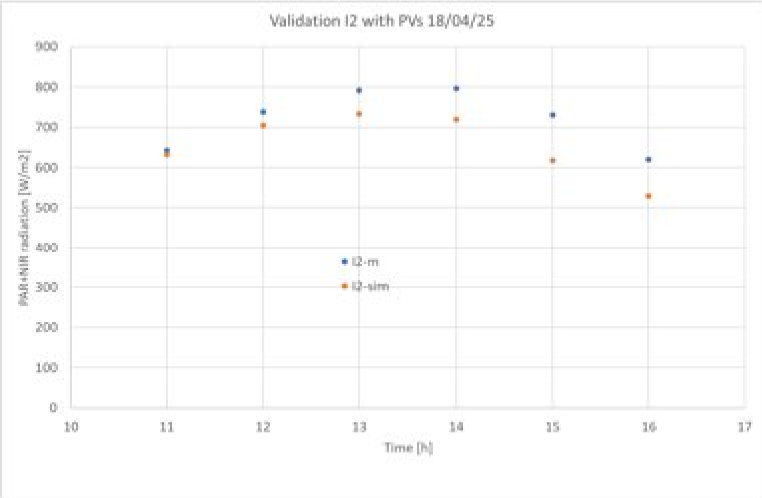


Fig. D.9 Comparison of measured and simulation predicted PAR+NIR in the location of sensors I2, 18/4/25

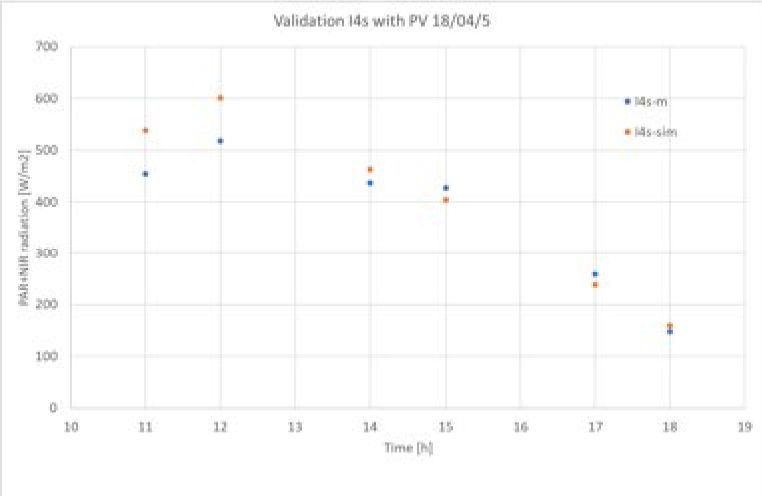


Fig. D.10 Comparison of measured and simulation predicted PAR+NIR in the location of sensors I4s, 18/4/25

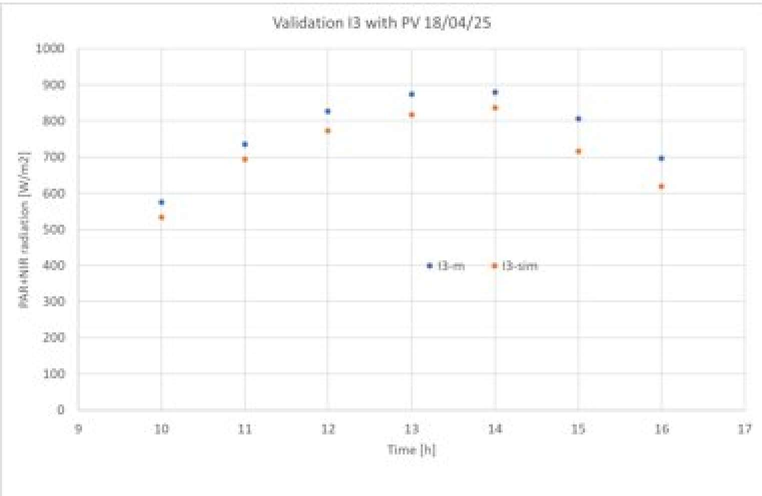


Fig. D.11 Comparison of measured and simulation predicted PAR+NIR in the location of sensors I3, 18/4/25

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

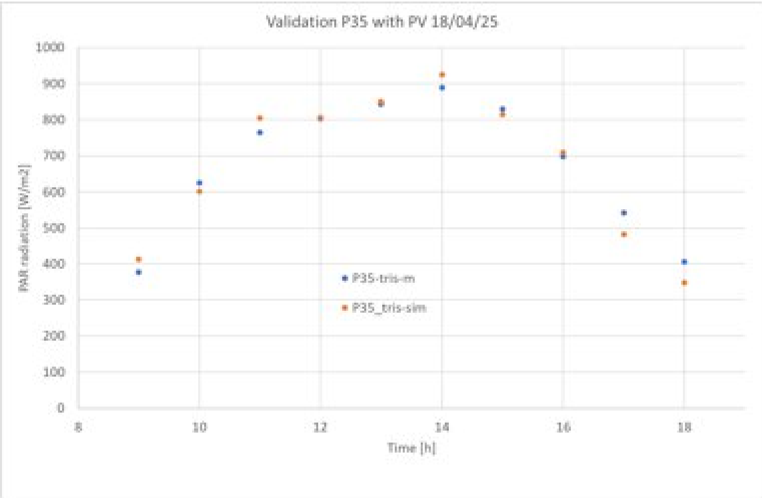


Fig. D.12 Comparison of measured and simulation predicted PAR in the location of sensors P35, 18/4/25

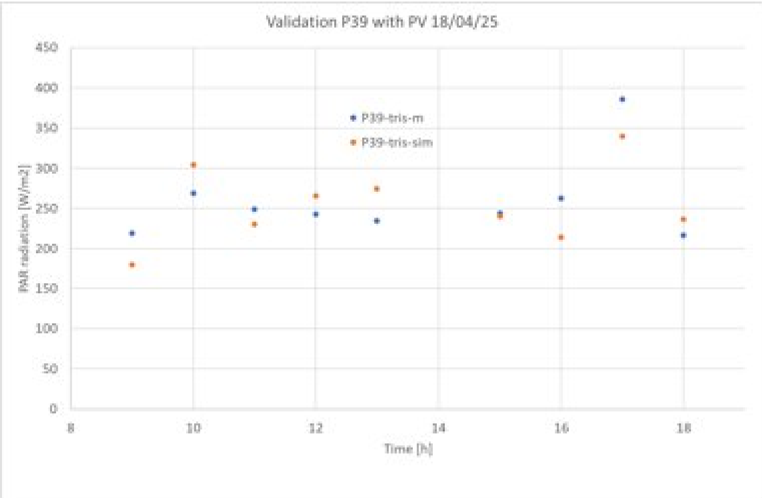


Fig. D.13 Comparison of measured and simulation predicted PAR in the location of sensors P39, 18/4/25

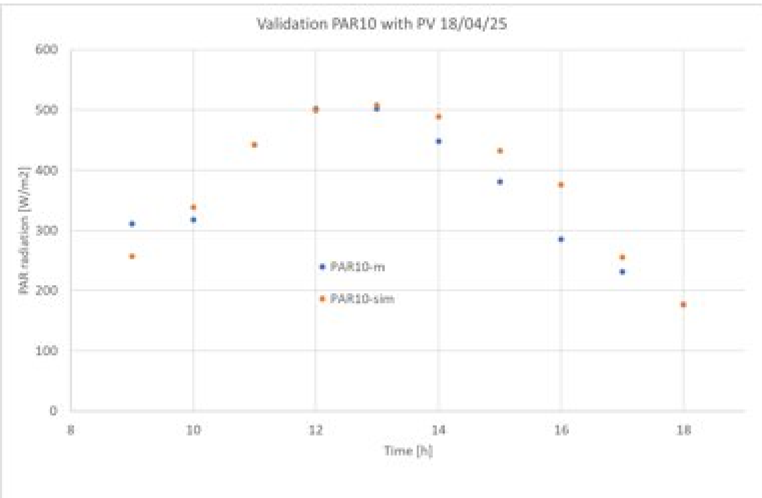


Fig. D.14 Comparison of measured and simulation predicted PAR in the location of sensors PAR10, 18/4/25

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

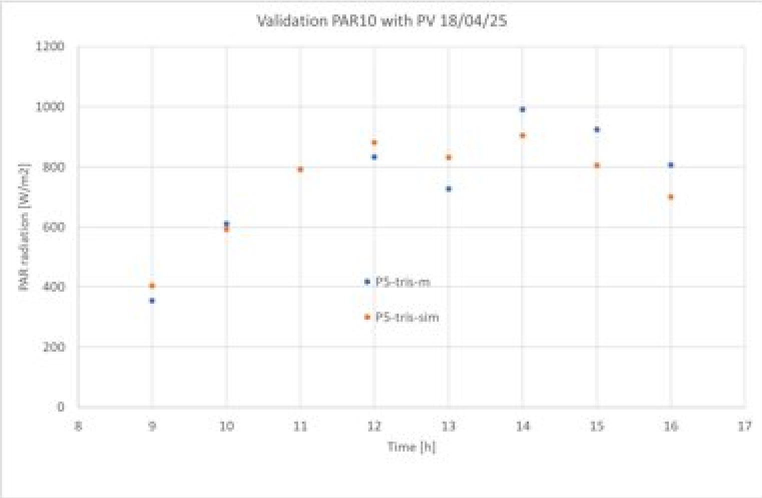


Fig. D.15 Comparison of measured and simulation predicted PAR in the location of sensors P5, 18/4/25

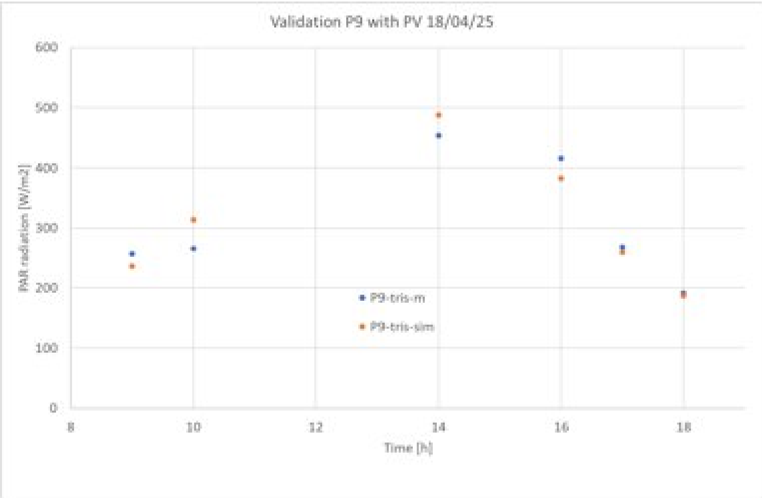


Fig. D.16 Comparison of measured and simulation predicted PAR in the location of sensors P9, 18/4/25

