D5.3 Research Advancements for the Pilots

HiDALGO2
CENTRE OF EXCELLENCE

Date: November 23rd, 2023

EuroHPC
Joint Undertaking

Document Identification			
Status	Final	**Due Date**	31/10/2023
Version	1.0	**Submission Date**	23/11/2023

Related WP	WP5	**Document Reference**	D5.3
Related Deliverable(s)	Insert Related Deliverables	**Dissemination Level (*)**	PU
Lead Participant	MeteoGrid (MTG)	**Lead Author**	Luis Torres
Contributors	PSNC, SZE, UNISTRA, MTG	**Reviewers**	Dennis Hoppe, USTUTT
			László Környei, SZE

Keywords:
Research, Research Lines, Research Management, Challenges, State of the Art, Pilot Roadmap, Research Advancements

Document Information

List of Contributors	
Name	**Partner**
Luis Torres	MTG
David Caballero	MTG
Jaime Ribalaygua	MTG
Luca Berti	UNISTRA
Christophe Prud'homme	UNISTRA
Michał Kulczewski	PSNC
Wojciech Szeliga	PSNC
Zoltán Horváth	SZE

Document History			
Version	**Date**	**Change editors**	**Changes**
0.1	23/10/2023	MTG	First version of the document.
0.2	01/11/2023	MTG	Major changes including restructuring the document contents.
0.2	02/11/2023	UNISTRA	Response to review comments and inclusion of UB roadmap.
0.2	06/11/2023	PSNC	Response to review comments and inclusion of RES roadmap.
0.2	07/11/2023	SZE	Inclusion of UAP roadmap.
0.2	08/11/2023	MTG	Response to review comments and inclusion of WF roadmap. Format setting.
0.3	14/11/2023	MTG	Response to review comments on headers and Document Information section.
0.3	15/11/2023	SZE	Response to review comments,
0.3	15/11/2023	MTG	Shorten Chapter 8 length, review Chapter 9. Conclusions.
0.3	15/11/2023	SZE	Review of references.
0.3	16/11/2023	PSNC	Review of references.
0.3	16/11/2023	MTG	New roadmap addition in Conclusions. Formatting and proofreading.
0.4	21/11/2023	MTG	Roadmap update and comments review.
1.0			FINAL VERSION TO BE SUBMITTED

Quality Control		
Role	**Who (Partner short name)**	**Approval Date**
Deliverable Leader	Luis Torres (MTG)	21/11/2023
Quality Manager (deputy)	Dennis Hoppe (USTUTT)	21/11/2023
Project Coordinator	Marcin Lawenda (PSNC)	23/11/2023

Table of Contents

List of Tables

List of Figures

List of Acronyms

Abbreviation / acronym	Description
AI	Artificial Intelligence
CFD	Computational Fluid Dynamics
CNNs	Convolutional Neural Networks
CoE	Centre of Excellence
DBN	Deep Belief Networks

DEIM	Discrete Empirical Interpolation Method
DESA	United Nations Department of Economic and Social Affairs
DEM	Data Elevation Model
Dx.y	Deliverable number y belonging to WP x
EC	European Commission
ECMWF	European Centre for Medium-Range Weather Forecasts
FMI	Functional Mockup Interface
FMUs	Functional Mockup Units
GFS	Global Forecast System operated by US National Weather Service
GIS	Geographic Information System
GMSH	Finite-element mesh generator (software https://gmsh.info)
HPC	High Performance Computing
IFC	International Foundation Class
HPC	High-Performance Computing
LOD	Level Of Detail
POD	Proper Orthogonal Decomposition
RAWS	Remote automated weather stations
RedSIM	RedSIM is the name of the software referring to Reduced Simulation
RES	Renewable Energy Sources
ROM	Reduced Order Method
SVD	Singular Value Decomposition
SVT	Sparse Volume Textures
UAP	Urban Air Project
UB	Urban Building
UBEM	Urban Building Energy Modelling
UN	United Nations
UQSA	Uncertainty quantification and sensitivity analysis
VDB	OpenVDB file format
VR	Virtual Reality
VVUQ	Verification, Validation and Uncertainty Quantification
WF	Wildfires
WFDS	Wildland-Urban Interface Fire Dynamics Simulator
WRF	Weather Research Forecast Model
WP	Work Package
WUI	Wildland-Urban Interface

Executive Summary

In pursuit of addressing global challenges and to provide stakeholders and decision makers tools that allow to mitigate tragic consequences, HiDALGO2 has embarked on a mission to integrate, implement, and optimize selected environmental use cases using advanced simulation systems. These endeavours, owing to their intricacy, require the utilization of HPC (High-Performance Computing) and big data/AI technologies, ultimately propelling the research and development of novel approaches, methods, and tools.

HiDALGO2 presents a coherent offer in which scientific institutions and SMEs of the consortium can work together in the development of innovative simulation models both for their scientific excellence and their technological performance. The technological mastery and development allow us to use the potential of the use cases contributing in the advance towards the idea of digital twins, which in turn generates the need to use computing resources at a much larger scale than before (Exascale).

This deliverable offers a comprehensive foundation of the scientific aspects, establishing the baseline for the chosen use cases and highlighting the associated research challenges. These use cases integrated in thematic pilot studies are instrumental in achieving the set objectives. Moreover, it outlines a strategic roadmap for the execution of research activities, the design of a research coordination plan, and the delineation of measurable outcome indicators for monitoring and evaluation.

In addition, this deliverable provides an overview of the initial research progress in each pilot use case. This groundwork is essential for shaping future developments, setting the stage for the formulation of a roadmap that includes specific objectives, tasks, and milestones. Each use case will undergo the creation of at least one scenario, serving as a foundation for future enhancements and research outcome implementations, ultimately paving the way toward scientific excellence.

HiDALGO2 considers the development of Use Cases in the following pilots:

> **Urban Air Project (UAP).** To achieve the main goals of the UAP, i.e. to create, operate, and provide computational solutions to city policy makers for planning and regulation, and to inform citizens about the state of air quality and wind comfort, the following research challenges must be addressed: considering new physics in the air quality simulation solution, improving current solvers, coupling the solution with wildfires (WF) and urban buildings (UB), and improving workflow orchestration.
>
> **Urban Buildings (UB).** The primary objectives of the UB pilot are to estimate and describe the energy consumption in the building sector. Two research lines are considered in this pilot: 1) the generation of the city's geometry from GIS data, which will provide a watertight mesh that will be the basis for both building simulation and

coupling with the UAP pilot; 2) the building description, for which different LOD (Level of Detail) have been identified: from a simplified bounding box description to the generation of a multi-zone model from an IFC file.

Renewable Energy Sources (RES). With the increasing number of renewable energy sources installations, such as photovoltaic and wind farms, is it important for DSOs, but also for individuals, to understand the physical phenomena and maximise the outcome. In RES focus is given on understanding how weather conditions influences the energy production and provide more detailed prediction of both. Three specific challenges are tackled: Energy production from the wind farms, Energy production from photovoltaic systems, Damages prediction to the DSO infrastructure.

Wildfires (WF). WF pilot aims to extend forest fire simulation to two scales, the landscape scale and the wildland-urban interface scale, and to estimate their potential consequences. This dual approach involves a twofold commitment: firstly, the development of bespoke solutions tailored to enable superior, more efficient simulations in HPC environments. Secondly, a concerted effort to enhance the methods and procedures used in practical scenarios, with a special focus on operational perspectives.

In the next iteration of this deliverable, the FAU use case will be considered too, as at the time of writing this document, the legal contract was not finalized.

1. Introduction

The key aspect of HiDALGO2 project is tackling global challenges via HPC and big data/AI technologies, which then needs work on integration, implementation, and optimization of selected use cases from the environmental domain. This leads to the research and development of new approaches, methods, and tools.

In this report a scientific and technical background (technological and scientific base line) is provided, as well as the identification of the research challenges of each of the selected use cases (pilot studies) to solve the stated objectives. Also, a strategy for the implementation of the research activities, including the assembling of thematic working groups, the design of a research coordination plan, the establishment of collaboration activities and the description of indicators of measurable results for monitoring and evaluation.

Besides, a description of the preliminary research advances in each pilot use case is provided. All this sets the scene for future developments by drafting a roadmap which includes specific goals, tasks, responsibilities, and deadlines. At least, one scenario per use case will be considered and developed, as a foundation for future improvements and implementations of the research outcome, paving the path for scientific excellence.

1.1 Purpose of the document

The objective of this document is to serve as a reference for the research activity within the project life cycle, providing a consistent guideline, a time framing, and a thematic rationale to set-up the scope of the planned research and the expected outcomes. Also, to highlight possible bottlenecks in software, specify details regarding data transfer, reflecting on software and hardware requirements defined in D2.1, and identifying scenarios to be use case as a foundation for the future developments and code improvements towards scientific excellence.

The scope of the document and, consequently, the research activity is closely related to the objectives and challenges of HiDALGO2. In this sense, the research outcome will be justified by and limited to the requirements and challenges identified in the use cases, notwithstanding the above opening new research lines to be shared and jointly developed in the future with other research groups. Some of these challenges include improvements in the efficiency and reliability of the results, in the resolution of the phenomena described, the finding, development and coupling of new solutions, etc.

1.2 Relation to other project work

This deliverable establishes the research lines to be followed in each of the use cases of the project and includes a roadmap of the main results expected in the development of the WP5 activities. It also states what are the functions to be performed in the Scientific Coordination of the project, so it is also related to other WPs in which research results are expected as WP3 and WP4, as well as with WP6 by incorporating within these tasks the publication of scientific papers and the establishment of scientific and technical collaborations with other entities and projects.

1.3 Structure of the document

This document is organized into several chapters, the contents of which are shown below.

Executive Summary. It frames the context of the document in the project, summarizes the objectives of this document and briefly describes its main contents.

Chapter 1. *Introduction* It establishes the purpose of the document, how its fit within the project framework and the structure of the document.

Chapter 2. *HiDALGO2 Pilots*. It contains a brief description of each of the pilot cases and their use cases.

Chapter 3. *Methodology*. It includes the functions to be carried out by the Scientific Coordination, the indicators defined for monitoring the activities and a summary of the expected results to be achieved in HiDALGO2 from the research point of view.

Chapters 4 to 7. Includes the research to be conducted by each of the pilots. Each of these chapters focuses on a single pilot and describes its scientific baseline, its main requirements, the results obtained up to M10 and the results expected during the life of the project.

Chapters 8. *Conclusions*. The main scientific and technological challenges that frame the main lines of research in each pilot are presented in the form of conclusions.

2. HiDALGO2 Pilots

This section includes a comprehensive summary of the most relevant technologies and developments of the research areas related to each pilot, to establish the background and identify the yet unsolved challenges.

2.1 Urban Air Project (UAP)

2.1.1 Use Case Definition

Urban challenges affect the majority of the Earth's population since 55% of the world's population lives in urban areas now, and it is estimated that 68% will live in urban areas by 2050 [1uap]. Urban air pollution is a critical factor for societal well-being in urban areas. Urban air pollution poses significant health risks, leading to respiratory and cardiovascular diseases, environmental degradation, and economic burdens. The WHO states that 6.7 million deaths are attributable each year to exposure to ambient and household air pollution [2uap]. Furthermore, urban wind comfort is another critical factor that directly influences pedestrian safety, the overall urban experience, and energy efficiency [3uap].

These issues are central to urban planning and design, too, as they impact public health, quality of life, and the sustainability of urban environments. Addressing these challenges is essential for creating healthier, safer, and more liveable cities.

The HiDALGO2 Urban Air Project (UAP) application creates, operates, and provides HPC, HPDA, and AI-based computational solutions for city policymakers, acts as a source of information for citizens, and offers tools for researchers to assist and improve urban planning, daily well-being, and environmental research, respectively. Namely, UAP solutions simulate the urban airflow and the pollutant dispersion under actual urban conditions, which are characterized by real 3D building geometry, emissions, and the neighbouring, rural weather. Thanks to the applied HPC technology, the urban airflow and pollutant concentration simulations are more accurate than those given by traditional computational technologies and at the same time the results are faster at users' disposal. The applied HPDA technology allows the users a fast evaluation of the simulation results for getting essential end-user measures like the number of pollution limit exceedance days from a full year simulation results and wind comfort parameters, which needs a histogram computation from wind speed data for the entire simulation domain. Recently, AI technologies has been improved to such a level that AI-based predictions are competitors even for the flagship HPC operational solution of the global weather simulations [3uap]. In HiDALGO2, physics-based (e.g., POD-DEIM) and purely data-based AI technology (special deep learning methods) will be applied. In the resulting solutions, HPC is mainly used for compute-intensive training, in order to get a computationally cheaper operational solution than the traditional simulations.

