Project Deliverable Report

DELIVERABLE 4.3

DIGITAL TWIN

WORK PACKAGE NUMBER: WP4

WORK PACKAGE TITLE: MODELLING & DIGITAL TWINS

TYPE: GITHUB REPOSITORY (DESCRIPTION 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/

Funded by the European Union

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

Document Information

Deliverable Information	
Work package number	**4**
Work package title	**Modelling & Digital Twins**
Deliverable number	**4.3**
Deliverable title	**DIGITAL TWIN**
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**
Revision number	**1**
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**

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

Document History			
Revision	**Date**	**Modification**	**Author**
0.1	2025/12/11	Outline	Cristina Cornaro
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

Funded by the European Union

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

Table of Contents

Funded by the European Union

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

Executive Summary

The activities described in this deliverable focus on the development and organisation of the software infrastructure supporting the Digital Twin of a novel photovoltaic (PV) greenhouse system within the REGACE project.
This deliverable presents the structure and content of the dedicated GitHub repository where all software developed for the Digital Twin is stored and maintained. The repository serves as a centralised and reproducible workspace for modelling activities, including data preparation, exploratory analysis, model training, validation, and performance assessment. It contains a collection of Jupyter notebooks implementing machine-learning and physics-informed modelling approaches, alongside reusable Python utilities that standardise data handling and model execution across different experiments and computing environments.
Due to size and licensing constraints, large input datasets are managed externally to the repository, in accordance with project data management policies, while the modelling logic and workflows remain fully version-controlled. The repository is designed to operate within a standard Python scientific computing environment and supports execution in common development platforms such as JupyterLab and Visual Studio Code.

Chapter 1 – Repository description

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.

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

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 the repository. This ensures that only the modelling logic is versioned, while the heavier artefacts remain local to the user environment.

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

Conclusion

This deliverable documents the establishment and structure of the software repository supporting the development of the REGACE Digital Twin for photovoltaic greenhouse systems. The repository provides a coherent, reproducible, and well-organised environment for implementing, testing, and validating the advanced modelling techniques developed within the project.

By centralising modelling workflows, data-processing utilities, and predictive model implementations, the repository ensures transparency, traceability, and reusability of the Digital Twin components, while complying with data management constraints through external handling of large datasets. The use of standard scientific Python tools and documented Jupyter notebooks facilitates accessibility and future extensibility.

Overall, this deliverable lays the technical foundation for continued Digital Twin development, collaboration among project partners, and future integration of the modelling framework into operational and decision-support applications for agrivoltaic greenhouse systems.

Funded by the European Union

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