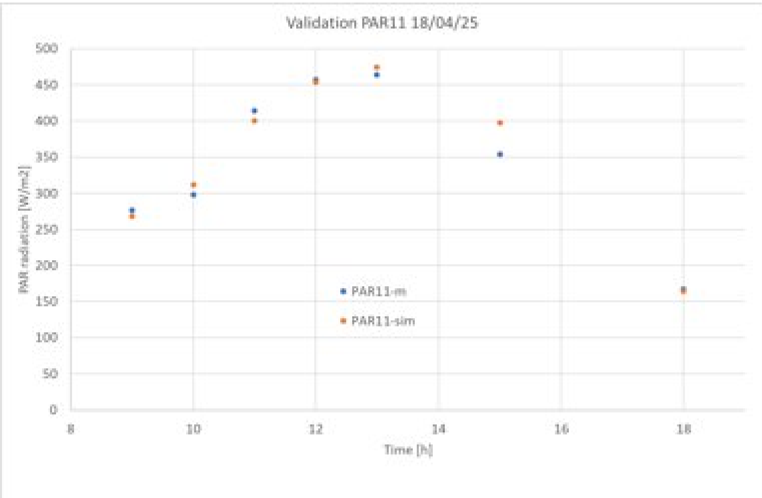


Fig. D.17 Comparison of measured and simulation predicted PAR in the location of sensors PAR11, 18/4/25

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

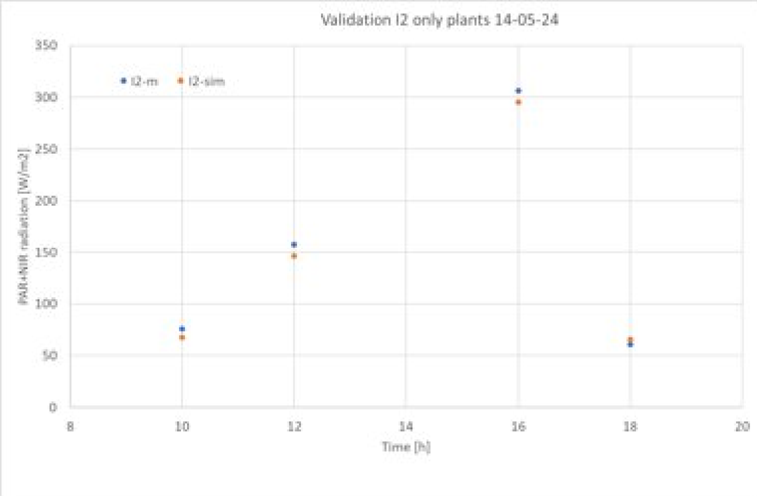


Fig. D18 Comparison of measured and simulation predicted PAR+NIR in the location of sensors I2, 14/5/24

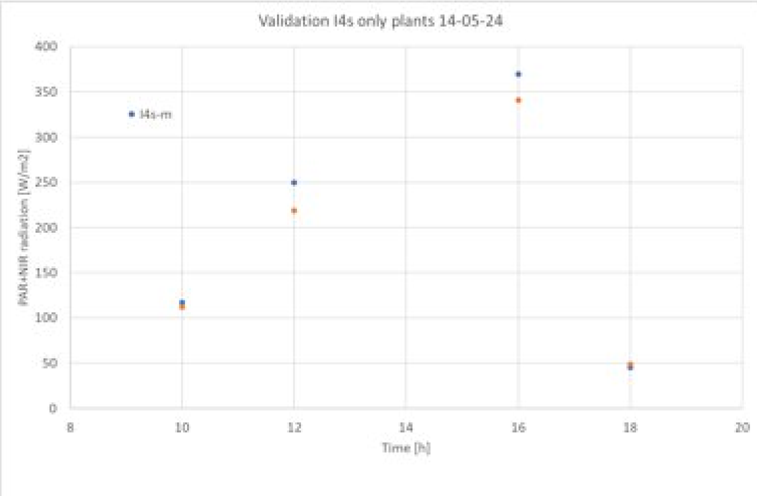


Fig. D.19 Comparison of measured and simulation predicted PAR+NIR in the location of sensors I4s, 14/5/24

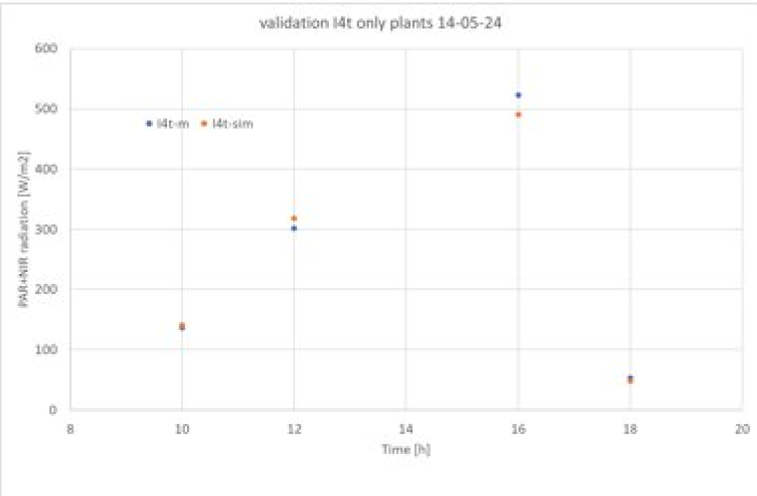


Fig. D.20 Comparison of measured and simulation predicted PAR+NIR in the location of sensors I4t, 14/5/24

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

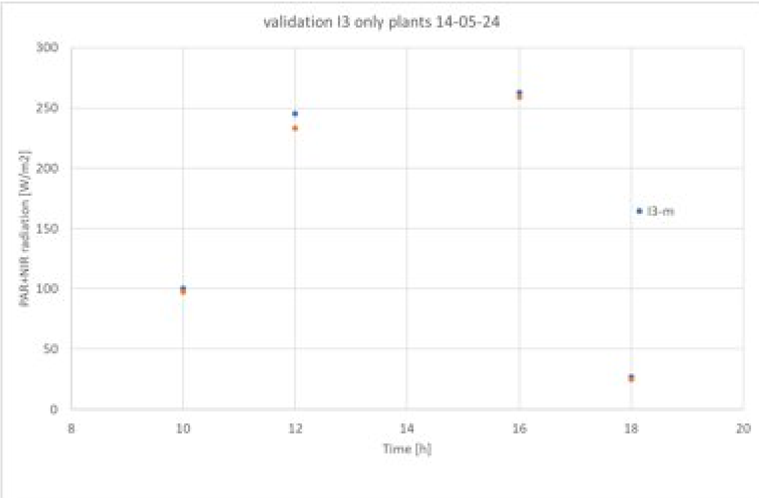


Fig. D.21 Comparison of measured and simulation predicted PAR+NIR in the location of sensors I4t, 14/5/24

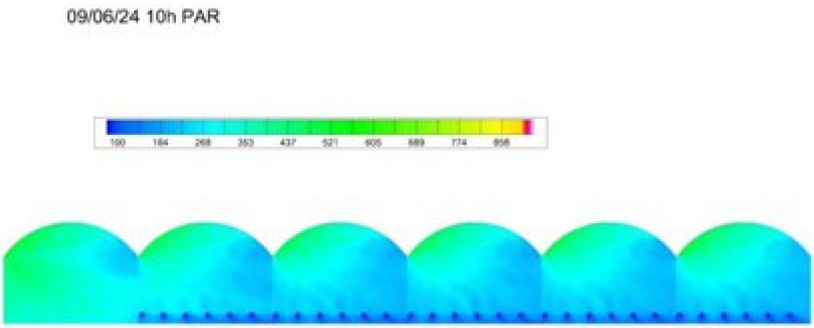


Fig. D.22 PAR isocontours for 09/06/24 without PV, without crops and without shading consideration at 10.00 [h]

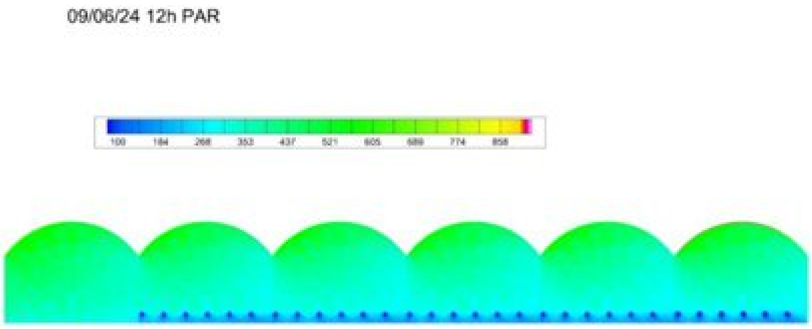


Fig. D.23 PAR isocontours for 09/06/24 without PV, without crops and without shading consideration at 12.00 [h]

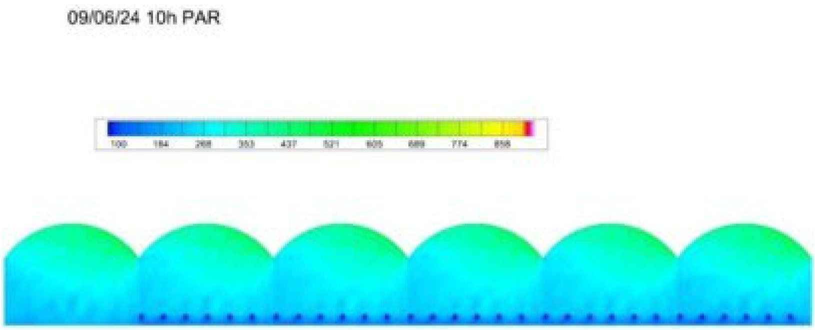


Fig. D.24 PAR isocontours for 09/06/24 without PV, without crops and without shading consideration at 16.00 [h]

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

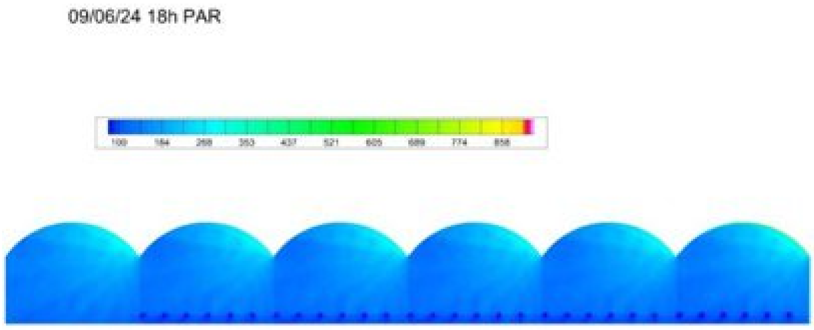


Fig. D.25 PAR isocontours for 09/06/24 without PV, without crops and without shading consideration at 18.00 [h]

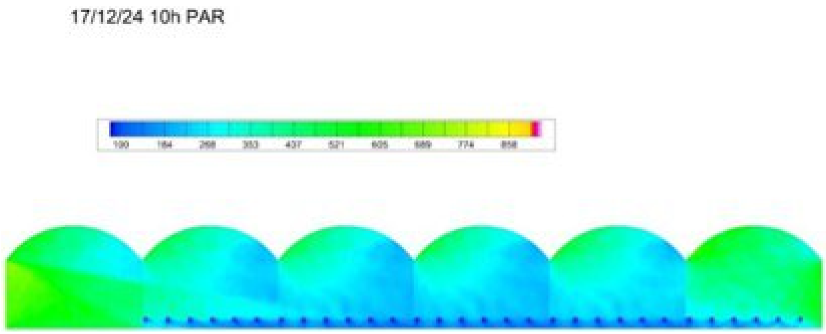


Fig. D.26 PAR isocontours for 17/12/24 without PV, without crops and without shading consideration at 10.00 [h]