Main algorithms and requested data

The core of UAP is the airflow computation in cities, which needs computational fluid dynamics (CFD) algorithms for 3D unstructured meshes. The physical conservation laws describing fluid flow can be formulated with the well-known Navier-Stokes equations. For urban air flow some other physical effects play also important role, such as solar heat radiation, modelling properly thermal processes, modelling inhomogeneity of the air (e.g., humidity), and taking into account the detailed geometry including, among others, trees. Coupling with the airflow simulation, the gaseous pollutant dispersion is modelled by the linear advection-diffusion equations with emission sources. The main challenge here is the proper selection of the diffusion coefficients, which are affected by the general weather conditions (e.g., solar radiation) and the airflow turbulence.

UAP takes boundary conditions from an external weather database, external global or regional weather simulation model (often from WRF simulations), or real-time sensors (for digital twin applications). Since WRF is used by all other use cases of HiDALGO2, an in-house WRF installation will serve the HiDALGO2 pilots.

As of project month 10, for the airflow simulations, the Navier-Stokes equations have been solved; other physical properties will be modelled later in the project. The main challenge in solving the Navier-Stokes equations is that the Reynolds number of the airflow in cities is huge hence the airflow is turbulent. This challenge is still under scientific research by many projects and research teams. There is no unique solver that could be optimal for all features (scale, accuracy, running time, usability, etc.) [4uap]. For the UAP application, two solvers, the UAP-openFOAM and RedSIM have already been integrated, and currently, a third solver, xyst is under testing and integration into the UAP framework; these solvers are introduced below. Notice that in HiDALGO2 solvers for airflow simulations, general purpose solvers have been chosen in order to be able to integrate modelling further physical features to assist the actual societal challenge and also be able to apply for other purposes (e.g., other global and industrial challenges).

The first Navier-Stokes solver in HiDALGO2 assumes incompressibility of the air, which is a reasonable assumption under normal conditions, and models the turbulence of air with the k-epsilon closure equations. The well-known OpenFOAM software solves this model, in particular, the pimpleFOAM and simpleFOAM solvers are used for unsteady and steady equations, respectively. Notice that several EuroHPC projects use OpenFOAM for different applications, for examples, see[1].

[1] https://exafoam.eu/, https://services.excellerat.eu/searchcodes/core

The second solver is a fully SZE-developed code called RedSIM[2]. RedSIM solves the compressible Navier-Stokes and the compressible Euler equations. RedSIM is genuinely designed and implemented for GPUs and from the same source code it supports multi-GPU and MPI+Pthread GPU platforms. In HiDALGO2 an MPI+multi-GPU implementation and a low-Mach solver will be implemented, too.

The third solver, the xyst (see [5uap]) is under testing in HiDALGO2. It is a free, open-source finite element-based CFD code. One of the most developed features of the xyst is the automated load balancing feature, which was exploited by several previous exascale projects of the Qinoa-code from which xyst has been forked and further developed.

Scalability

The initial scalability of the OpenFOAM and RedSIM solvers are well documented in the predecessor project HiDALGO's deliverables, of which several are publicly available [6uap]. As a result, the OpenFOAM solver was well-scalable up to 100,000 CPU cores on Hawk, USTUTT's supercomputer. In HiDALGO2, the initial and current status of the solvers were benchmarked on EuroHPC machines LUMI, Meluxina, VEGA, Discoverer and the Hungarian NCC's Komondor[3] machine; these benchmarks confirmed the scalability of the OpenFOAM-based solver up to 100,000 CPU-cores and the GPU-scalability of RedSIM up to 8 GPUs of the Komondor.

2.2 Urban Buildings (UB)

2.2.1 Use case definition

The Urban Buildings pilot aims to offer simulation tools to predict energy consumption, thermal comfort, and air quality at the building and urban scales. Evaluating these quantities is fundamental for addressing energy losses and enhancing the well-being of citizens. Indeed, in 2020, the energy inefficiencies of the European building stock have been reported by the European Commission, and it has been estimated that 40% of the total energy consumption is attributable to the building sector [1ub]. Moreover, buildings are estimated to generate 36% of EU greenhouse gas emissions [1ub] and studies indicate that the main reason is their energy use [2ub]. Yet, every year, only 1% of the building stock undergoes energy renovation.

The approach proposed by the UB pilot is two-fold: on the one hand, simulations at the city level will enable decision makers to identify the regions of higher energy consumption and higher pollutant emissions. The building models that will be employed in this simulation mode will be simplified in order to treat large geographical regions; on the other hand, simulations

[2] Note that in the predecessor project HiDALGO this code was named Fluid-Solver; the change of the name of the code was motivated by the new reduced order operational module of the code.

[3] https://hpc.kifu.hu/en/komondor-one-of-the-greenest-supercomputers-in-the-world

at the building scale could feature more precise physical models to better understand the sources of energy loss and suggest actions to improve the overall efficiency.

In both cases, the shading effects of the neighbouring buildings will be taken into account in order to evaluate the actual solar radiation that each building receives, as well as the energy sources that are consumed by each building and their impact on the pollution generation.

Main algorithms and requested data

The UB pilot uses multi-zone models for the LOD-0 and LOD-1 description of buildings, ensuring a simple yet comprehensive modelling of the thermal phenomena, energy transfers and the pollution generated by buildings. More accurate parametric models can be provided using model order reduction (based on finite element analysis), especially the reduced basis method. Other algorithms that are considered in the use case perform data assimilation and parameter estimation using ensemble methods, as well as Monte Carlo algorithms for the computation of solar shading charts (solar masks). The following tools are used to generate the models and to simulate them: Dymola (based on Modelica language), fmpy (Python) and fmi4cpp (C++) for the execution of FMUs in co-simulation mode, Ktirio-city and GMSH for geometry generation, and Feel++ for the core components as well as reduced basis models, mesh handling and solar mask computation.

The UB application requires at least the following input data:

- GIS description of district/city to extract geometric and material information on the building envelope;
- Climate data for the building models;
- Building construction standards;
- IFC files (for the higher level of detail among building models); and
- Use/occupation scenarios, energy consumption and source (electric, gas, fuel).

Further data on construction assemblies, HVAC systems, and also sensor data would be used to increase the accuracy of the models.

The output data that the UB application generates are:

- heat flux and temperature on building surfaces;
- greenhouse gas emissions (especially CO_2 and NOx) per building;
- city geometrical discretization; and
- solar shading mask per building surface.

Scalability

The matter of scalability is addressed using domain decomposition and parallel-in-time algorithms. The first approach consists in partitioning the computational domain and simulating the buildings of each subdomain on different cores. Difficulties arise if the buildings on different partitions belong to a network (electrical, geothermal), as communication among subdomains is required, while in the presence of isolated buildings the problem becomes embarrassingly parallel. The second approach leverages the studies on the parallelization in the time dimension, where the simulation's time interval is also partitioned, and the problem is simultaneously solved on the temporal subdomains.

The scalability study of the UB application is under development in work package 3 of the HiDALGO2 project, using the domain decomposition approach. First results are expected at the end of Y1 since most of the software used by the UB application has been developed during the first year of the HiDALGO2 project and there are no prior results on its scalability.

2.3 Renewable Energy Sources (RES)

2.3.1 Use case definition

Social technological development and increase in the number of people causes a significant increase in the demand for electricity. Alternative energy sources are an important part of the electricity production process in the world. Most of all, they reduce the dependence on fossil fuels, increasing the share of environmentally friendly energy sources and limiting the greenhouse gas emissions. There are some challenges related to the emerging wind farms and photovoltaic system. It is of the utmost importance to understand the physical phenomena behind renewable energy sources in order to be able to control and increase productions of such energy. Individual users are keen to know how much savings there will be comparing to traditional electrical energy sources. Wind and photovoltaic farms owners seek for means to increase their efficiency. They look for a solution to i) find the best location for a new site, ii) understand the interaction between wind turbines, or iii) control wind turbines and solar panels position to maximize energy production. Knowledge of how much energy will be produced is important to Distribution System Operators (DSOs) to stabilize the grid and optimize incomes from trading energy in the energy market. Detailed weather predictions is an important source of information for plant operators, as they allow to estimate energy production in selected future time periods. Many wind energy DSOs rely on statistical approaches and general weather forecasts. The statistical approach is based on analysis of the amount of produced energy, trying to find a correlation with actual weather conditions. More sophisticated solutions are required by the market, taking into account more detailed weather prediction and AI solution in prediction of energy production.

Use case

The Renewable Energy Sources (RES) pilot is driven, among others, by the requirements of one of the largest DSO in Poland. It addresses its needs by providing a high-resolution, detailed weather prediction, which can be then correlated with data from wind farms and photovoltaic panels to predict how much energy can be produced.

Components

RES is a multiscale application which combines mesoscale and regional weather forecast model with local detailed one for accurate modelling of wind flows over complex terrain topography and through building structures for urban areas. The same models are applied for solar energy plants owners and operators for detailed sun/shade forecasts. Yet another scenario RES implements is the prediction of damages to infrastructure. The aforementioned models are WRF and EULAG. The former is an open-source, community-based model. The latter is an all-scale geophysical flow solver, which in RES is tailored towards simulating flows over complex terrain topography and in urban areas. Figure 1 presents workflow between components. Data obtained from WRF forecast is put into EULAG. Initial data with a resolution of several kilometres is supplemented with topography data and information about buildings, in case an urban scenario is considered. RES.WIND and RES.WIND_URBAN allow to receive separate wind data for each wind turbine, and the final results are combined by RES.WIND_ENERGY to calculate energy production. For solar energy, RES.SOLAR and RES.SOLAR_ENERGY are used respectively.

Data

RES aims at using publicly available data, which are used as boundaries for input parameters. Global Weather Forecast data is used for initial weather data, but it can be substituted with another model. SRTM and EU-DEM repositories are used for the terrain topography. The land cover information is obtained from the publicly available CORINE repository. For the urban areas, the building information (shape and height) is delivered by interested cities, but publicly available data can be used. The used data and its flow are presented in Figure 1.

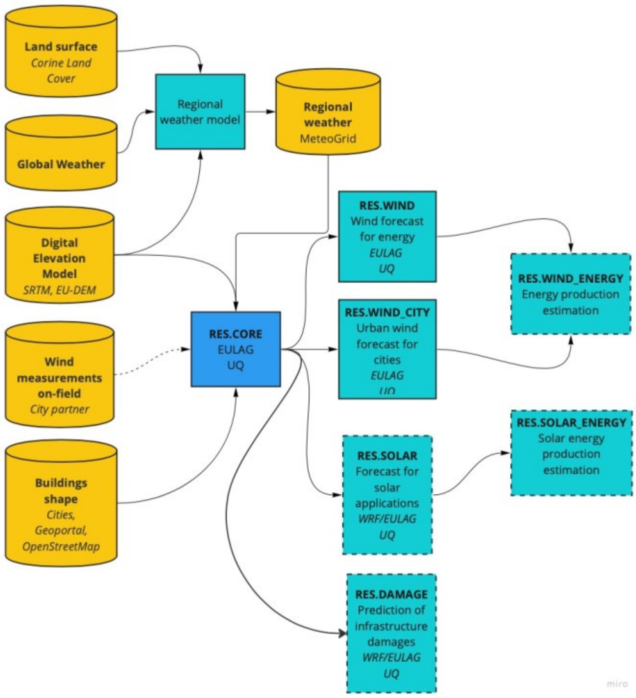

Figure 1. UAP Components and data flow.

Scalability

RES models are mature and known for their scalability. It depends on the size of the domain and mesh resolution. For the scenarios foreseen by this pilot, current runs were limited by the means of the coarser mesh to find a trade-off between the cost of the computations and time-to-solution. Production energy estimation for a single wind farm usually requires thousands of CPU core hours. In case finer mesh is required, even more CPU core hours are demanded. The demand for CPU cores also depends on time step between simulation iterations. For a fixed mesh resolution and time step, making the domain size twice larger in one dimension, requires twice more amount of time for the computations. For a fixed domain size, increasing mesh resolution by two horizontally boosts computation time by four to eight, and vertically by two. In the initial study on uncertainty quantification (same models but different scenario not concerned within HiDALGO2), we were able to use 98,000 CPU cores simultaneously for 256 ensembles.

2.4 Wildfires (WF)

2.4.1 Use case definition

Fire growth models integrate spatial information on fuel types and terrain factors, along with temporal data on evolving weather conditions and fire risk levels. These models work in conjunction with fire behaviour prediction models to replicate the propagation and evolution of fires across the terrain. Fire growth simulation refers to the process of amalgamating diverse fire behaviour models with multi-dimensional mathematical models for forecasting rates of fire spread in intricate, spatially and temporally shifting environmental circumstances. These simulations predict the probable progression of fire over specified time periods using fire behaviour models that encompass aspects such as fuels, weather, fuel moisture, and topography to forecast the spread rate, flame height, and intensity in the vegetation affected by the fire.