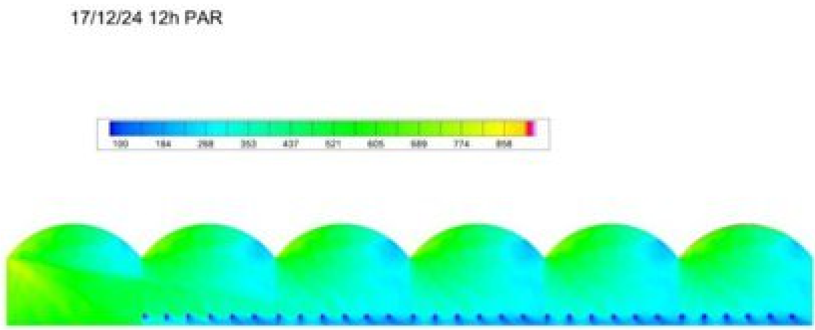


Fig. D. 27 PAR isocontours for 17/12/24 without PV, without crops and without shading consideration at 12.00 [h]

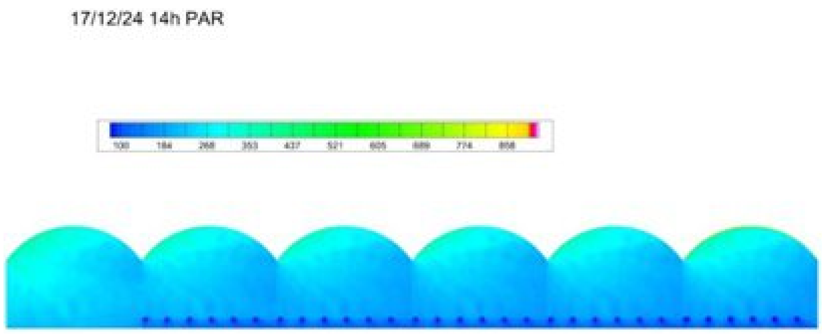


Fig. D.28 PAR isocontours for 17/12/24 without PV, without crops and without shading consideration at 14.00 [h]

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the
European Union

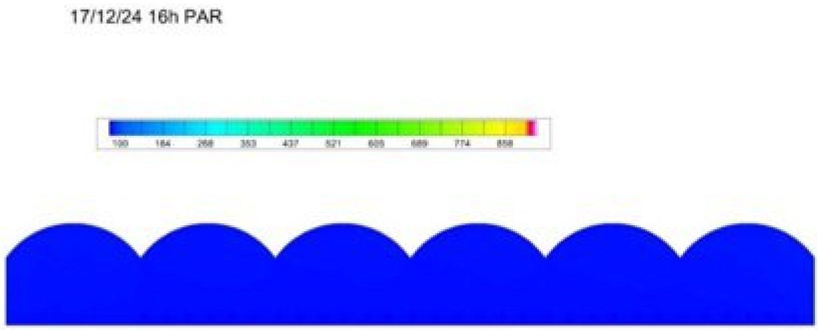


Fig. D.29 PAR isocontours for 17/12/24 without PV, without crops and without shading consideration at 16.00 [h]

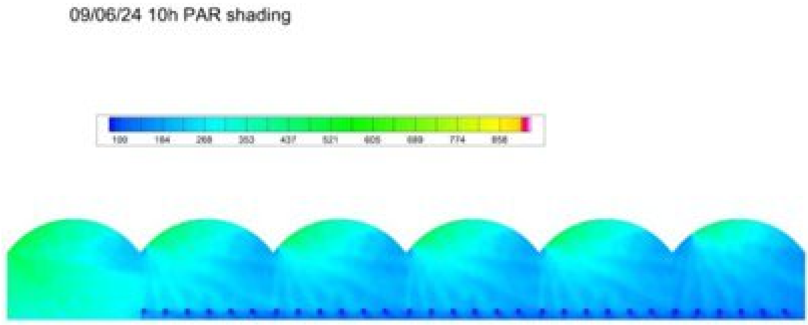


Fig. D.30 PAR isocontours for 09/06/24 without PV, without crops accounting shading at 10.00 [h]

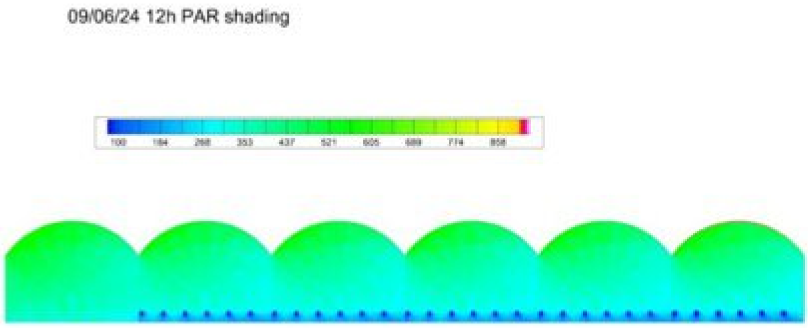


Fig. D.31 PAR isocontours for 09/06/24 without PV, without crops accounting shading at 12.00 [h]

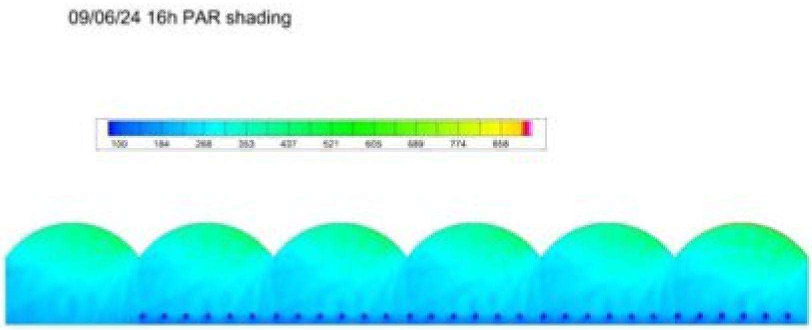


Fig. D.32 PAR isocontours for 09/06/24 without PV, without crops accounting shading at 16.00 [h]

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

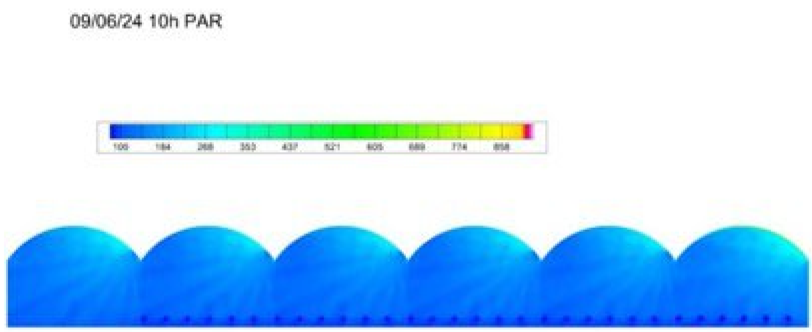


Fig. D.33 PAR isocontours for 09/06/24 without PV, without crops accounting shading at 18.00 [h]

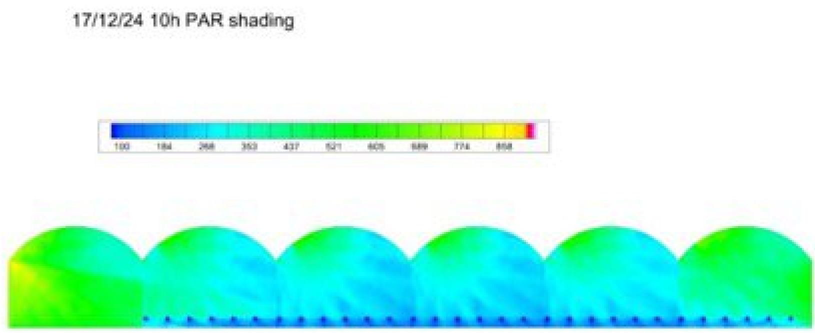


Fig. D.34 PAR isocontours for 17/12/24 without PV, without crops accounting shading at 10.00 [h]

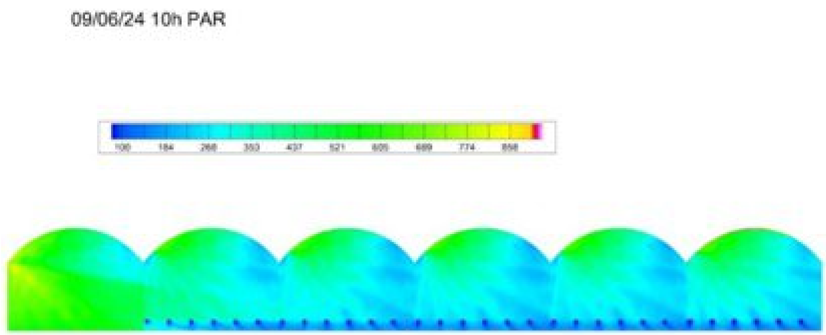


Fig. D. 35 PAR isocontours for 17/12/24 without PV, without crops accounting shading at 12.00 [h]

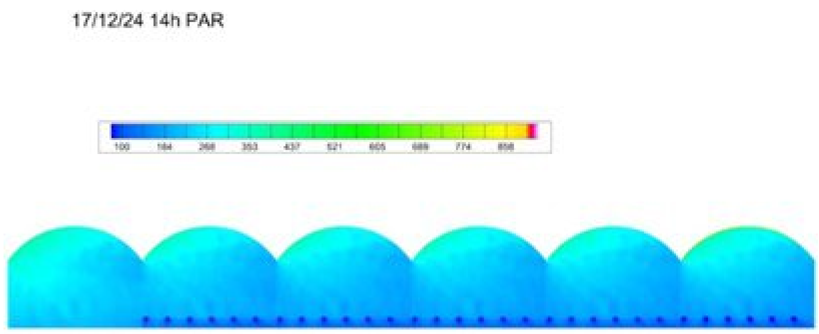


Fig. D. 36 PAR isocontours for 17/12/24 without PV, without crops accounting shading at 14.00 [h]

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

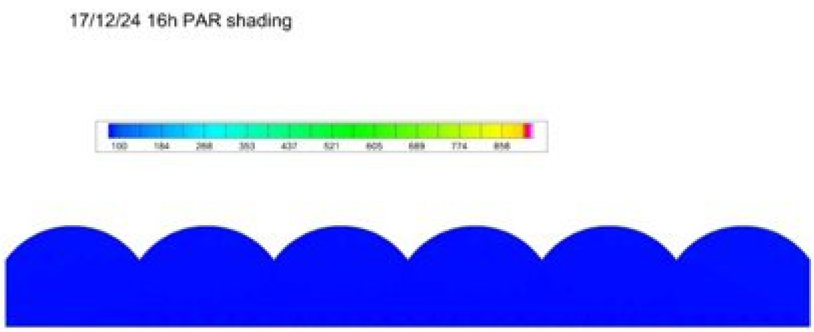


Fig. D.37 PAR isocontours for 17/12/24 without PV, without crops accounting shading at 16.00 [h]

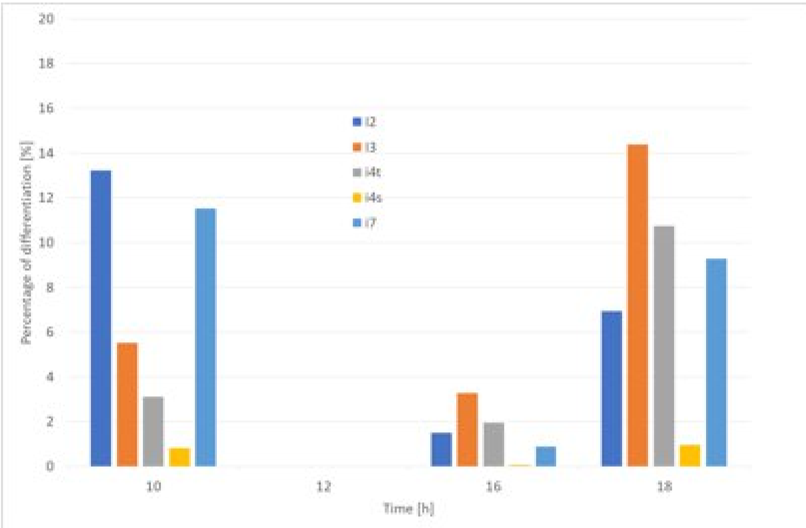


Fig. D.38 Differentiation at predicted PAR radiation due to shading consideration for the 09/06/24

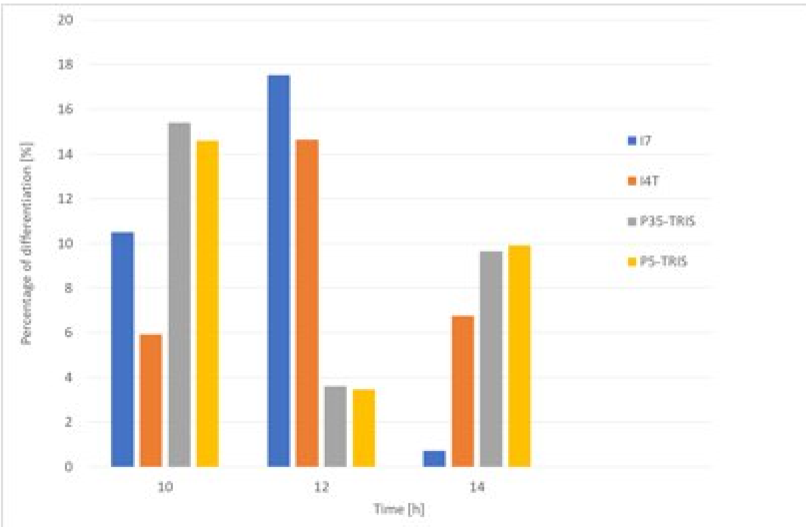


Fig. D. 39 Differentiation at predicted PAR radiation due to shading consideration for the 17/12/24

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.
Funded by the European Union

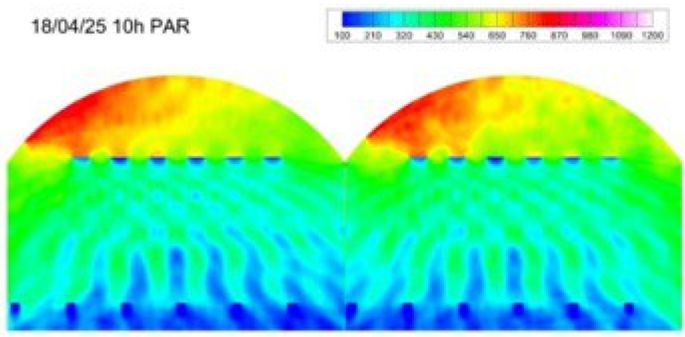


Fig. D.40 PAR isocontours for 18/04/25 with PV at horizontal position, without crops accounting shading at 10.00 [h]

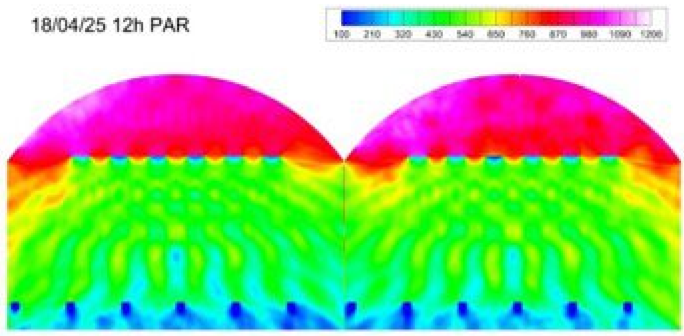


Fig. D.41 PAR isocontours for 18/04/25 with PV at horizontal position, without crops accounting shading at 12.00 [h]

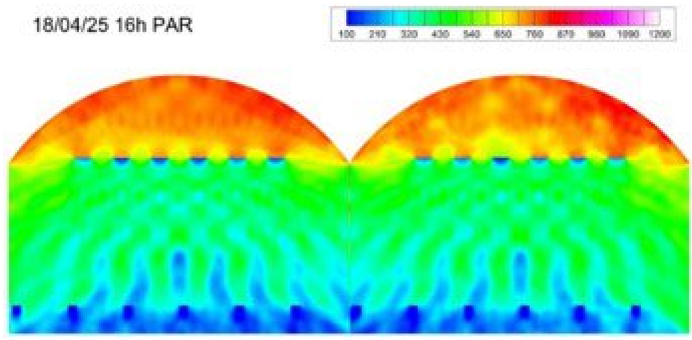


Fig. D.42 PAR isocontours for 18/04/25 with PV at horizontal position, without crops accounting shading at 16.00 [h]

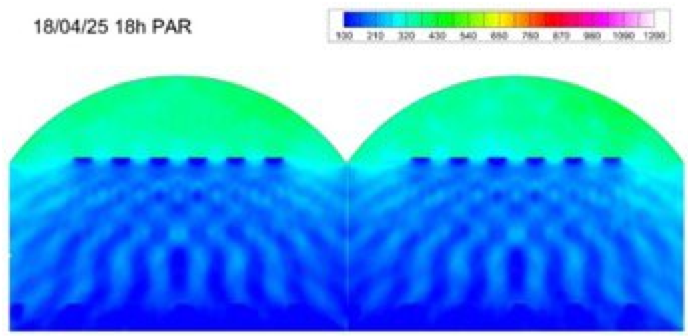


Fig. D. 43 PAR isocontours for 18/04/25 with PV at horizontal position, without crops accounting shading at 18.00 [h]

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

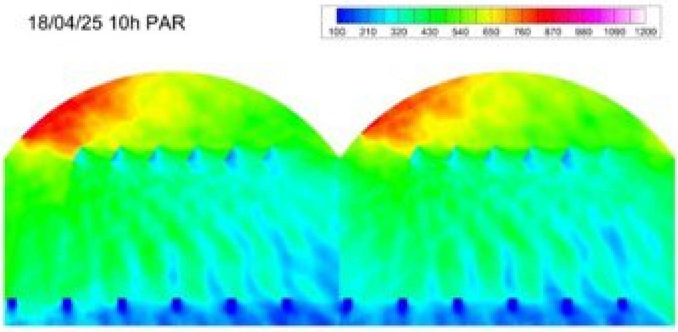
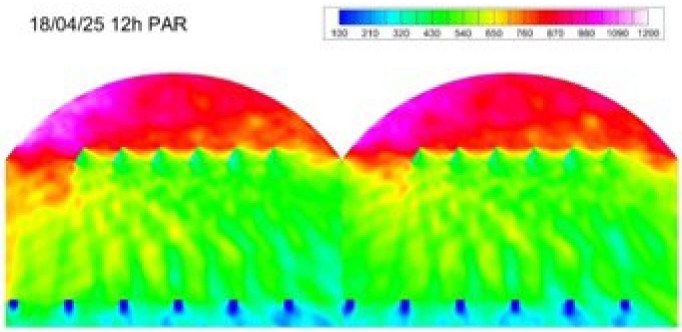


Fig. D.44 PAR isocontours for 18/04/25 with PV inclined, without crops accounting shading at 10.00 [h]



4

Fig. D.45 PAR isocontours for 18/04/25 with PV inclined, without crops accounting shading at 12.00 [h]

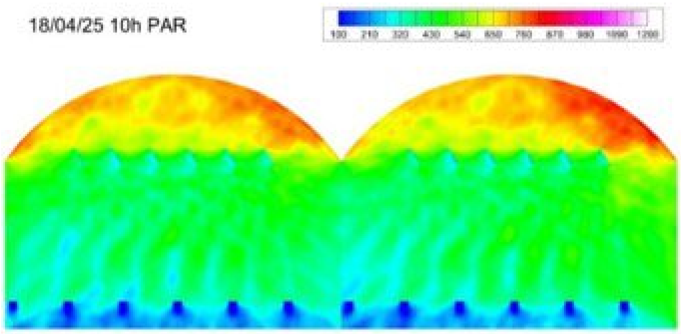


Fig. D.46 PAR isocontours for 18/04/25 with PV inclined, without crops accounting shading at 16.00 [h]

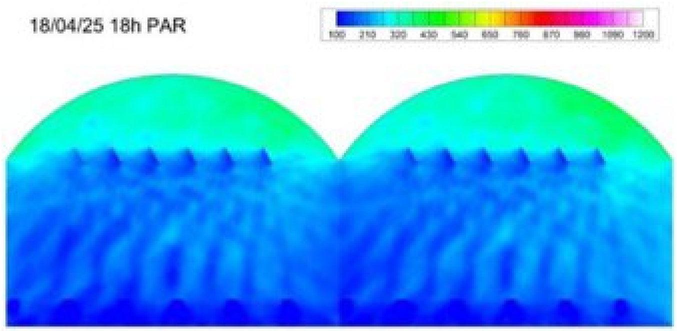


Fig. D.47 PAR isocontours for 18/04/25 with PV inclined, without crops accounting shading at 18.00 [h]

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

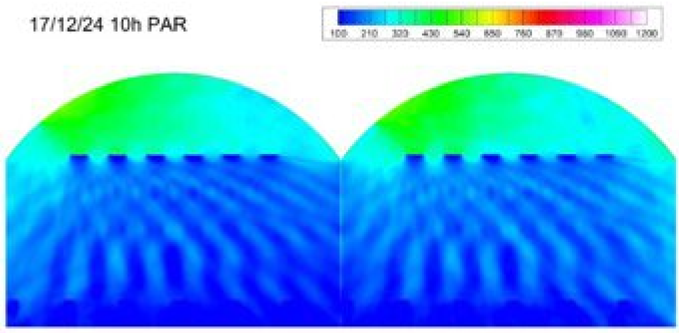


Fig. D.48 PAR isocontours for 17/12/24 with PV at horizontal position, without crops accounting shading at 10.00 [h]