Broadly speaking, fire simulation models can be categorised into two groups based on the technique employed: those reliant on raster analysis and those reliant on vector analysis. Until now, the prevailing method employed has been the raster or cellular-automata strategy, as detailed by Finney in 2004 [1wf]. This approach simulates fire expansion as a distinct process involving ignitions distributed across a regularly spaced grid covering the landscape. In the cellular technique, each grid cell possesses a specific status (burning, burnt, or unburnt), and established rules dictate how fire advances from burning cells to unburnt ones, contingent on cell conditions and attributes such as fuel slope, weather, and more.

Two use cases have been identified), with two specific scales and scenario definitions, which may be of interest: the landscape level and the settlement level.

Scenario 1. Landscape level. The main objective is to provide simulations of wildfire progression, the energy released and coupled atmosphere-fire interactions, particularly the disturbance of wind fields and the generation of pyro-convective movements, as well as smoke release and dispersion. The generation and dispersion of incandescent particles (sparks) are also included. For this purpose, coupled fire-atmosphere models, WRF-SFIRE, will be used, including smoke dispersion with models such as WRF-CHEM.

Scene 2. Urbanization level. The main objective is to provide simulations of fire behaviour at a flame scale (a few meters). This includes consideration of atmospheric interactions, including airflow disturbance and the local effect of buildings (geometric bodies) and vegetation (diffuse porous media). Also considered are the generation, transport and emission of sparks, as well as the production and dispersion of local smoke. Comprehensive CFD models such as OpenFOAM and the specific fire module fireFoam are applied.

Data

The following data was identified as required to conduct research:

- High-resolution Digital Elevation Model (DEM), Lidar data and photogrammetry imagery,
- ECMWF and GFS forecast data,
- Meteorological observations.

Scalability

Although the scalability of the WRF and OpenFOAM models has been tested and documented, no tests have been performed before HiDALGO2 with the WRF-SFIRE coupled model in HPC environments. During this first year of the project, tests have been performed with different configurations of this coupled model that allow us to believe that its behaviour will be as expected, although we will have to wait for the first benchmark tests to obtain data on the efficiency of its scaling. We have just finished to couple Reframe and WRF-SFIRE and we expect to have initial benchmark results before the end of this year.

3. Methodology

3.1 Research Coordination

The main objectives of the Scientific Coordination are 1) to ensure coherence between approaches, methods, scientific and technical procedures within and among pilots and working groups, 2) to create and manage an observatory of new technologies and scientific advances that are relevant to the project, 3) to coordinate together with the TCC the dissemination of scientific and technical results in the most relevant international forums, congresses and publications in each domain, 4) to identify and promote possible bridges of international collaboration with other entities, working groups, research centres or companies.

Coordination in research is carried out through participation in the follow-up meetings of the different WPs as well as the TCC and PCC. In addition to the specific advances expected in each of the pilot cases, which are monitored in the periodic WP5 meetings, HiDALGO2 aims to advance in the following research areas: Benchmarking Methodology and Optimization, HPDA and AI, Uncertainty Quantification and New computation technologies.

Benchmarking Methodology and Optimization

Advancing the transition towards Exascale capabilities is one of the most important ambitions of HiDALGO2. However, the process is complex and requires a systematic approach. There are many aspects that cause application scalability to be limited in efficiently running on HPC systems. In addition to in-depth expert knowledge, an appropriate methodology is required to identify them and determine the methods of removal. In addition to the procedure and tools, it specifies the metrics to be tested, which improves the entire process and enables the identification of bottlenecks,

High-Performance Data Analytics and AI

Data analytics and AI methods implemented as part of project tasks also apply to tasks outside the current area of exploration. The enormous datasets resulting from the operation of simulation applications contain information whose importance and significance have not yet been discovered.

This requires automated methods supported by machine learning solutions. The discovered dependencies can potentially set a new perspective on current problems and other, non-standard research directions. HiDALGO2 will build, from an AI point of view, on state-of-the-art algorithms and frameworks such as PyTorch, TensorFlow, or Theano in order to implement the intelligent workflow composition mechanism. From the HPDA perspective, well-adopted tools and their integration ability will be investigated in order to enable in-situ analytics of produced data.

Specifically, the algorithmic aspect of the data analytics part is talked about, in order to implement more HPC-like functionality. HiDALGO2 focuses on the AI application itself in order to improve the training mechanism and in particular, the precision of the tools developed. However, it is of utmost importance that the model training phase encompasses a wide range of data sources to generalize well. Consequently, the implementations will rely on Convolutional Neural Networks (CNNs) or Deep Belief Networks (DBN), which will be tailored to the HiDALGO2 and with this, HPC requirements.

Uncertainty Quantification and ensemble methods

Use case owners will add features to their models, so it will be possible to improve their accuracy and performance, representing a significant progress beyond the state of the art. The project will deploy a Verification, Validation and Uncertainty Quantification (VVUQ) framework on the supercomputer infrastructures for sensitivity analyses. The characterising input uncertainties are propagated through a computational model in order to perform statistical or interval assessments on the resulting problem solutions, which need a sufficiently high number of HPC simulations ensemble.

New technologies

Another area where HiDALGO2 goes beyond state-of-the-art is research into new technologies provided by HPC vendors for project participants. As part of the cooperation between such companies offering processors (Intel, AMD, Huawei) and graphics accelerators (NVIDIA, AMD, Huawei), it will be possible to study the impact of newly introduced solutions on the performance of pilot applications and the related potential correction in the code optimization strategy.

Furthermore, the applications will be continuously benchmarked so that different levels of designs and implementations can be evaluated; the newly established EuroHPC JU supercomputers will serve as the foundation for deployment, improving scalability and benchmarking new infrastructures.

Other areas of interest.

Finally, HiDALGO2 also addresses other technical areas like visualisation and Orchestration Solutions for the simulations done, doing research in both areas with the purpose of improving existing tools and incorporating new ones not used before in these areas (i.e., usage of the Unreal Engine 4 for wildfire representation).

To better monitoring their advances and in order to favour the sharing of solutions that can be used in several pilots, the creation of working groups has been promoted to ensure coherence between approaches, methods, scientific and technical procedures within and among pilots:

- WRF Exchange on global numerical modelling of the weather with WRF and provide WRF simulations for the CoE partners. As a first step, several meetings have been held to harmonize the installation of WRF on the different HPC platforms available and installation procedures are being documented to facilitate its use by the different partners and to allow other external groups to follow them. A detailed analysis of the different parametrizations that can be used to improve the reliability of very high-resolution predictions and their sharing for use in wildfires, renewable energies and urban environments is planned.
- Coupling urban simulations. Couple the Urban Air Pollution and Urban Building models and simulations. In this coupled scenario UB will deliver building heat and pollutant emissions for UAP.AQ; and conversely UAP.AF will provide urban air flow information for UB.
- Coupling wildfire and air pollution. Couple the WildFire, the UAP, and maybe some other pilots, and create a demonstration with a domain around Valencia, Spain. This will be achieved by coupling the WRF-SFIRE-CHEM resulting data files as boundary conditions for the urban-pollution modelling. Besides, the geometry and meshing of the urban models will be shared at the urban scale to ensure consistency in the modelling.
- Visualization. Develop visualization tools for the CoE for visualizing results from the portal and dashboard. Three advanced visualization applications will be used in HiDALGO2 use cases: SZE Visualizer, a visualization tool for the postprocessing of general-purpose simulations, written mostly in C++ and compiled via Emscripten into WASM that uses SDL2 and IMGUI as external libraries, an open source visualization application developed by USTUTT that has been designed for parallel processing of huge data-sets from simulations on supercomputers, and an adaptation of Unreal Engine 4 for wildfires and CFD 3D fields simulations.

3.2 Measurable Results and Indicators

HiDALGO2 has among its objectives to establish a working environment that allows the analysis of the different use cases proposed in the Hidalgo2 pilot studies by establishing the most appropriate simulation models for different working resolutions, performance metrics that allow deciding the most efficient solutions in each case, installation and operation procedure guides that facilitate the replication of the results to other groups outside the project, and technical and scientific publications that help disseminate the results obtained and share them with the scientific and technical community.

Use cases that are part of HiDALGO2 present a significant methodological and IT advancement. Scientists from the fields of interest of HiDALGO2 (CFD modelling, meteorology, air quality, energy building efficiency, renewable energies, protection from fire or water management) are looking for new opportunities and support from specialists in the field of HPC, data analytics, AI or advanced workflow orchestration to expand the research workshop.

- Urban Air Pollution project aims to reach pre-exascale performance by running a single simulation on 250,000 cores and ensemble simulations for uncertainty quantification on 475,000 cores. The idea will be put into practice by means of cooperation with four European cities: Győr, Illkirch, Poznań, and Stuttgart.
- Urban Buildings will develop building energy and indoor air quality models to simulate the contribution of the buildings at the city scale, in terms of heat, GHG and NOx emissions. Information about gas emission enriches the UAP pilot by new, never-before-analysed data.
- Renewable Energy Sources will develop more advanced models for wind and solar energy and study the impact of uncertainties in the model, which was not ever done before. The goal in scalability is to develop and run ensemble uncertainty quantification on 98,000 CPU cores.
- Wildfires will concentrate on providing a computational capacity for the simulation of wildfire-atmosphere interactions and smoke dispersion at several scales (a new approach) aimed at the assessment of risk and potential impacts overpopulated areas. Moreover, a new model for probabilistic fire spread in WUI areas will be delivered along with adaptation of FDS algorithms and algorithms for the visual simulation of animated volumetric smoke.

In order to facilitate the monitoring of the achievement of the proposed results, a series of measures have been defined and are included in Section 4.3.

In the list of the different expected results, the related area of research has been identified, indicating to which area the result belongs: New functionalities, Coupling models advancements, Data Analytics, AI procedures, Model Order Reduction methods, Scalability, Data assimilation, Data management and Application efficiency improvements, or Verification, Validation and Uncertainty Quantification (VVUQ).

Although this categorization may be too detailed in principle, it will serve throughout the project to identify the area or groups of interest among the different Pilots and consequently to homogenize the solutions proposed in all of them, thus helping in the coordination tasks. It is expected that this will be reflected in future versions of this document and will help in the follow-up of the activities carried out.

3.3 Expected results summary

The following table shows the expected results according to the Grant Agreement. Although the scalability results are included in the table, these are not scientific results strictly speaking, but technological results, we show them below as they appear in the Grant Agreement.

Table 1. Expected results as stated in the Grant Agreement.

Use Case	Key Framework	Key Algorithms or Models	Current State	Objectives in HiDALGO2	Means to improve scalability
Urban Air Project	- OpenFOAM	- CFD - solving Navier-Stokes equations - model order reductions	Parallel efficiency of OpenFOAM simulation of 69% on a 14M mesh with 16,384 cores	Achieve a per-node parallel efficiency of at least 80% on more than 100,000 CPU-cores with at least a 60M mesh	Improve the fluid-solver in the model reduction module by porting it to GPUs (requires highly scalable SVD methods)
Urban Buildings	- Feel++ - PETSc - Dymola - fmpy	- Heat transfer - model order reduction - uncertainty quantification	Simulation is limited to a district of a city. Execution is sequential and tested for up to 144 buildings	Expand the simulation to multiple cities. Improve scalability by parallelizing the simulation	Partition of city buildings in order to leverage domain decomposition for speed-up. Parallelization of time-dependent algorithms
Renewable Energy Sources	- RES	-WRF -EULAG -includes modelling of wind flows -uncertainty quantification	-Strong trade-off between the cost of the computations and time-to-solution is the resolution of the underlying mesh - Mesh resolution is the current limiting factor	Leverage up to 100,000 CPU cores for simulations	Significantly increase the mesh size to have precise predictions on wind farms

Wildfires	- FireStation - Clark - FDS - OpenFOAM - Unreal Engine 4	- Fire-spread models - flame-atmosphere interactions - CFD - real-time visualisation algorithms	Currently using multi-core PC architectures, from 4 to 8 cores each, and CUDA-based computation in GPUs around 2,000+ cores each	- Increase mesh size to better model interactions with air and smoke - Coupling of weather, fire, and smoke models - Full simulations limited to hundreds of meters.	- Develop a probabilistic fire-spread model - Adaptation of the FDS algorithms - Enhance visual simulation with animated volumetric smoke

The research objectives foreseen within each Pilot are summarized in Table 2, indicating the type of novelty involved[4]. A more detailed description of the lines of research for each Pilot is given in the following chapters.

Table 2. Expected Research Results.