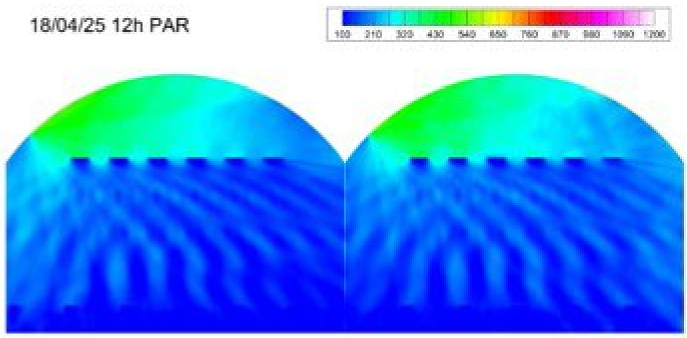


Fig. D.49 PAR isocontours for 17/12/24 with PV at horizontal position, without crops accounting shading at 12.00 [h]

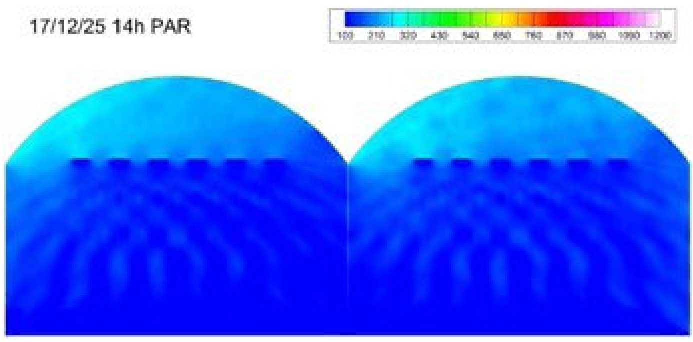


Fig. D.50 PAR isocontours for 17/12/24 with PV at horizontal position, without crops accounting shading at 16.00 [h]

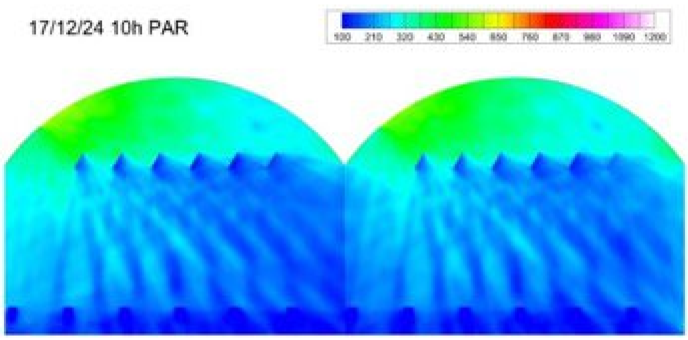


Fig. D.51 PAR isocontours for 17/12/24 with inclined PV, without crops accounting shading at 10.00 [h]

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

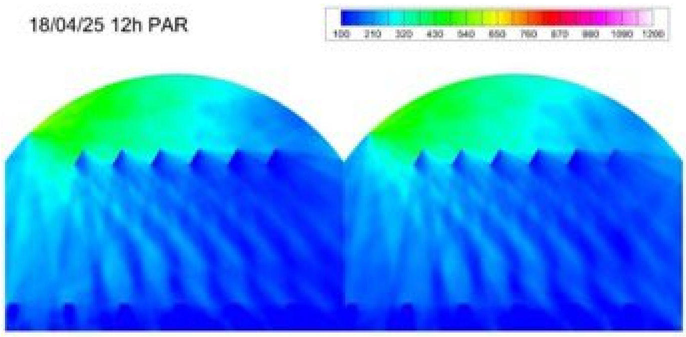


Fig. D.52 PAR isocontours for 17/12/24 with inclined PV, without crops accounting shading at 12.00 [h]

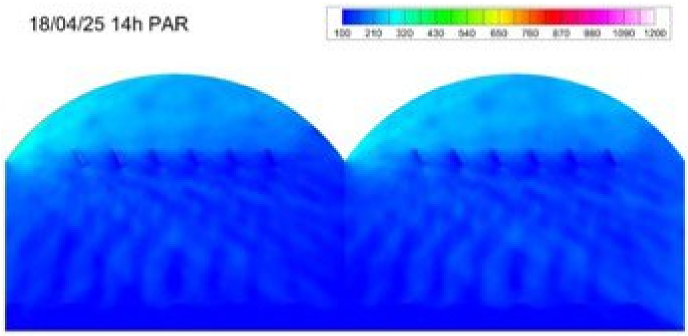


Fig. D.53 PAR isocontours for 17/12/24 with inclined PV, without crops accounting shading at 14.00 [h]

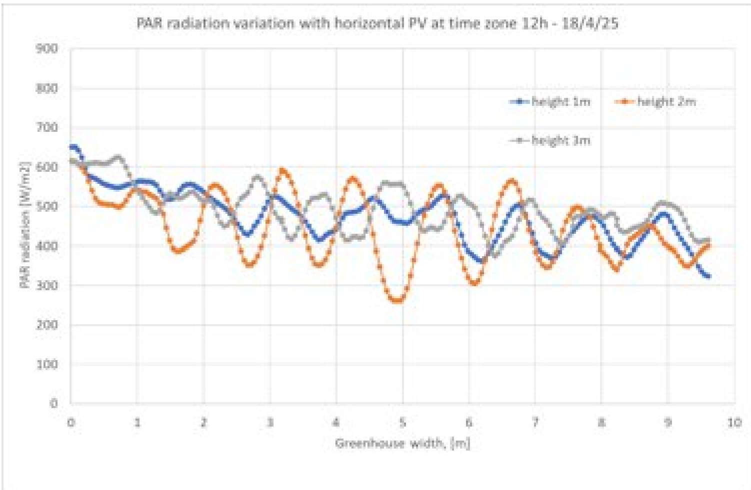


Fig. D.54 PAR radiation variation with horizontal PV at time zone 12h - 18/4/25

Funded by the European Union
by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

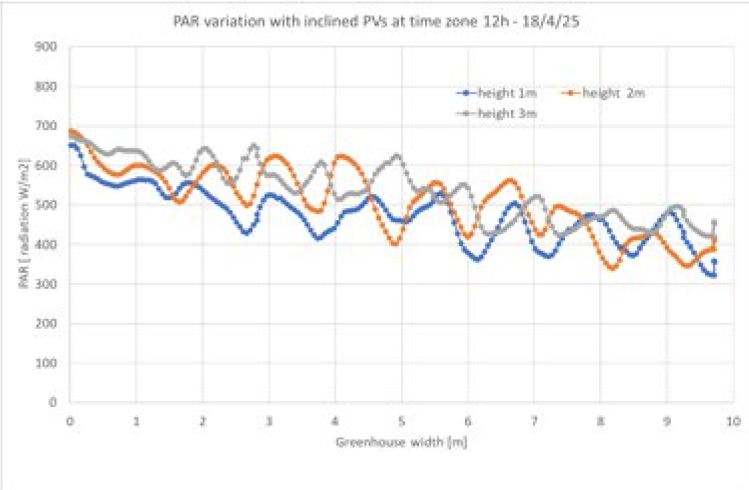


Fig. D.55 PAR radiation variation with inclined PV at time zone 12h - 18/4/25

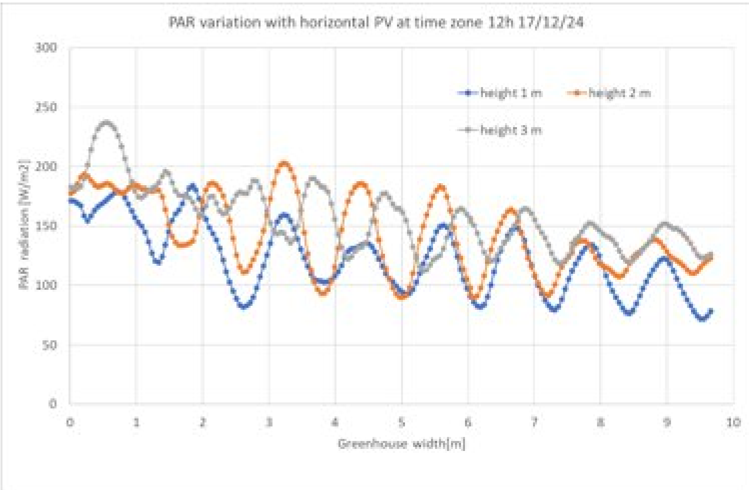


Fig. D.56 PAR radiation variation with horizontal PV at time zone 12h - 17/12/25

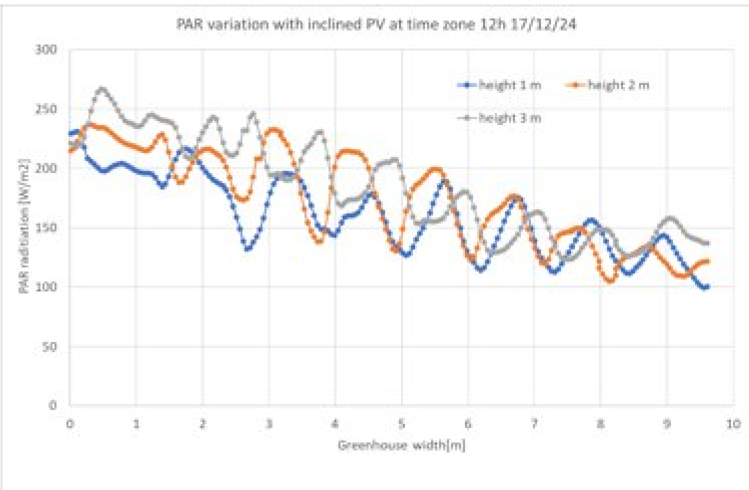


Fig. D.57 PAR radiation variation with horizontal PV at time zone 12h - 17/12/25

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

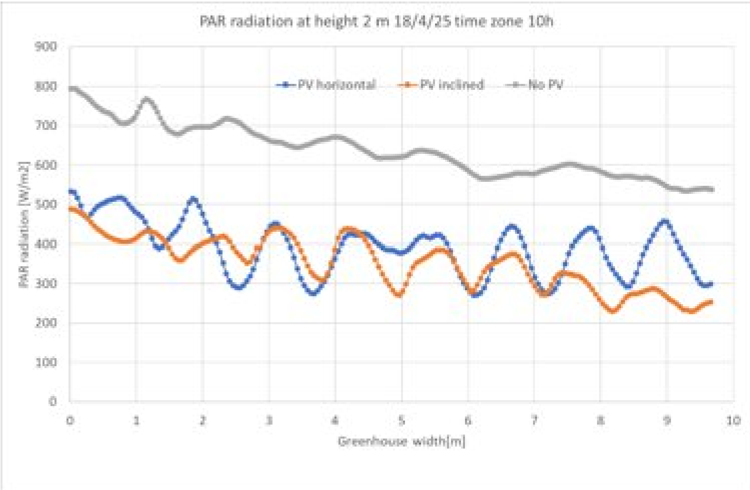


Fig. D.58 PAR radiation variation at height 2 m across chamber at 10h of 18/4/25

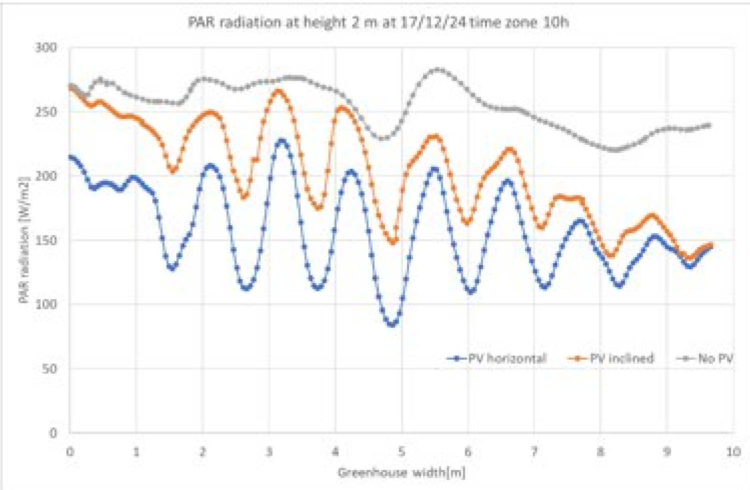


Fig. D.59 PAR radiation variation at height 2 m across chamber at 10h of 17/12/25