Pilot	**Key framework**	**Research Goal (refer which type of novelty is involved)[3]**	**Current status as of M10**	**Foreseen date of achievement**
Urban Air Project	RedSIM	Scalable compressible CFD solver for 3D unstructured mesh on compute architectures MPI+Pthread, multi-GPU, MPI+multi-GPU, efficient up to 100k CPU cores	Scalable multi-GPU, tested on JU-machines, developed MPI+Pthread done	M18
Urban Air Project	UAP workflow development	Develop UAP workflow to make EuroHPC deployment possible from the portal	Deployment to LUMI is possible from the portal,	M10

[4] New functionality / Coupling models / Data Analytics – AI / MOR methods / Scalability / Visualization advances / Data assimilation / Data management / efficiency / Uncertainty

			under some restrictions	
Urban Air Project	UAP OpenFOAM solver development	Develop the O3-NOx rection mechanism into the OpenFOAM air-flow and emission dispersion	Develop the coupled solver for chemistry to the operational code and its validation, for the latter ReactingFOAM-solver has been developed	M18
Urban Air Project	UAP digital twin workflow	Develop features of coupling sensors and visualization on-the-fly	Continuous sensor integration is done for RedSIM	M36
Urban Air Project	UAP model order reduction module	Develop a POD-DEIM framework for UAP CFD	Implementation in progress, enhanced POD for performance	M36
Urban Air Project	RedSIM MPI+multi-GPU version	Develop a multi-GPU and an MPI+multi-GPU version of the RedSIM	Multi-GPU version has been done and benchmarked	M18
RES	EULAG	Coupling models. Coupling using orchestrator	Static coupling	M24
RES	WRF and EULAG	New functionality. Improved prediction of the damages	Basic version	M18
RES	EULAG, energy module	New functionality. Cloud solution	Planned	M36
RES	EULAG, energy module	Visualization advances. Improve visualization with project tools	Post-processing done with some python and QGIS scripts	M36
RES	EULAG, energy module	Uncertainty quantification. Applied to models	Proof-of-concept	M36

RES	EULAG, energy module	Scalability. Improvement	Planned	M36
RES	Energy module	New functionality. Prediction of the energy produced by renewable energy sources with HPDA/AI	A first version without HPDA/AI, for testing analysis has been developed	M36
UB	Feel++	Efficiency/scalability. Parallelization of shading mask generation	Functional in sequential	M18
UB	Ktirio-city	New functionalities. Watertight mesh generation	Planned	M18
UB	Ktirio-city	New functionalities. LOD-1 mesh generation	Partially available	M12
UB	Dymola	New functionalities. LOD-1 building model generation	In progress	M18
UB	Feel++ Dymola Ktirio-city	Efficiency/scalability. LOD-0 parallel simulation	In progress	M12
UB	Feel++	Data analytics/AI. Leverage of shading mask data for prediction/clustering of buildings	In progress	M24
Wildfires	WRF-SFIRE	Benchmarking. WRF-Sfire preliminary benchmark tests with Reframe	In progress	M11
Wildfires	WRF-SFIRE Reframe	Benchmarking. Testing **WRF-Sfire** scenarios with **ReFrame** for performance evaluation	In progress	M15
Wildfires	WRF-SFIRE Reframe	Data Analytics/AI/ WRF-Sfire HPDA/AI first release	In progress	M12
Wildfires	WRF-SFIRE UQ-QCG	Uncertainty quantification. First installation of WRF-Sfire coupled with uncertainty quantification – first ensembles, without QCG	Planned	M18

Wildfires	WRF-SFIRE UQ-QCG	Uncertainty quantification. Second releases of WRF-Sfire coupled with uncertainty quantification	Planned	M24
Wildfires	WRF-SFIRE UQ-QCG	Uncertainty quantification. WRF-Sfire + UQ/QCG final release	Planned	M48
Wildfires	SFIRE	New functionality. Sfire performance improvement	Planned	M36

4. UAP Pilot

4.1 Scientific Baseline

Figure 2 below shows an overview of the current status of UAP, including the UAP control and data flow. The numerous input data for each actual application are handled in a pre-processing workflow.

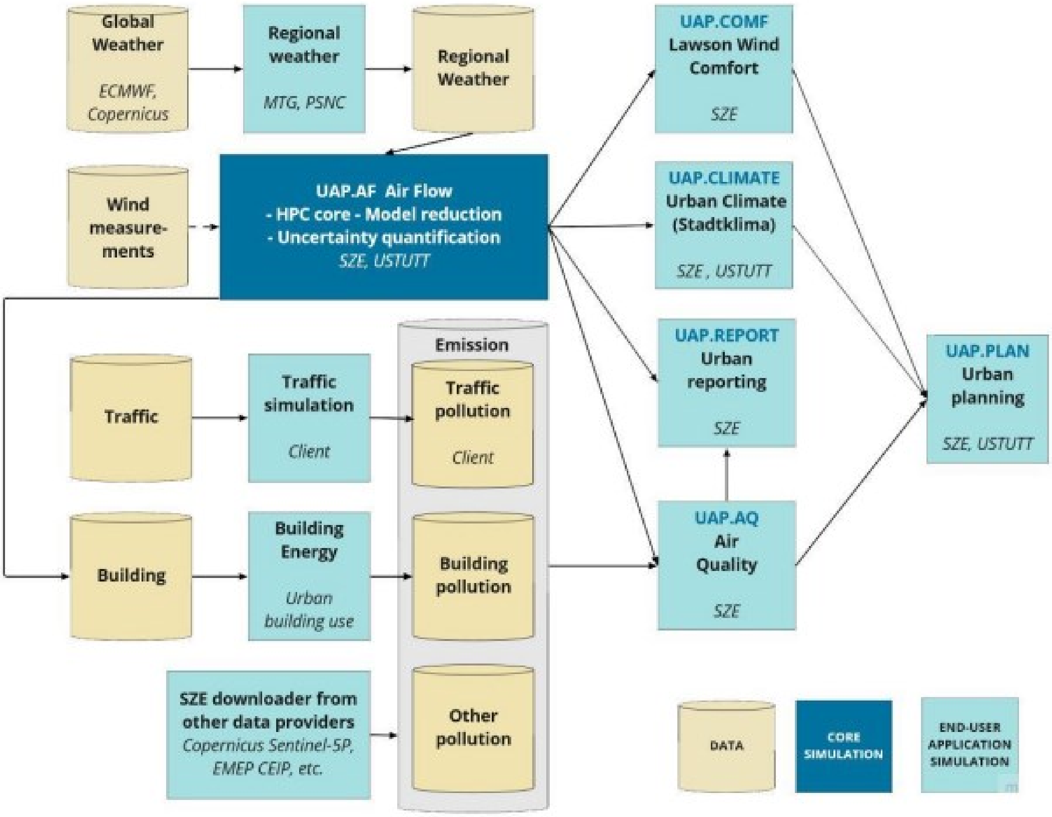

Figure 2. Diagram presents control and data flow for Urban Air Project use case (Grant Agreement, p. 144).

Airflow simulation

As is mentioned in Section 2.1.1, UAP uses OpenFOAM and RedSIM solvers for the airflow computations. From the OpenFOAM framework the pimpleFOAM and the simpleFOAM solvers have been used, for unsteady and steady simulations, respectively. These solvers use a Reynolds-averaging technology with the k-epsilon turbulence modelling, which seems to be a good compromise between accuracy and running time since for their application a relatively coarse mesh tackles the accuracy requirements in most cases.

RedSIM solves the compressible Euler and Navier-Stokes equations on 3D unstructured, conform tetra- or polyhedral meshes. Vijayasundaram's flux-vector splitting method and its second order reconstruction scheme is used for the spatial discretization. The time stepping is solved by the Explicit Euler method, or by a second order total variation diminishing (TVD) optimal explicit scheme. The selected spatial and time discretization methods are implemented for GPUs in a genuine, straightforward and efficient way, which was required by the fact that RedSIM does not use turbulence modelling and thus for an appropriate result relatively fine meshes are needed.

Real-time, or even faster than real-time simulations either need an expensive very large-scale simulation leveraging HPC, or a reduced order method (ROM) that is trained on HPC and maybe ran on cheaper hardware. For RedSIM, the Proper Orthogonal Decomposition (POD) method is the core for ROM. The main idea of the technology is that the HPC-simulation results with relevant parameters are sampled and stored in a snapshot matrix offline. Then a small number of left-singular vectors, r, of the singular value decomposition (SVD) of the snapshot matrix are selected according to the decay of the largest singular values, and then the corresponding *r* left-singular vectors provide an optimal basis of the space of snapshots. Then, each state vector of the time iteration is approximated in the selected basis and only the r coefficients must be computed in each iteration step. Though the degree of freedom is dropped significantly, this method for larger r is very demanding computationally, due to the projections between the original state space and the coefficient space, some interpolation method is often used. In HiDALGO2 the discrete empirical interpolation procedure method, DEIM, has been implemented and used to make the projections computationally effective [7uap]. By using AI methods this part may be significantly accelerated, which research is part of the HiDALGO2 project, for similar approaches see [8uap], [9uap].

4.2 Discussion of Requirements

The most demanding initial requirement of UAP is the simulation of the airflow and air pollution at high resolution in an entire city. Such a domain is of physical size 5 kilometre by 5 kilometre by 0.8 kilometre and at urban ground level 1-2 metre mesh resolution is expected. Though the mesh may be coarsened far from the buildings and high above the buildings, the mesh size is at least millions, or tens of millions of cells. Since for each mesh cell several physical parameters must be computed and stored, several tens of millions degree-of-freedom systems arise. Doing all computations with such a big data is very demanding in particular for development purposes.

Development of the DEIM-part in the model reduction module is also very demanding. In the literature, results for very simple geometry and relatively small number of cells can be found [7uap].

4.3 Status as of M10

From the start of HiDALGO2, the following main results have been achieved. The order of challenges in the previous chapter has been applied here.

1. Validation of UAP airflow simulations.
    a. SZE team participated in FAIRMODE and submitted a manuscript about validation results in EU joint intercomparison exercise; this work is joint with about ten different research groups within the EU who deliver annual air quality reports according to EU regulations; the SZE team was the only team who could report from unsteady simulations for a whole year.
2. Implementation and optimization of the solvers on EuroHPC infrastructures.
    a. Benchmarking is done on EuroHPC for the OpenFOAM, the RedSIM, and another solver xyst[5] under investigation.
    b. RedSIM has been implemented to multi-GPU and OpenMPI+Pthread infrastructures. Detailed benchmarking is ongoing.
3. Mathematical algorithms development
    a. analyse and implementation of low-Mach number flow solvers is ongoing for xyst;
    b. first results for making DEIM consistent, investigation of AI methods for ROM is ongoing,
    c. the feasibility of time-stepping procedures, in particular that of the SVD procedure is partially done, results are communicated at ICIAM2023[6],
    d. UQ is ongoing, first results done with USTUTT,
    e. optimization of the OpenFOAM sub-solvers (e.g., linear algebra) is partially done, with PSNC;
4. Digital twin implementation of UAP:
    a. an early prototype already exists and it is foreseen to be showcased at HiPEAC2024;
5. Coupling UAP and HiDALGO2 use cases:
    a. WRF implementation is ongoing, first implementation is done with MTG.

4.4 Beyond State-of-the-Art, Challenges and Roadmap

During the development of the UAP application, some simple test problems have been created. For example, for the development CFD solvers and their model order reduction, 1D and 2D SOD shock-tube problems, and flow around cylinders have been applied. Also, test problems from literature such as FitzHugh-Nagumo test problem [7uap] have been investigated, operationally for much larger meshes than in the papers in related literature.

The main objective of the project is to be able to support actual urban decision-making processes. This requires simulation of wind comfort and air pollution for entire cities. We identified the following use case: Simulation of air pollution in Stuttgart and in Stockholm.

---

[5] https://xyst.cc/
[6] https://iciam2023.org/

Research Lines and Challenges.

To achieve the main goals of the UAP, i.e., to create, operate, and provide computational solutions for city policy makers for planning and regulations, and inform citizens about the air quality and wind comfort status, the following research challenges must be tackled.

- Validation of UAP airflow simulations. Compare simulation results and measurements in the main use cases and determine the conditions for validation, in terms of mesh resolution and quality, quality of external weather conditions, solver configuration properties, model reduction parameters.
- Implementation and optimization of the solvers on EuroHPC infrastructures to the greatest extent possible. Reach near exascale performance with validated codes.
- Mathematical algorithms development. The utmost optimization of the code requests development of the numerical algorithms, in particular
    - analyse and implement different algorithms for exploiting the physics properties of urban airflow, e.g., low-Mach number flow solvers, Large Eddy Simulations (LESs), artificial diffusion techniques and others;
    - the model reduction methodologies, including development of
        - the state-of-the-art DEIM procedures (to make them consistent),
        - AI solutions for getting rid with DEIM in POD-DEIM,
        - AI solutions for the whole time steppings;
    - develop the time-stepping procedures:
        - find and implement minimally invasive limiters,
        - analysis of time-stepping to provide physically relevant solutions for the transport and the chemistry modules, both for full order and reduced order simulations,
        - implementation of implicit time-stepping;
    - UQ for UAP development,
    - optimization of the OpenFOAM sub-solvers (e.g., linear algebra).

2. Digital twin implementation of UAP:
    - choose appropriate data assimilation methodologies and implement them,
    - interactive visualization of digital twin simulations;
3. Coupling UAP and HiDALGO2 use cases:
    - develop a general coupling mechanism to the UAP framework,
    - develop coupling with WRF,
    - develop coupling to UB and WILDFIRES.
    - Insert the UAP into the scope of the main European programmes:
        - implement interfaces to the Destination Earth solutions (digital twin engine, data lake, etc.);
4. Workflow and service development
    - use HPDA methods to be able to address fast the planning requirements,
    - develop, optimize, apply in production, and document workflows according to user requirements.

Roadmap

Figure 3 shows the main research activities in the UAP Pilot and the main milestones in each of them. Milestones in blue refer to deliveries of partial functionality, deliveries that will make it possible to evaluate the progress of the activity and check the progress in functionality. Milestones in red refer to deliveries of the research lines that had been explained before in this chapter.

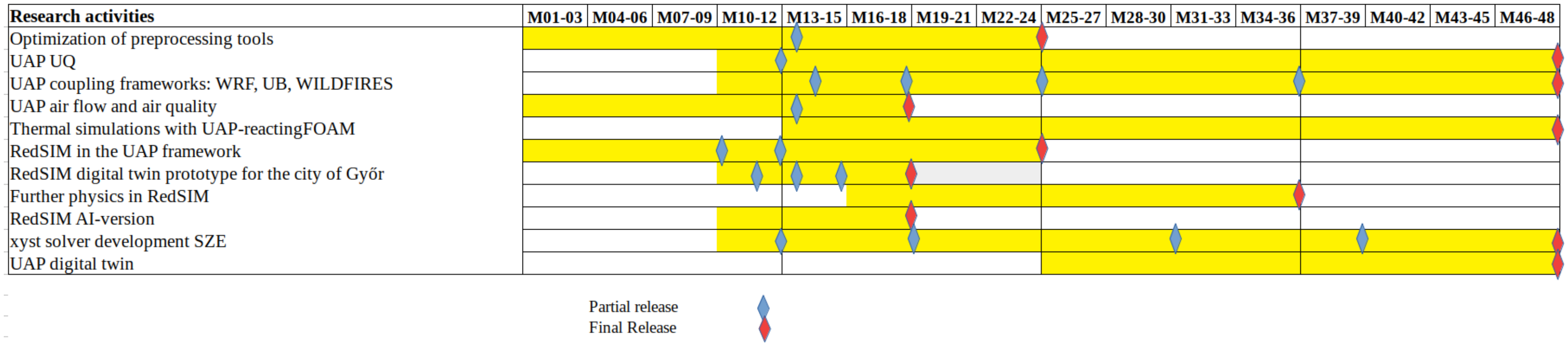

Figure 3. UAP Pilot Roadmap.

5. UB Pilot

5.1 Scientific Baseline

Urban Building Energy Modelling (UBEM) is the domain in which the UB application operates. Much like the UB pilot, the broader UBEM community emphasizes reducing energy consumption and greenhouse gas emissions through simulations at the district/city scales and beyond thanks to exascale computing.

Research groups often develop their own UBEM simulation tools which may use already existing software for parts of the simulation workflow. This leads to studies that use building models with different levels of detail or that may neglect shading effects among buildings.

In [3ub] the authors use the "DIMOSIM" tool to analyse the effects of level of detail in both building and solar shading models to accurately predict the annual heating demand and, at the same time, reduce the simulation time. They show that the choice of mono-zone (or multi-zone with one zone per floor) building models and shading mask data per building facade are parsimonious yet satisfactory choices for thermal simulation at the city scale. In the UB pilot similar choices were made for building and solar models.

In [4ub] the authors present an urban energy modelling approach that is based on the automatic extraction of GIS data and enrichment based on national building stock statistics. In order to avoid impractical runtimes, the importance of choosing an appropriate level of detail for building models is stressed, and the importance of automatic retrieval and completion of information for the building models is also highlighted. These points are of particular importance for the UB application as well, and for this reason, it automates the generation of the geometry and the download of GIS data with building information. In this study, the authors mention that a limitation of their study is the lack of inter-building shading modelling, an aspect that the UB pilot aims to take into consideration.

In [5ub] the authors present the UBEM "CityBES" software, which is developed at the Lawrence Berkeley National Laboratory and is able to retrieve data and simulate any U.S. city. In their application, GIS input data is used, and through parallel simulations of building models (created using EnergyPlus) they can evaluate the impact of certain energy-saving measures on the city's energy demand.

In [6ub] authors also state that solar shading masks are not computed efficiently, and that the shading and transpiration effects of vegetation are not considered. Moreover, they suggest using CFD simulations to evaluate the impact of micro-climates on urban building performance. The UB application aims to cover these problems and explore coupled simulations with CFD models via the interaction with the UAP pilot of the Hidalgo2 project.

In [6ub] the authors focus on the building simulation of two districts in Macau to suggest directions in energy saving and pollution reduction. They also use GIS data to extract building information, and propose two methods to analyse carbon emissions: land use-based method, which does not take into account the shape and height of the buildings and assumes constant energy intensity for each land use type (residential, commercial, etc.) , and the "simulation" method which accounts for additional geometric information such as building height and shape, whose results were more in line with the measurements they could access for comparison.

The articles above mention some of the open research challenges in the UBEM community that the UB application aims to tackle, namely:

- automatic generation of geometry and building models from GIS data;
- inclusion of inter-building and vegetation-induced solar shading information into the thermal models;
- coupling with CFD models for the exploration of micro-climates and their effects in urban energy simulation;
- highly parallel and large-scale urban simulations that go beyond one district.

5.2 Discussion of Requirements

The main general initial requirements of the UB pilot were:

– the access to HPC resources for both benchmarking and simulations, which has been ensured by the available local, PSNC and EuroHPC supercomputers (Lumi, Vega, Karolina);
– the access to real input data and models for benchmarking, which has been ensured by the development of a software - during the first months of the HiDALGO2 project - capable to automatically generate mesh and building models generated from GIS data. This input data can be modified/simplified to generate synthetic inputs for benchmarking;
– the leverage of HPDA methods to efficiently process simulation data, which is under study for the case of shading masks in collaboration with ICCS;
– installation of the pilot's software to HPC clusters, which has been carried out thanks to Apptainer (formerly Singularity) containers for both PSNC and EuroHPC machines;
– Integrate simulation and data analytic tools in a manner that avoids IO (input/output) bottlenecks to enhance computational efficiency and incorporate uncertainty quantification tools into the pilot.

Other initial requirements that are specific to the UB pilot and that are necessary for its execution are:

- the availability of tools for building model generation, which is ensured by the installation of Dymola (Modelica language) on Cemosis local cluster, and model execution, which is ensured by the usage of the FMI standard for the simulation of FMUs; Dymola is used to generate template building models, which are later parametrized using the interface of FMUs;
- the automatic generation of mesh and building models, which is ensured by the platform Ktirio-city that is being developed by Cemosis;
- the development of two families of building models, multi-zone and 3D reduced models, which are ensured by the Modelica and Feel++ libraries, respectively.

5.3 Status as of M10

The following results have been achieved up to month 10 of the project:

- Geometry generation: development of a tool capable of extracting GIS data for any city and reconstructing the building geometry in MSH file format. Two LOD descriptions are available: LOD-0 (bounding boxes) and LOD-1 (polygonal description of the buildings). The resulting meshes are not yet watertight, but they can be used for solar mask computation and visualisation.
- Building models: development of a parametrized multi-zone building model of residential type using the Modelica language. Inclusion of solar shading mask information to model the mutual shading effects of surrounding buildings.
- Solar masks: development of algorithms, data structures and visualisation tools for the computation of solar shading masks at the district and city level. The current version of the code is sequential, and its parallelization is under development. It has been applied successfully on portions of Strasbourg and New York.
- Workflow: automatic generation of geometry and building models. Partitioning of the geometry and parallel simulation of the multi-zone models associated with each building.

5.4 Beyond State-of-the-Art, Challenges and Roadmap

Considering the significant energy consumption attributable to the building sector, the primary objectives of the UB pilot are to estimate and describe these energy exchanges. This insight is intended to equip decision-makers with tools and data to support their interventions. Specifically, the pilot seeks to assess the aggregate energy consumption of buildings at the district or city level. It also aims to provide building-specific energy balances to pinpoint regions or individual structures with elevated energy consumption. The latter may require more information about the buildings and the associated energy consumption through additional data from databases, building owners or measurements. Achieving this requires the UB application to run simulations under varying meteorological conditions and different Levels of Detail (LOD) for building models. The focus of the analysis determines the required LOD – broader city-level studies might necessitate a lower LOD, whereas detailed building-specific analyses demand a higher LOD.

Since the building sector is responsible for a third of the EU greenhouse gas emissions, the UB application aims to provide insight into such production of pollutants, at the city level. The user of the UB pilot will be able to evaluate the emissions of greenhouse gases, based on the building's usage, and identify pollution sources and zones of high concentration. This can be obtained through simulations of the UB application alone, by considering different meteorological conditions and correlating the energy consumption with the energy sources and their pollution levels. Coupling the UB application with the UAP application could give further insight into the transport of pollutants that are generated by buildings, and how the geometry of the city might create zones with different concentrations.

Research lines.

The objectives of the UB pilot require the development of new and efficient tools for the numerical simulation of buildings at the city scale. The first research line focuses on the generation of the city's geometry from GIS data, which will provide a watertight mesh that will be the basis for both building simulation and coupling with the UAP pilot. The second research line is on the building description, for which different LOD have been identified: from bounding-box representation (LOD-0) to polygonal exterior geometry (LOD-1) and eventually faithful geometry from IFC files (LOD-3) for a selection of relevant buildings. Depending on the level of detail, the modelling of physical phenomena comes also with different accuracies, as the internal and external architectural elements are modelled with increasing precision. The third research line is in the efficient computation of solar masks, which model the portion of solar radiation that is shaded by the surrounding constructions. Lastly, the simulation of building energy exhaust will be improved by coupling it with the UAP pilot. This will offer more precise boundary conditions. Additionally, thanks to the integration of temperature and air quality sensors, a data assimilation strategy will enhance the alignment of our models (both UAP and UB) with observed data.

Research Challenges.

The planned objectives require the improvement of several tools that are already under development at UNISTRA. These improvements include:

- the automatic generation of district/city mesh from GIS data, and in particular the generation of a watertight version enabling the exchange of information between our pilot application and the UAP one;

- the automatic generation of building models for all levels of detail, starting from a simplified bounding box description (LOD-0) to the generation of a multi-zone model from an IFC file (LOD-3);
- the specialisation of building models according to their usage (residential, commercial, factory) as retrieved from the GIS data, which includes the estimation of pollutant generation;
- the automatic creation of a network (electrical, heating) that connects buildings and improves the estimation of energy consumption and greenhouse gas emissions;
- the parallelization of the simulation through geometry partitioning and the parallel execution of building computational models;
- the generation of 3D reduced order models for buildings in the unsteady regime;
- the usage of data assimilation algorithms (in part already developed) to include available sensor data in order to estimate parameters of the building models or improve the simulation results.

D5.3 Research Advancements for the Pilots

Roadmap

Figure 4 shows the main research activities in the UB Pilot and the main milestones in each of them. Milestones in blue refer to deliveries of partial functionality, deliveries that will make it possible to evaluate the progress of the activity and check the progress in functionality. Milestones in red refer to deliveries of the research lines that had been explained before in this chapter.

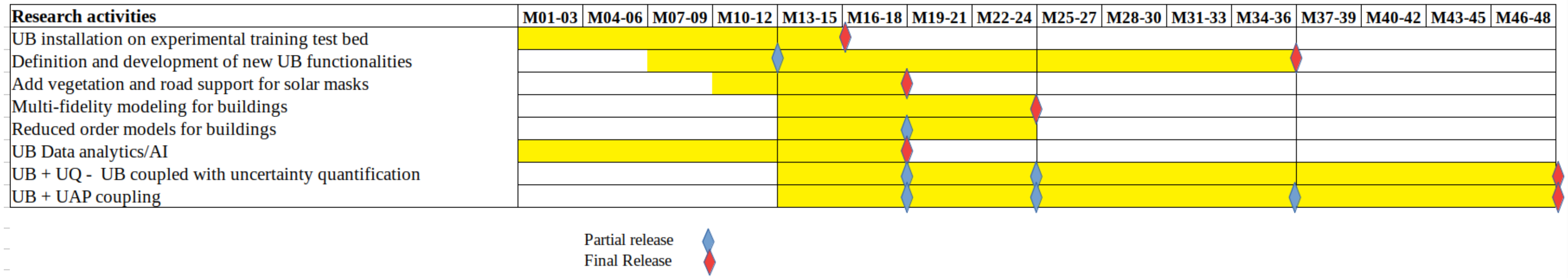

Figure 4. UB Pilot Roadmap.