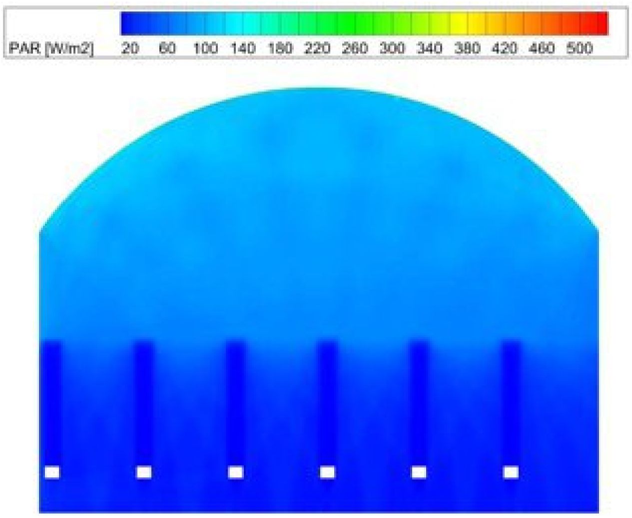


Fig. D.60 PAR isocontours for 14/05/24 with cucumber plants, without PV panels at 10.00 [h]

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

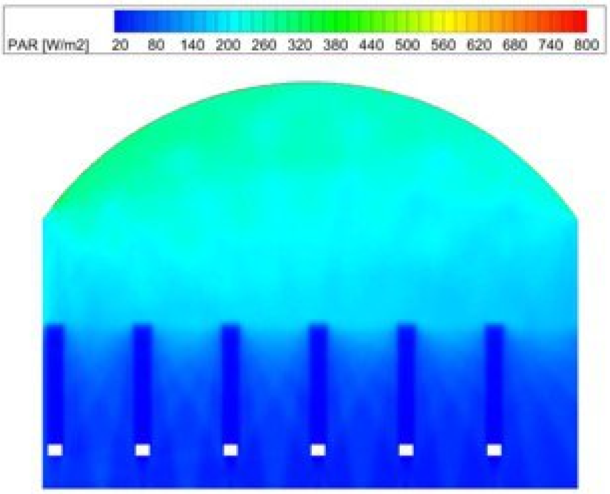


Fig. D.61 PAR isocontours for 14/05/24 with cucumber plants, without PV panels at 12.00 [h]

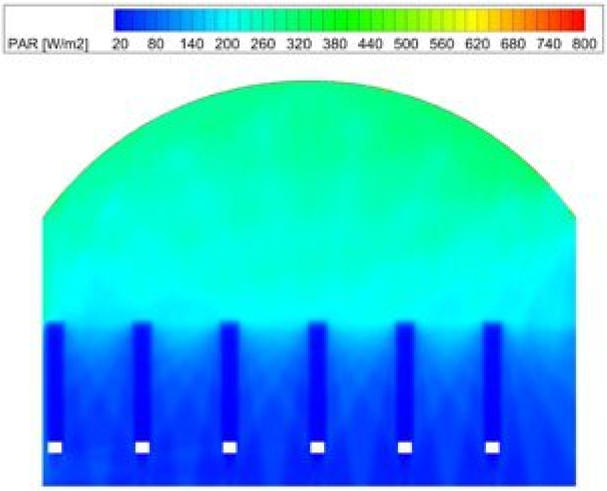


Fig. D.62 PAR isocontours for 14/05/24 with cucumber plants, without PV panels at 16.00 [h]

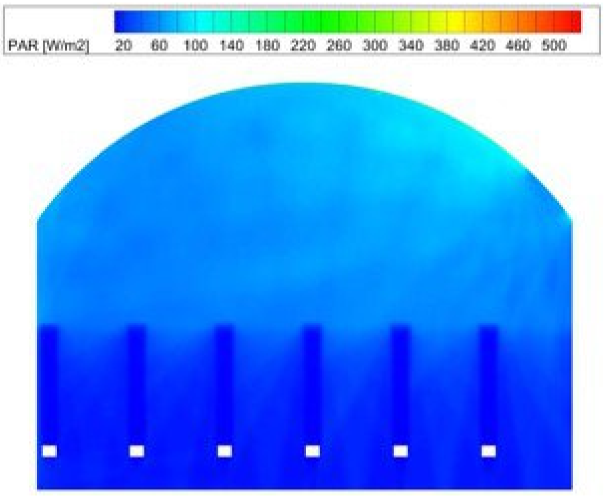


Fig. D.63 PAR isocontours for 14/05/24 with cucumber plants, without PV panels at 18.00 [h]

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

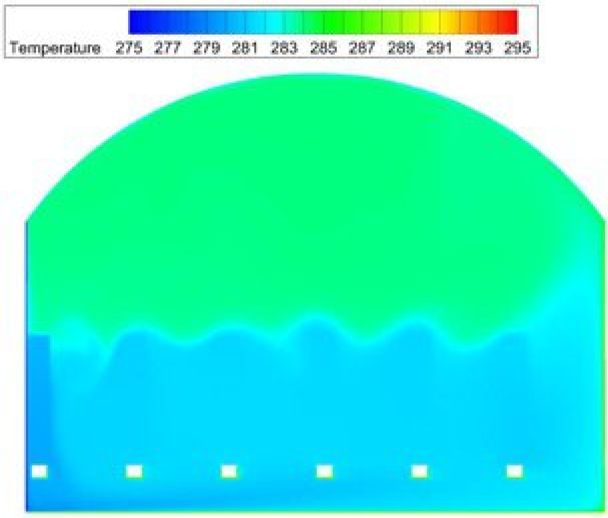


Fig. D.64 Temperature isocontours for 14/05/24 with cucumber plants, without PV panels at 10.00 [h]

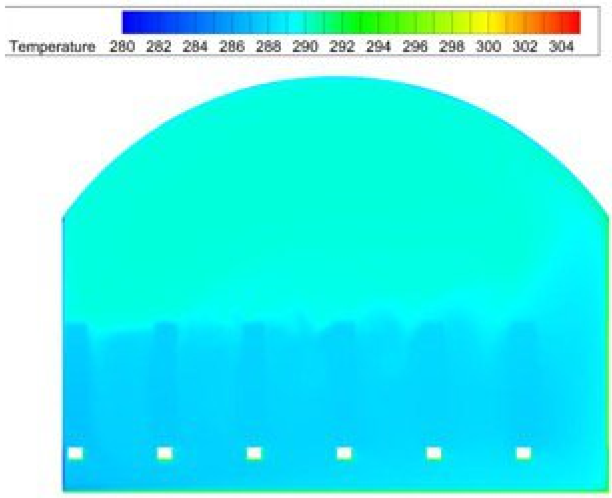


Fig. D.65 Temperature isocontours for 14/05/24 with cucumber plants, without PV panels at 12.00 [h]

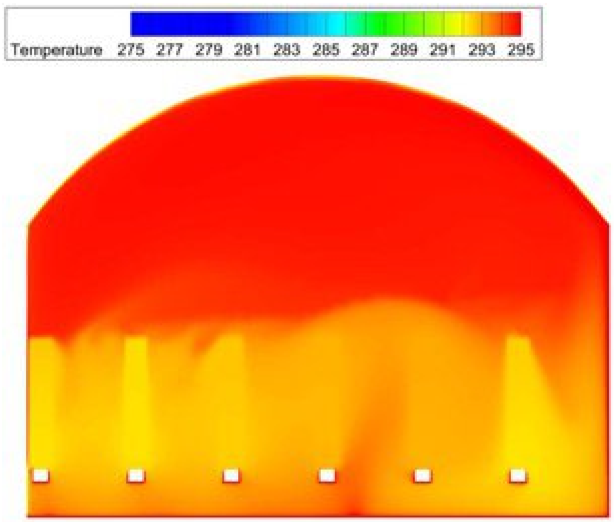


Fig. D.66 Temperature isocontours for 14/05/24 with cucumber plants, without PV panels at 16.00 [h]

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

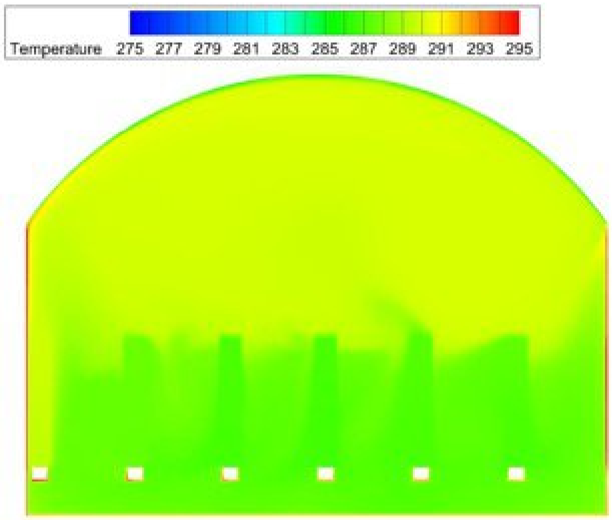


Fig. D.67 Temperature isocontours for 14/05/24 with cucumber plants, without PV panels at 18.00 [h]

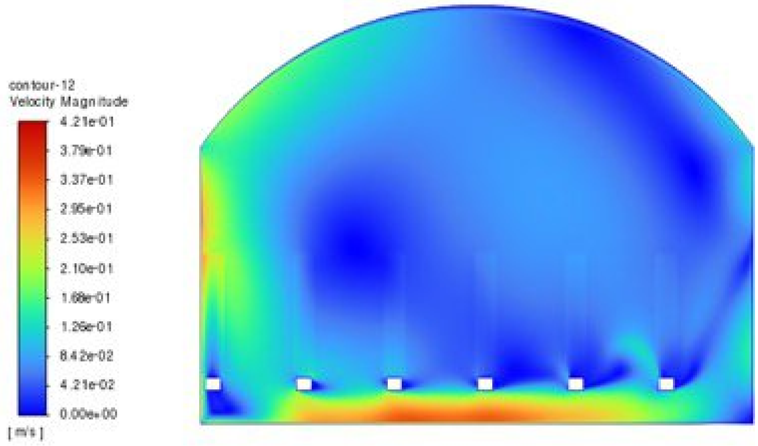


Fig. D.68 Velocity magnitude isocontours for 14/05/24 with cucumber plants, without PV panels at 12.00 [h]