6. RES Pilot

6.1 Scientific Baseline

RES is a tool for analysing and solving problems related to renewable energy sources, such as estimation of energy production or prediction of the damages to the infrastructure. The aim is to combine weather prediction model - WRF - coupled to all-scale geophysical flow solver - EULAG - with energy prediction module, uncertainty quantification analysis and visualization, as depicted in Figure 5.

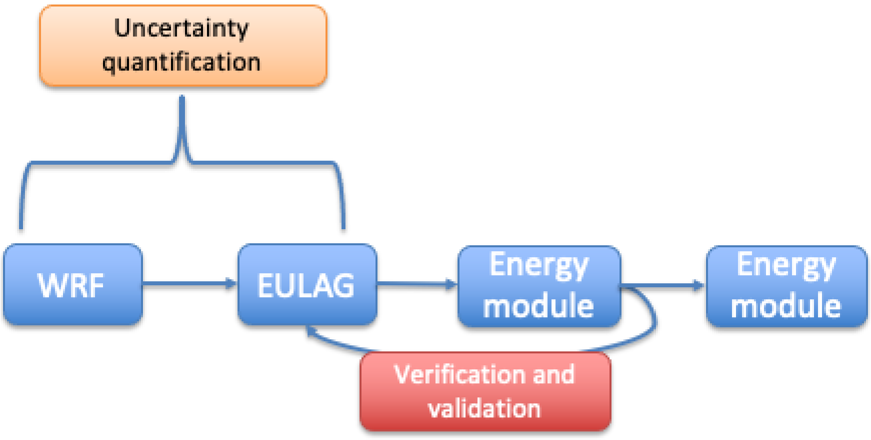

Figure 5. RES main components.

WRF

WRF is the initial model in RES to provide overall mesoscale weather prediction to EULAG. The challenge is to run WRF at larger domain with higher spatial resolution, while preserving good quality of results. Although this challenge is not explicitly solved within RES pilot, it is being solved by Wildfires pilot and the results are directly applicable to the RES.

From the functionality perspective, WRF-Solar will be used to improve sun/shade model and support the energy production from photovoltaic systems. Sun/shade model for cities being delivered by UB pilot is used for solar panels installed in urban areas.

EULAG

EULAG is the core simulation module in RES. From the functionality perspective, the vertical wind profile needs to be enhanced for scenarios dedicated to wind farm modelling. From the performance perspective, runs over large domain with increased spatial resolution need to be conducted to check the quality of results and computational bottlenecks to overcome. Another challenge would be to replace some parts of the solver with AI/DNN techniques to speed up computations.

Energy module

The energy module is responsible for predicting how much energy will be produced by wind farms or photovoltaic system. Currently it is based on some simple historical data, where wind direction and speed at ground level was given and map to the energy produced at that time. The next step is to use synthetic data to use HPDA techniques to process the data and AI to propose a learning model for finding correlation between forecasted weather data and amount of produced energy. Then, real-world data will be used to improve the energy module.

Uncertainty quantification and sensitivity analysis (UQSA)

UQSA has been applied to EULAG models in different scenarios, not considered in the Hidalgo2 project, as a proof-of-concept [1res]. It is based on EasyVVUQ toolkit [2res] (for UQSA) and QCG tools [3res] (for automation in HPC environment). In HiDALGO2, UQSA need to be applied at least to the EULAG model to: i) study the uncertainties related to input parameters and parametrization, ii) decide which uncertainty parameters influences the results the less thus lowering required number of ensembles and the required computational resources, iii) study to what extend increasing spatial or time resolution yields in better simulation results, iv) replace some parts of the EULAG model with surrogates to speed up computations.

Visualization

RES is currently using some Python-based tools and QGIS scripts to visualize the results. The challenge is to adapt visualization tools proposed in the Hidalgo2 project to improve analysis of results to users and customers.

6.2 Discussion of Requirements

Hardware

Increasing the spatial and time resolution of weather prediction and energy productions, as well as running the ensembles, requires vast of computational resources. Although PSNC owns a HPC cluster, ranked in TOP500 list, RES pilot requires access to pre-exascale machines such as these available in EuroHPC initiative. Such large machines are required also for benchmarking the improvements made.

Workflow orchestrator

RES pilot comes with two simulation models, one -way coupled, where integration is done via a dedicated framework. Developments aim at having the models exchangeable (e.g., from other pilots). Similar approach is required for module for energy prediction, pre- and post-processing tools. RES requirement is to have a workflow orchestrator, which allows for loosely coupling of different parts of the solution.

Data management

Each run of RES requires GBs of weather data to be downloaded to initialize the models. This is often cumbersome because downloading takes quite amount of time due to limited bandwidth at source side. To speed up this process, the data should be downloaded in separate process and stored in near-to-computation storage before RES is executed. The data should be well described and there should be tools to efficiently search for it and download it. In this way RES could use similar data but generated with other models or pilots.

Uncertainty quantification

Uncertainty quantification and sensitivity analysis is planned to be applied to RES by the mean of EasyVVUQ toolkit. Requirement is to have this toolkit coupled with HPC orchestrator for automatic ensembles generation, run and analysis.

HPDA/AI

Energy production model will rely on data coming from the sensors and monitors installed at wind farms and photovoltaic systems. To find a correlation between weather situation and amount of energy produced, there is a requirement to have tools to process these sensor data in an efficient manner, as well tools to train the AI model.

6.3 Status as of M10

Coupling WRF and EULAG with the python framework

Since the beginning of the project a key component of Renewable Energy Sources pilot was created - RES software. It is a Python-based framework that orchestrates both WRF and EULAG models into a workflow offering a multiscale approach. While setting up aforementioned models in a standalone manner is a tedious process, RES is designed in a way that allows the user to set up a simulation workflow with a minimal amount of information in a single .SON file, defining such basic parameters as location, domain size, time frame, resolution etc. The framework automatizes all the steps necessary in every study:

1. substantive check of correctness of input data given by the user,
2. connecting to FTP server with global meteorological forecast data, selection of the most suitable result files and downloading the to the working directory,
3. setting up WRF model for the scenario including all the defined levels of domain nesting,
4. creating necessary source files and compilation of EULAG model,
5. coupling WRF output data with EULAG and configuration of EULAG model for each separate domain that was defined by the user,
6. running simulation on:
    a. local machine
    b. HPC site with SLURM queuing system; in such a case, RES manages and controls the status of all the related SLURM jobs sent to the queue

7. checking of completeness of the results,
8. post-processing including generation of pictures, animations and CSV files with probe data

RES allows to adjust its configuration for compatibility with various HPC environments, including the resources of Euro HPC sites. The code is also covered with unit tests to ensure its reliability. The final module including energy production estimation is currently under development.

Damages case study

This year PSNC started a collaboration with one of the largest Polish DSO related to predictions of damages of the overhead network due to extreme weather conditions in one of the major Polish cities. The performed study, which gave an opportunity to utilize RES framework, was focused on the effects of excessive wind gusts on the overhead infrastructure. In the workflow there were used three levels of domains nesting, where the outer domain of the scale of the whole country had a resolution of 3.6km while the final and most fine-grained domain was discretized with the spacing of just 100 meters. In the study a detailed description of the location of all the analysed infrastructure given by the DSO was used as part of the input data in the workflow. The final resolution was fine enough to provide meteorological simulations results individually for each component of the infrastructure and asses the likelihood of damage. An exemplary visualization of the study is presented in Figure 6.

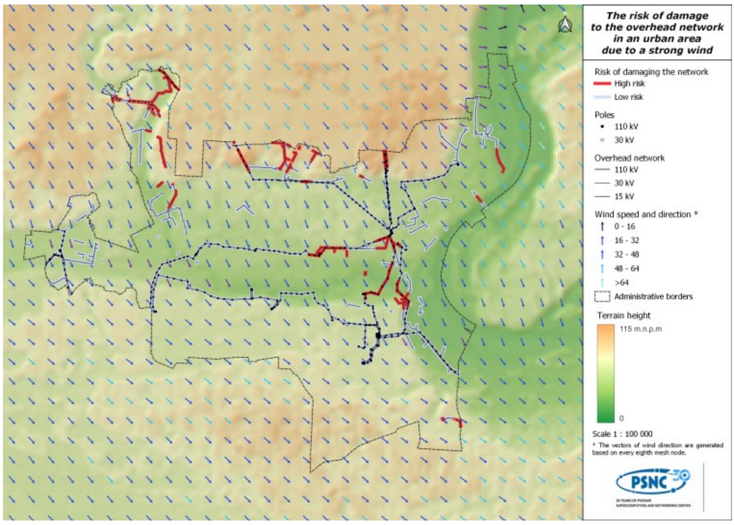

Figure 6. Overhead network damage prediction based on weather conditions.

All the simulations are executed daily on the Altair supercomputer owned by PSNC. This scenario is scalability prepared for LUMI for benchmarking and further analysis. For the scenario of damage likelihood analysis in RES, an additional module including QGIS-based visualization is currently under development.

6.4 Beyond State-of-the-Art, Challenges and Roadmap

With the increasing number of renewable energy sources installations, such as photovoltaic and wind farms, is it important for DSOs, but also for individuals, to understand the physical phenomena and maximise the outcome. In RES focus is given not on modelling the interaction between wind turbines or how to plan steer the farm in near-to-real-time manner. Rather it is given on understanding how weather conditions influences the energy production and provide more detailed prediction of both. AI and HPDA tools can be used to find the correlation between weather conditions and amount of energy produced. Because of the many variables and uncertainties, applying uncertainty quantification can improve results even more. Particular focus will be given for models related to solar energy: WRF-solar is considered to be used instead of WRF, and UB pilot solar-shade model could be used for urban environment.

The Renewable Energy Sources (RES) pilot is driven, among others, by the requirements of one of the largest DSO in Poland. Three specific challenges are tackled:

- Energy production from the wind farms to estimate how much energy a wind farm can produce taking into account its infrastructure, localization and forecasted weather conditions (wind direction and speed in particular).
- Energy production from photovoltaic system to estimate how much energy photovoltaic system can produce taking into accounts its infrastructure and forecasted weather conditions (sun angle and cloud cover in particular).
- Damages prediction to the DSO infrastructure to predict potential damages of the overhead network due to icing, excessive heat, or wind gusts. This scenario can be extended to photovoltaic systems to estimate damage probability due to hailstorm.

Research lines.

Following new functionalities are foreseen at this stage of the project:

Wind and solar energy module, aimed in finding a correlation between forecasted weather and energy produced based on data from RES farms on-site sensors, where HPDA/AI techniques will be used. PSNC is finalizing building of the large photovoltaic system (almost 1MWp), and the data will be available at hand.

Prediction of damages to the DSO infrastructure; subsequent enhancements to improve the probability of damages by comparing with actual damages, and including new hazards such as icing or excessive heat.

Visualization. Use proprietary and in-project post-processing solutions to demonstrate product outcomes for scientific communities and customers.

Uncertainty quantification. Limit uncertainties and improve quality of results. Sensitivity analysis is to limit the number of ensembles while preserving the given degree of quality results.

Contenerized solution. Be able to run RES on different EuroHPC sites without a hassle, and in cloud environment upon stakeholders' request.

The HPC systems are primary target to run RES in high-resolution for more accurate results. On the other hand, there is a money factor taken into account by potential customers. The goal is to provide a cloud solution, not only for the sake of compatibility with many systems, but most of all to make the product affordable at expense of less accurate predictions.

New coupling between WRF and EULAG; introduction of QCG and use of the workflows to run each pre-, post-processing tools and different models in automatic fashion.

Sensitivity analysis: Identify which ensemble parameters can be omitted resulting in less ensembles and less computational time required.

Integration with Destination Earth initiative by using provided products and API

Coupling with UB pilot; possibility of use building and solar shading data.

Research Challenges.

Uncertainty quantification

Uncertainty quantification gives opportunity to deal with uncertainties which occur at different stages of the workflow. The uncertainties come with input data, coupling methods between the models, parametrization of the model or its numerical schemes. RES pilot aims at reaching 80,000 CPU cores with ensembles. The utmost importance is to have use UQ tool which can process the ensembles in an efficient manner. Previously EasyVVUQ toolkit was used for uncertainties purposes, and QCG for efficient and seamless ensembles execution in HPC environment, and it will be applied to RES.