Fig. D.69 Streamlines for 14/05/24 with cucumber plants, without PV panels at 12.00 [h]

Funded by the European Union

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

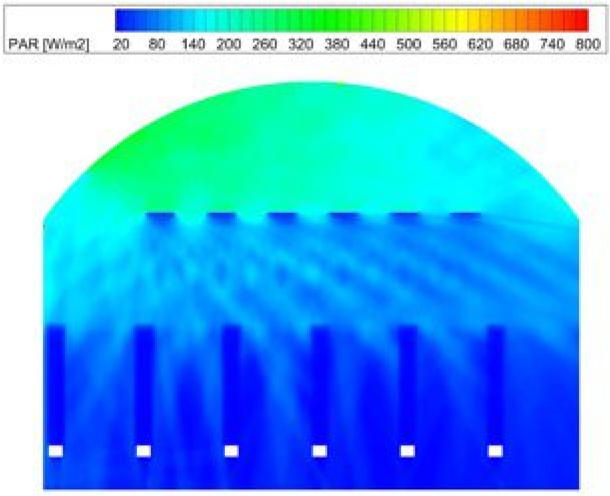


Fig. D.70 PAR isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 10.00 [h]

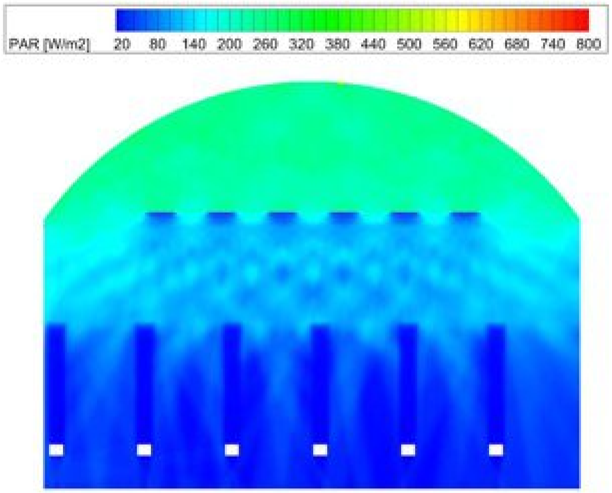


Fig. D.71 PAR isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 12.00 [h]

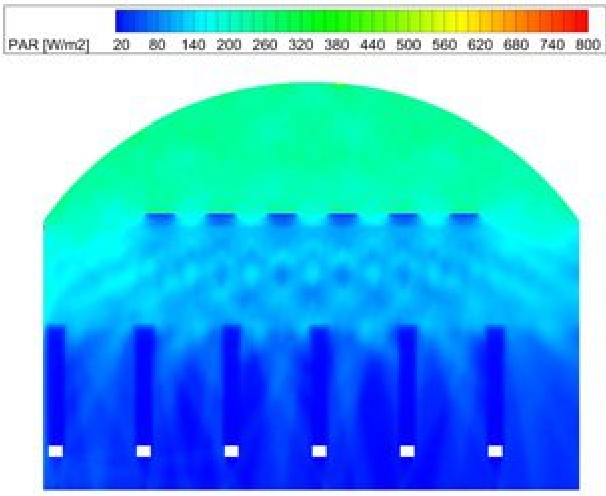


Fig. D.72 PAR isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 14.00 [h]

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

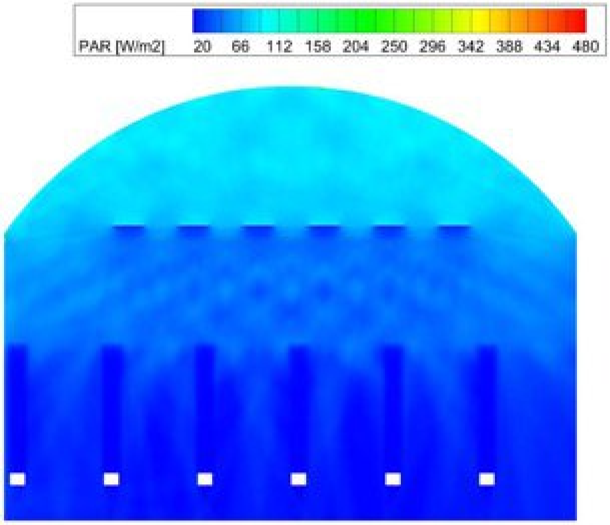


Fig. D.73 PAR isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 16.00 [h]

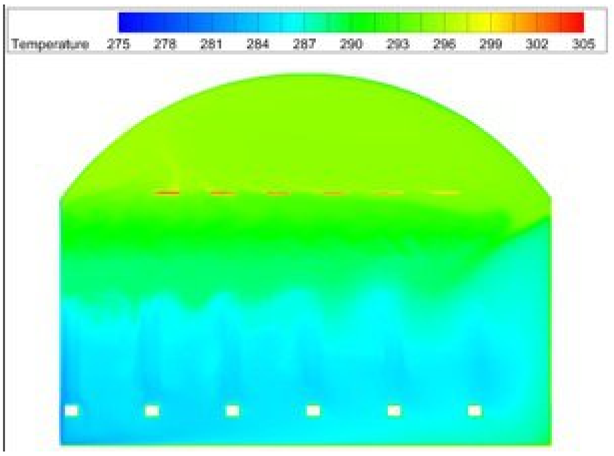


Fig. D.74 temperature isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 10.00 [h]

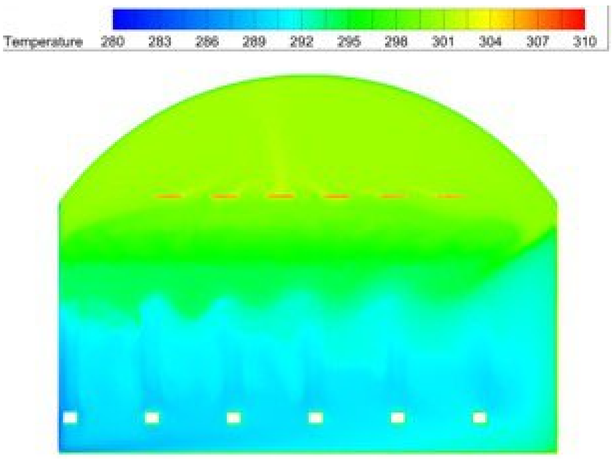


Fig. D.75 temperature isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 12.00 [h]

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

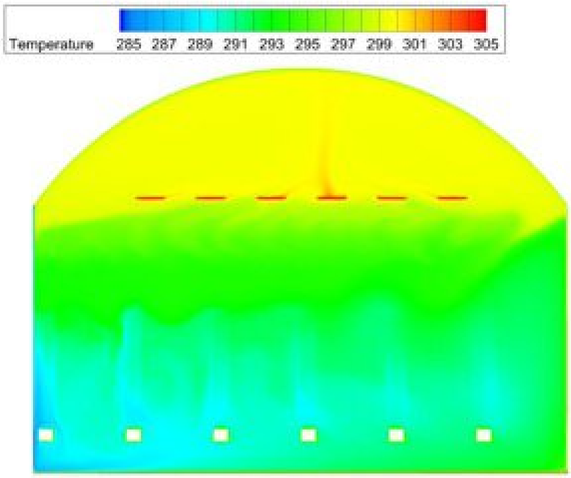


Fig. D.76 temperature isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 14.00 [h]

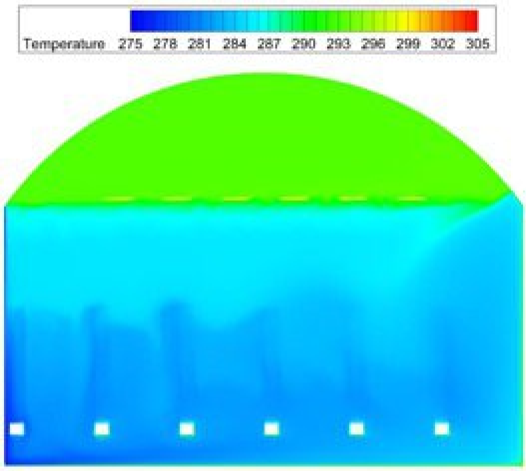


Fig. D.77 temperature isocontours for 11/10/24 with cucumber plants, and PV panels in horizontal position at 16.00 [h]

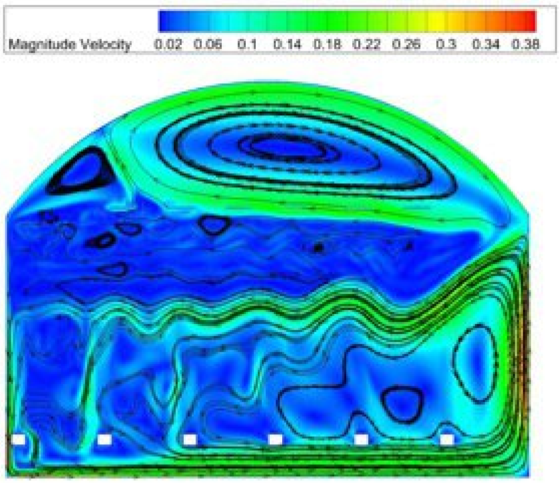


Fig. D.78 Velocity magnitude isocontours and streamlines for 11/10/24 with cucumber plants, and PV panels in horizontal position at 10.00 [h]

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

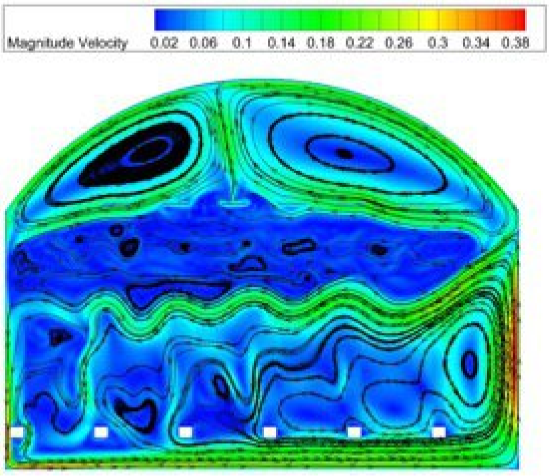


Fig. D.79 Velocity magnitude isocontours and streamlines for 11/10/24 with cucumber plants, and PV panels in horizontal position at 12.00 [h]

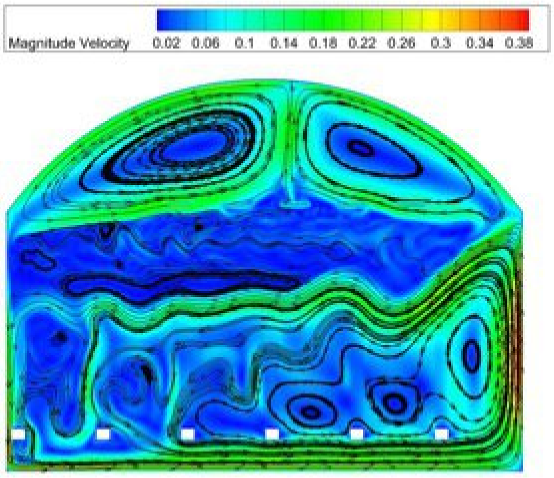


Fig. D.80 Velocity magnitude isocontours and streamlines for 11/10/24 with cucumber plants, and PV panels in horizontal position at 14.00 [h]

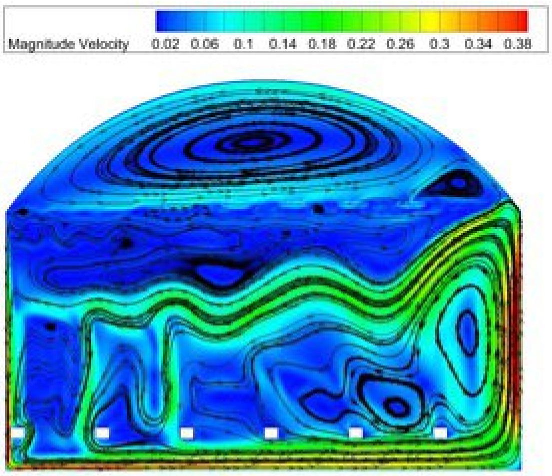


Fig. D.81 Velocity magnitude isocontours and streamlines for 11/10/24 with cucumber plants, and PV panels in horizontal position at 16.00 [h]

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

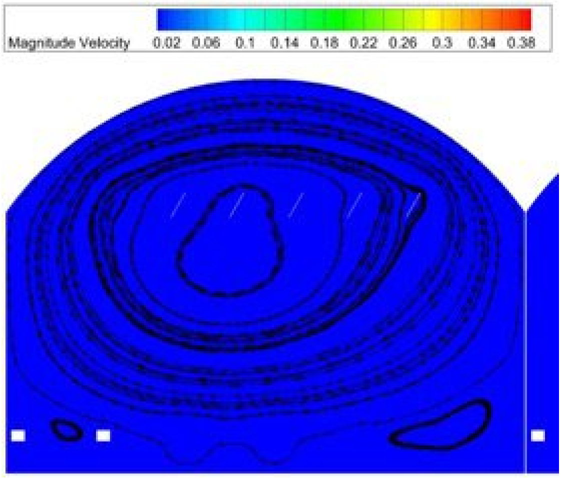


Fig. D.82 Velocity magnitude isocontours and streamlines for 18/4/25 11h with inclined PV and without plants

References for Annex D

[1] ASHRAE FUNDAMENTALS

[2] Technical Chamber of Greece (TCG) 2017. TOTEE 20701-3-2017- Climatic Data of Creeck Teritory. 2945B, National Printing Office, Athens.

[3] https://www.buteline.com/my/my/buteline-pe/technical-information/thermal-characteristics-of-polyethylene

[4] https://www.engineeringtoolbox.com/specific-heat-polymers-d_1862.html

[5] BPF. "Polyethylene (Low Density) LDPE, LLDPE". www.bpf.co.uk. Retrieved 2018-06-14.

[6] https://matmatch.com/materials/mbas006-linear-low-density-polyethylene-lldpe-

[7] Arblaster, John W. (2018). Selected Values of the Crystallographic Properties of Elements. Materials Park, Ohio: ASM International. ISBN 978-1-62708-155-9.

[8] https://www.matweb.com/search/datasheet.aspx?bassnum=AMEAL00&ckck=1

[9] Miguel Ángel Sanjuán , Ángel Morales and Aniceto Zaragoza, Effect of Precast Concrete Pavement Albedo on the Climate Change Mitigation in Spain, Sustainability 2021, 13, 11448. https://doi.org/10.3390/su132011448

[10] An index based road feature extraction from LANDSAT-8 OLI images

[11] Simone Kotthaus, Thomas E.L. Smith, Martin J. Wooster, C.S.B. Grimmond, Derivation of an urban materials spectral library through emittance and reflectance spectroscopy, ISPRS Journal of Photogrammetry and Remote Sensing 94 (2014) 194–212

[12] Optical and thermal properties of commercial polymer film, modeling the albedo effect, J of Applied Polymer Sci, Volume: 138, Issue: 24, First published: 12 February 2021, DOI: (10.1002/app.50581)

[13] "Density of water and temperature - Physical Sciences, Chemistry and Biology". 10 February 2009.

[14] Brewster MQ (1992) Thermal radiative transfer and properties. Wiley, New York

[15] Fazal M. A., A. S. M. A Haseeb, Farazila Yusof, Laser-based Surface Modifications of Aluminum and its Alloys, Critical Reviews in Solid State and Material Sciences · October 2015

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.

Funded by the European Union

Annex E – List of variables, features, and target names of the models (Chapter 6)

List of variable used for the training of the low-cost model:

T_ext

Sun

RH_ext

Wind

Hour

Month

T_ext_lag_1

T_ext_lag_2

T_ext_lag_6

T_ext_lag_12

Sun_lag_1

Sun_lag_2

Sun_lag_6

Sun_lag_12

Delta_T_lag_1

Delta_T_lag_2

Delta_T_lag_6

Delta_T_lag_12

Sun_mean_1h

Hour_sin

Hour_cos

Target: Delta_T

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union

Annex F – List of varible used for all the other models (Chapter 6)

CO2_env	Temp_5_ground	Temp_2_below	T_110	TDS_101	Target_lag_5m
LUX_env	LUX_5_adj	RH_2_below	T_111	EPS_101	GHI_lag_10m
GHI_env	PAR_5_adj	CO2_2_ground	T_112	EC_102	TempExt_lag_10m
RH_env	Temp_8_ground	CO2_2_adj	RH_80	SS_102	Target_lag_10m
Temp_env	RH_8_ground	CO2_2_below	RH_81	TDS_102	GHI_lag_30m
Ws_env	Temp_8_adj	LUX_2_adj	RH_82	EPS_102	TempExt_lag_30m
Wd_env	RH_8_adj	PAR_2_adj	RH_83	EC_103	Target_lag_30m
Wr_env	Temp_8_below	RH_10_above	RH_101	SS_103	GHI_lag_60m
Wind_speed_id_90_m_per_s	RH_8_below	Temp_10_above	RH_102	TDS_103	TempExt_lag_60m
Wind_force_id_90_m_per_s	CO2_8_ground	CO2_10_above	RH_103	EPS_103	Target_lag_60m
Wind_dir_id_90_files	CO2_8_adj	GHI_10_above	RH_104	EC_104	GHI_lag_120m
Wind_dir_2_id_90_deg	CO2_8_below	RHI_10_above	RH_105	SS_104	TempExt_lag_120m
RH_4_ground	LUX_8_adj	CO2_5_ground	RH_106	TDS_104	Target_lag_120m
CO2_4_ground	PAR_8_adj	CO2_5_below	RH_107	EPS_104	Delta_T_Int_Ext
CO2_4_adj	GHI_5_adj	CO2_5_adj	RH_108	EC_105	GHI_delta_5m
CO2_4_below	CO2_6_adj	RH_5_below	RH_109	SS_105	TempExt_delta_5m
RH_4_below	CO2_6_below	Temp_5_below	RH_110	TDS_105	GHI_roll_mean_2h
Temp_4_below	CO2_6_ground	RH_5_adj	RH_111	EPS_105	Target_roll_mean_1h
RH_4_adj	RH_6_ground	Temp_5_adj	RH_112	EC_106	Target_lag_5m
Temp_4_adj	Temp_6_ground	RH_5_ground	WF_WEST_PV	SS_106	GHI_lag_10m

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

LUX_4_adj	RH_6_below	Temp_7_adj	WF_EAST_PV	TDS_106	TempExt_lag_10m
PAR_4_adj	Temp_6_below	RH_7_adj	WF_EAST_no_PV	EPS_106	Target_lag_10m
CO2_1_below	RH_7_ground	Temp_7_below	WF_WEST_no_PV	EC_107	GHI_lag_30m
Temp_1_ground	RH_6_adj	RH_7_below	P	SS_107	TempExt_lag_30m
RH_1_ground	Temp_6_adj	CO2_7_ground	WF	TDS_107	Target_lag_30m
Temp_1_adj	LUX_6_adj	CO2_7_adj	T	EPS_107	GHI_lag_60m
RH_1_adj	PAR_6_adj	CO2_7_below	East_no_PV	EC_108	TempExt_lag_60m
Temp_1_below	Temp_7_ground	LUX_7_adj	WEST_no_PV	SS_108	Target_lag_60m
RH_1_below	RH_8_ground	PAR_7_adj	EAST_PV	TDS_108	GHI_lag_120m
CO2_1_ground	Temp_8_adj	RH_9_above	WEST_PV	EPS_108	TempExt_lag_120m
CO2_1_adj	Temp_5_ground	Temp_9_above	EC_80	EC_109	Target_lag_120m
LUX_1_adj	LUX_5_adj	CO2_9_above	SS_80	SS_109	Delta_T_Int_Ext
PAR_1_adj	PAR_5_adj	GHI_9_above	TDS_80	TDS_109	GHI_delta_5m
CO2_3_ground	Temp_8_ground	RHI_9_above	EPS_80	EPS_109	TempExt_delta_5m
CO2_3_below	Hour_cos	Potenza_W	EC_81	EC_110	GHI_roll_mean_2h
CO2_3_adj	Hour_cos	T_80	SS_81	SS_110	Target_roll_mean_1h
RH_3_below	RH_8_adj	T_81	TDS_81	TDS_110	Target_lag_5m
Temp_3_below	Temp_8_below	T_82	EPS_81	EPS_110	GHI_lag_10m
RH_3_ground	RH_8_below	T_83	EC_82	EC_111	TempExt_lag_10m
Temp_3_ground	CO2_8_ground	T_101	SS_82	SS_111	Target_lag_10m
RH_3_adj	CO2_8_adj	T_102	TDS_82	TDS_111	GHI_lag_30m
Temp_3_adj	CO2_8_below	T_103	EPS_82	EPS_111	TempExt_lag_30m

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting ty can be held responsible for them.
Funded by the European Union

LUX_3_adj	LUX_8_adj	T_104	EC_83	EC_112	Target_lag_30m
PAR_3_adj	PAR_8_adj	T_105	SS_83	SS_112	GHI_lag_60m
Temp_2_ground	GHI_5_adj	T_106	TDS_83	TDS_112	TempExt_lag_60m
RH_2_ground	CO2_6_adj	T_107	EPS_83	EPS_112	Target_lag_60m
Temp_2_adj	CO2_6_below	T_108	EC_101	GHI_lag_5m	GHI_lag_120m
RH_2_adj	CO2_6_ground	T_109	SS_101	TempExt_lag_5m	TempExt_lag_120m
Hour_sin					

by the European Union under Grant Agreement No 101096056. Views and opinions expressed are however those of the
s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting
ty can be held responsible for them.

Funded by the
European Union