Applying UQ and SA will help in:

- Determining which input data and/or model parameters can be omitted to lower the number of ensembles needed;
- Improving the results by lowering the uncertainties;
- Analysing if and to what extent domain size and mesh resolution impacts the results.

Scalability and efficient processing

RES use case aims at reaching 50k CPU cores in a single run. WRF is subjected to Wildfires pilot, therefore a focus will be given to EULAG model and energy module. There are several steps and tools to be used in order to reach that objective. First, bottlenecks have to be identified to steer the further developments towards increase of the scalability. The plan is to test different domains sizes, mesh resolutions and time steps and seek for the scalability limits. Once identified, profiling tools will help to identify the bottlenecks and suggest some optimization strategies. A possible improvement is to increase the efficiency within a single node.

Another improvement may come with uncertainty quantification studies. Figure 7 presents how UQ is going to be applied to RES. While the primary target is EULAG, applying UQ to overall input data or WRF model is not excluded at the moment. Running UQ analysis may propose a surrogate model to replace the simulation model with a less expensive one, and yet more efficient. It is important to study what is the price to be paid in terms of quality of results, and to find the best trade-off between these two factors.

Yet another increase in scalability may come from applying AI/DNN to GCRK solver of the EULAG model, as presented in Figure 7. The GCRK is an iterative solver, and the idea is to replace some iterations with AI/DNN in the following manner: 1) solver initiate computations, 2) agent decides whether to interrupt solver, 3) AI/DNN takes over, 4) upon agent decision, iterations continue to be solved by solver, 5) iterations continues with AI/DNN and solver until stop condition occurs. It is expected that by following this direction quality of results will be preserved, while increasing the efficiency.

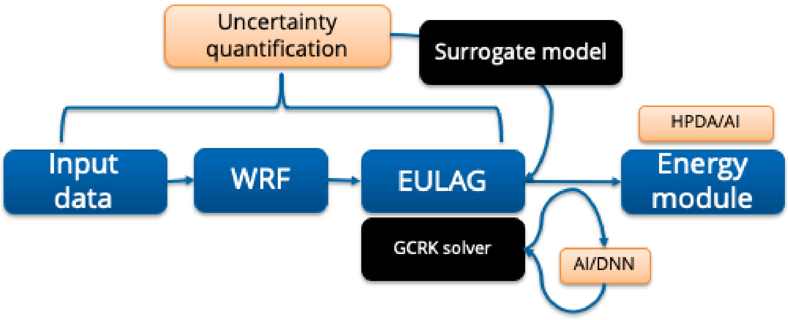

Figure 7. Uncertainty Quantification and AI for RES.

D5.3 Research Advancements for the Pilots

Roadmap

Figure 8 shows the main research activities in the RES Pilot and the main milestones in each of them. Milestones in blue refer to deliveries of partial functionality, deliveries that will make it possible to evaluate the progress of the activity and check the progress in functionality. Milestones in red refer to deliveries of the research lines that had been explained before in this chapter.

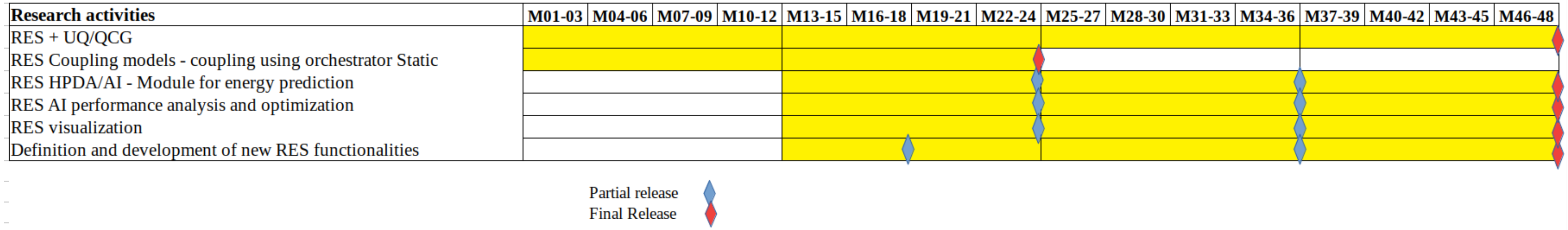

Figure 8. RES Pilot Roadmap.

7. Wildfires Pilot

7.1 Scientific Baseline

Use of mesoscale atmospheric models for fire spread simulation

WRF

The Weather Research and Forecasting (WRF) model is a versatile and widely utilized tool in the fields of atmospheric science and meteorology.

Dynamical Cores

The WRF model operates with two primary dynamical cores: the Advanced Research WRF (ARW) and the Nonhydrostatic Mesoscale Model (NMM). These cores are Eulerian mass dynamical cores, underpinned by components such as advection, pressure gradients, Coriolis forces, buoyancy, filters, diffusion, and time-stepping. However, it's essential to mention that the chemistry component functions exclusively with the ARW core.

Pre-processing System

The WRF Pre-processing System comprises real-data interpolation for Numerical Weather Prediction (NWP) runs (WPS) and programs for augmenting analysis with additional observations (obsgrid). The WRF Model, encompassing ARW and NMM dynamical cores, involves initialization programs for real and idealized data (real.exe/ideal.exe) and a numerical integration program (wrf.exe). In addition to these elements, the WRF Pre-processing System extends to WRF Data Assimilation (WRFDA), WRF-Chem, and WRF-Fire as specialized modules.

Multi-Domain Nesting

The WRF model supports nested domain configurations, allowing for runs with varying resolutions in nested areas. Parent domains receive specified boundary conditions from the wrf-boundary file, while nested boundary conditions are derived from the parent domain.

WRF Data Assimilation (WRFDA)

WRFDA offers data assimilation techniques including variational data assimilation (3D-Var and 4D-Var), ensemble data assimilation, and hybrid variational/ensemble data assimilation. Its primary function is to assimilate observations to enhance WRF input analysis generated by the WRF Pre-processing System.

Simulation of wildfire-atmosphere interactions

WRF-SFIRE

The WRF-SFIRE-CHEM model represents a cutting-edge, integrated fire-atmosphere model infused with chemical processes. This model interconnects the Weather Research and

Forecast model (WRF-CHEM) from the National Centre of Atmospheric Research (NCAR) with a fire propagation model known as SFIRE. WRF-SFIRE, an open-source community model accessible[7], employs a two-dimensional semi-empirical rate-of-spread model to parameterize the progression of fires. Numerical weather prediction (NWP) models are highly valuable for prediction and projection due to their scaled physics, dynamic parametrization, and data assimilation capabilities. To seamlessly incorporate wildfire modelling into an operational weather model, a physics option is simply added to the Weather Research Forecast (WRF) model, as demonstrated by Patton and Coen [2wf], Mandel et al. [3wf], and Coen et al. [4wf].

Several advancements have been made to enhance the performance of WRF-FIRE, including corrections to near-surface winds (Coen et al. [4wf]) and the introduction of a fuel moisture model in SFIRE (Kochanski et al. [5wf]), which runs independently at each grid point. This model estimates drying and wetting fuel equilibrium moisture content and is coupled with WRF-Chem to track fire emissions. Mandel et al. [6wf] exemplified the operational deployment of WRF-FIRE in Israel.

Parametrisation / improvement of the fire simulation module - WRF-SFIRE-CHEM

Installation, parametrisation, and benchmarking of the coupled WRF-SFIRE-CHEM models for use in forest fire simulation environments and their interaction with the lower atmosphere. An exhaustive analysis of the parametrisations that best model the behaviour of the lower atmosphere at very high spatial resolution will be carried out, testing the behaviour of the WRF-SFIRE coupled model with real 3D data in a highly scalable HPC environment.

WFDS

The Wildland-Urban Interface Fire Dynamics Simulator (WFDS) is an extension of the Fire Dynamics Simulator (FDS), originally developed at the U.S. National Institute of Standards and Technology (NIST). WFDS is a fully three-dimensional, physics-based, semi-coupled fire-atmosphere model that leverages approximations to govern the equations of fluid dynamics, combustion, and the thermal degradation of solid fuel (Mell et al. [7wf]). FDS operates as a large-eddy simulation (LES) model, and its extension, WFDS, incorporates a low Mach number approximation to eliminate acoustic wave propagation from the dynamics analytically. It boasts a fast and direct pressure solver, significantly reducing computational time.

The LES box model incorporates a Eulerian framework (level set) to simulate the fire front. Comparisons between WFDS and FIRETEC, using Lagrangian fire spread models, have demonstrated similar results in crown fire spread within the range of a comprehensive empirical crown fire dataset (Hoffman et al. 2016 [8wf]). Nonetheless, such comparisons have limitations due to the scarcity of detailed atmosphere and fuel data. This data limitation

[7] https://wiki.openwfm.org/

affects not only the initialization of simulations but also the evaluation of the processes involved.

Simulation of wildfire smoke plume emission and dispersion

The following smoke-related processes must be considered in the modelling, each one leading to a particular model: fire growth or burned area; smoke emissions, the buoyant rise plume rise driven by the fire, mixing between smoke plume and the ambient air outside of it, deposition processes, downwind smoke dispersion, and plume chemistry.

WFDS utilize a physics-based methodology to predict fire growth and the extent of the burned area. In contrast, WRF-SFIRE and WRF-FIRE adopt an empirical-based parametrization for forecasting fire growth, fuel consumption, fire heat fluxes, and smoke. An alternative approach to predict fire spread involves the use of empirical-based parametrizations, with the Rothermel surface fire spread model being the most employed method.

WRF-Chem

WRF-Chem, supported by NOAA/ESRL/GSD [9wf], encompasses chemistry species and related processes, offering numerous chemistry options. Although included in the WRF tar file, it requires a specialized compilation option. Users are expected to provide emissions data to leverage WRF-Chem's capabilities effectively.

Advanced visualization techniques for training and education

Use of Virtual Reality in fire-fighting training

The use of virtual reality (VR) and extended reality (XR) technologies for forest fire management is an emerging field with significant potential. These technologies offered innovative solutions for training, simulation, and decision support in the context of wildfire prevention, response, and mitigation.

Some examples of application of VR, XR and AR are:

- Firefighter Training Simulations. One such example is "Firefighter Training VR" by Chronos VR. Firefighter VR [10wf]
- Smoke Inhalation Training. Recent research in this area includes the work by Xu et al. (2014) in their paper "A virtual reality-based fire training simulator with smoke hazard assessment capacity" [11wf].
- Wildfire Management Training. VR offers training scenarios for managing wildfires, including strategies for containment, evacuation, and coordination among first responders.
- Team Coordination and Communication. "VRFire" [13wf] by the University of Maryland is an example of a VR platform for team training.
- Hazardous Material Handling. Recent work explores this in "Hazmat Virtual VR First Responder Training" [12wf].

Use of Unreal Engine for scientific data visualization, architecture and engineering

Unreal Engine, developed by Epic Games, stands as the paramount real-time 3D creation tool renowned for its openness and advancement. Evolving far beyond its origins as a cutting-edge game engine, it now empowers creators across various sectors to craft pioneering content, interactive journeys, and immersive virtual realms with unparalleled freedom and control.

Volumetric smoke

Current techniques for simulating volumetric smoke using ray marching are a significant advancement in creating realistic virtual environments, particularly in the context of virtual reality (VR). Ray marching is a rendering method that is well-suited for rendering complex volumetric effects, such as smoke, and it has various applications in VR environments.

To simulate smoke, modern algorithms utilize computational fluid dynamics (CFD) or other methods to model the behaviour of smoke particles and their interaction with the environment. Smoke columns exhibit complex, dynamic behaviour, resulting in data sets that are often characterized by a significant number of empty or sparsely populated cells. In the latest version of Unreal Engine (v5.3), native code is included to add functionality for the management of OpenVDB files, a format designed for the storage of volumetric data in a sparse grid that makes it an ideal choice for simulations that model smoke columns.

7.2 Discussion of Requirements

Hardware and Software requirements.

Provide a computational capacity for the simulation of wildfire-atmosphere interactions and smoke dispersion at several scales aimed at the assessment of risk and potential impacts over populated areas

Integrated models and results in visually oriented platforms for analysis, training and education to advance in the understanding and characterisation between fire and atmosphere.

Coupling weather data and models at different scales, such as mesoscale forecast models (WRF), numerical fluido-dynamic models of air movement (Open Foam) and smoke emission and dispersion models, with fire spread and fire-atmosphere interaction thermo-fluid-dynamic models (FDS).

Boundary conditions derived either from real world measurements or from couplings with external simulation data (wind, temperature, stability, …).

Technical documentation of the use cases must be available. This includes, among other things, the specification of available data and its format and processing methods.

The Wildfires models will be adapted, improved and developed as microservices for their integration as modules in coupled simulations in interoperable HPC environments and then interpreted visually in immersive and interactive spaces.

Adaptation of specific fire spread models and integration with WRF for improvements of fire front forecast and heat transfer computation.

Adapted FDS algorithms for the propagation of fire through non-uniform, built WUI areas (as follow-up of WFDS and current works of Graziani, Mell et Al. [14wf])

Data requirements

The following data are used to derive Wildfires simulations: high-resolution mdt, lidar data, land cover, forest fuels, and photogrammetry imagery, ECMWF and GFS forecast data, Meteorological observations.

Benchmarking

Prepare a historical dataset of known fires for benchmarks, Large-scale datasets for benchmarks, software which might attract broader GSS and data sciences audience.

Uncertainty Quantification

The study on uncertainty will validate which parameters of the model are crucial for the end results, and which have neglectable impact. Uncertainty quantification and sensitivity analysis is planned to be applied to Wildfires by the mean of EasyVVUQ toolkit.

Visualization

Visualize simulation results for validation, also for workflow demo.

Algorithms for the visual simulation of animated volumetric smoke, i.e., using compression techniques of 3D textures and packing them into MP4 and other video formats. Integration of raymarching methods for the visualization of smoke-light interactions (i.e., Mie and Rayleigh scattering).

Visualize 3D spatial variables, both scalar and vector valued functions in a moving time frame (with videos), cross sections or in innovative ways efficiently, be able to visualize from each module of the workflow.

7.3 Status as of M10

Fire-atmosphere interaction simulations

- Design of WFR/SFIRE simulation scenarios, with up to four nested domains and the use of WRF/LES/SFIRE coupled models for real fires simulation. This has been installed and tested in local MPI machines, Eagle HPC and EuroHPC servers Vega and Lumi.
- Design of a benchmark scenario and test environment setup, reframe with WRF-SFIRE has been installed and tested on local infrastructure Eagle and EuroHPC server Vega. First benchmark tests are scheduled in the first fortnight of November.
- Description of real wildfire/weather scenarios for the simulation in WRF/SFIRE

- Description of the geometry for the detailed simulation with OpenFOAM/fireFoam in 3x3 Km scenarios.

Visualization

- Integrating Unreal Engine 5.2 with GIS Data Sources (SHP) for Immersive Visualization. This integration empowers us to visualize diverse geospatial data layers, including polylines, polygons, and point data, within Virtual Reality (VR) environments.
- Utilizing Raster Layers for Dynamic Fire Spread Visualization. In our pursuit of more comprehensive visualizations, we suggest coupling multiple raster layers to depict the spread of fires in VR environments. This multifaceted approach allows us to create detailed, real-time representations of fire dynamics within immersive settings, offering valuable insights for various applications.
- Leveraging Vector Fields for Enhanced Realism. To implement vector fields as arrow fields in VR environments, each object featuring independent functionalities. This implementation enhances the overall realism of our immersive experiences by providing dynamic representations of vector-based data, making complex phenomena easier to grasp and analyse.
- Utilizing Vector Fields for Tracing Particle Movements. To incorporate vector fields to control the movement of tracing particles, a novel approach designed to visualize volumetric datasets in VR environments.
- Visualizing Volumetric Scalar Fields with Ray Marching. A notable aspect of our endeavour is involving the implementation of volumetric scalar fields. These fields enable the visualization of volumetric smoke and other atmospheric elements using advanced Ray Marching techniques.

7.4 Beyond State-of-the-Art, Challenges and Roadmap

Within the HiDALGO2 project, our aim extends beyond the current state of the art and conventional techniques in forest fire simulation at two scales, and to estimate the potential consequences. This dual approach signifies a twofold commitment: firstly, the development of bespoke solutions tailored to enable superior, more efficient simulations in HPC environments. Secondly, a concerted effort to enhance the methods and procedures used in practical scenarios, with a special focus on operational perspectives.

High-Resolution Weather Prediction

Particular attention will be dedicated to designing and implementing new parametrizations for the WRF model, as well as the simulation modules for forest fires (WRF-SFIRE) and smoke column emission and dispersion (WRF-CHEM), all in response to forest fires. Furthermore, the coupled utilization of computational fluid dynamics numerical models like OpenFOAM and its fire propagation module, fireFOAM, will provide us with high-resolution wind vector fields at the urbanization scale. Currently, this approach has not been applied to an urbanization scale and has only been used for experimental purposes with idealized building envelopes.

Forest Fire Simulation Enhancement

Our goal is to enhance the current forest fire propagation procedures (WRF-SFIRE) for integration into fire-atmosphere interaction models to achieve a more precise representation of fire spread across the tree canopy (crown fires), thereby incorporating the energy production resulting from combustion. In addition, we aim to enhance and further advance the simulation of wildfire progression within urbanized areas. This involves the integration of thermo-fluid dynamic models (such as OpenFoam/fireFoam), along with a detailed description of the existing buildings and vegetation.

Fire-atmosphere interaction

Our aim is to transcend the current state of development of fire-atmosphere interaction models (WRF-SFIRE), making them fully operational for use in landscape sensitivity analysis and comprehensive territorial preventive planning. To achieve this, it's imperative to meticulously define the parametrization of mesoscale meteorological models as well as the models for wildfire spread. We will pay close attention to energy emissions from the forest fire and their impact on the vertical wind component, as well as the formation of pyro-cumulus clouds.

Smoke production and dispersion

An additional breakthrough in the project HiDALGO2 is the integration of smoke components into the simulations, both at the landscape and urbanization scales. Smoke is a crucial element that has demanded extensive research efforts, and smoke modelling is still in a developmental stage. In our project, we aim to incorporate and enhance specific smoke emission and dispersion models (WFR-CHEM) to assess their influence on landscapes and urban communities.

Visualization

A significant aspect of implementing the outcomes of Project HiDALGO2, particularly in the context of the forest fire case study, involves the interpretation and visualization of results with realistic rendering. This approach is justified for their application in both, training for intervention teams and educational activities for public risk awareness raising about the factors, processes, and consequences of forest fires in urbanized areas. To accomplish this, an innovative approach is proposed, by building bridges between the outcomes of numerical simulations and immersive visualization platforms, particularly game development platforms such as Unreal Engine. The integration of these techniques aims to enhance the accuracy and realism of simulating fire behaviour and smoke, and their interaction with the environment in immersive settings.

Figure 9. An exploratory example of a realistic smoke plume visual simulation of an incipient forest fire near an urban area implemented in a VR environment, using ray marching techniques and OpenVDB sparse data management. Credits: David Caballero, MeteoGrid. Developed within the scope and framework of WUICOM-BCN project.

Research lines.

The following lines of research have been identified in the case of forest fires:

- Optimization of high-resolution WRFs in HPC environments as products and services for mesoscale and microscale weather prediction.
- Simulation of fires in HPC environments with operational scope.
- Improvement of the current WRF-SFIRE surface wildfire spread model.
- Simulation of forest fire assemblages in HPC environments for the study of landscape sensitivity oriented to preventive planning.
- Coupling of case studies and related models in HPC environments.
- Use of AI techniques and algorithms for searching and obtaining pre-calculated simulations of forest fires according to similarity analysis.
- Simulation of forest fire propagation in HPC environments and movement at the urbanization scale using CFD models (OpenFOAM/firefoam) for risk assessment and proposal of preventive measures.
- Visualization of scalar and vector volumetric fields in immersive environments at both landscape and urbanization scales.

Research Challenges.

The following challenges are identified in the wildfire case study:

Parametrization and Enhancement of the Fire Simulation Module - WRF-SFIRE-CHEM.

The installation, parametrization, and benchmarking of the integrated WRF-SFIRE-CHEM models are crucial for their effective utilization in forest fire simulation contexts and their dynamic interaction with the lower atmosphere.

Replacement and Enhancement of the Current WRF-SFIRE Surface Wildfire Propagation Model.

The current WRF-SFIRE surface wildfire propagation model, based on the Level Set method, will be replaced, and improved with a model developed in-house.

OpenFOAM Application.

The installation, parametrization, and wind simulation at a very local scale (urbanization) will be performed while considering the geometry of urban environments, including both solid geometric elements (buildings) and porous materials (trees and other vegetation). These simulations will be executed within exascale HPC environments. It's worth noting that while the OpenFOAM/fireFoam solution has seen widespread use, this is the first application of this approach for fire simulations in built-up areas.

Application of OpenFOAM and fireFoam for Thermo-Fluid-Dynamics Simulations.

The OpenFoam and its FireFoam extension will be applied to simulate the thermo-fluid-dynamics of wildland fire spread in wildland-urban interface (WUI) zones near dwellings. This approach will provide insights into fire behaviour in proximity to residential areas.

Simulations in urbanized areas.

fireFoam will be employed to simulate fire spread within housing estates, considering the influence of local heat emitters (such as combustion flames) on buildings and local smoke production.

Bridging of UE with GIS platforms.

To ensure the visualization of geographical items (points, lines, polygons) and the associated information in UE environments a development and integration of specific pieces of software for the bridging with GIS platforms through common data formats, and the use and improvement of existing plug-ins (ArcGIS, CESIUM) for the seamless integration of GIS servers.

Advanced Visualization Techniques.

- Methods and algorithms will be developed to visualize scalar fields in Unreal Engine, with a specific focus on representing forest fire volumetric smoke plumes.
- In addition, mixed Eulerian-Lagrangian models will be created for visualizing vector volumetric field tracers, particularly wind patterns, at both landscape and urbanization scales.
- Various components of wildland fire scenarios, spanning landscape and urbanization scales, will be integrated into Unreal Engine.

Roadmap

Figure 10 shows the main research activities in the WF Pilot and the main milestones in each of them. Milestones in blue refer to deliveries of partial functionality, deliveries that will make it possible to evaluate the progress of the activity and check the progress in functionality. Milestones in red refer to deliveries of the research lines that had been explained before in this chapter.

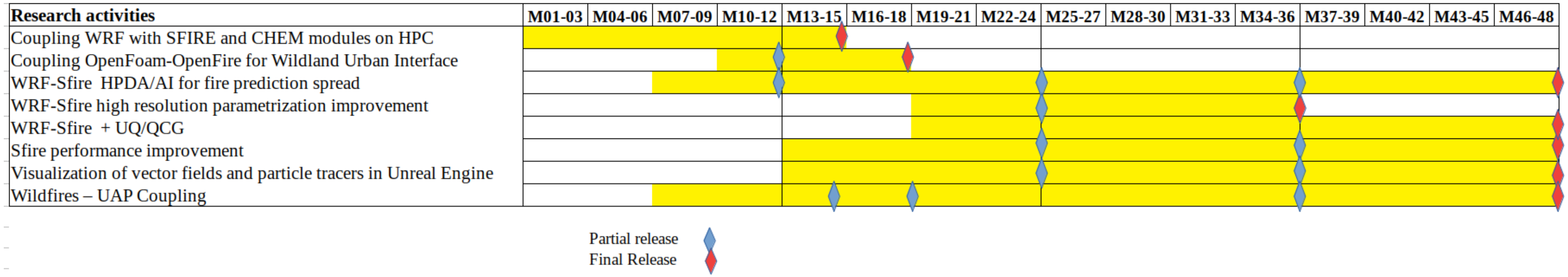

Figure 10. WF Pilot Roadmap

8. Conclusions

This deliverable serves as a cornerstone, laying a comprehensive foundation for the scientific aspects of the project. It establishes the baseline for the chosen use cases and highlights the associated **research challenges**, setting the stage for a journey toward scientific excellence.

The integration of four **use cases**, namely the Urban Air Project (UAP), Urban Buildings (UB), Renewable Energy Sources (RES), and Wildfires (WF), into thematic pilot studies is instrumental in achieving the project's set objectives. Each use case will undergo the creation of at least one scenario, providing a solid foundation for future enhancements and research outcome implementations.

Collaboration is at the heart of project's endeavour and is essential for expanding its research horizons. Scientists from various fields, including CFD modelling, meteorology, air quality, energy building efficiency, renewable energies, and protection from fire or water management, are actively seeking new opportunities and support from specialists in the field of HPC, data analytics, AI, and advanced workflow orchestration.

This deliverable outlines a **strategic roadmap** for the execution of research activities, a design for a research coordination plan, and the delineation of measurable outcome indicators for monitoring and evaluation, ensuring that that work remains focused and goal oriented.

The deliverable also provides an overview of the **initial research progress** in each pilot use case, offering insights into the ongoing work within the project.

The document also provides a summary of the research development status as of Month 10 (M10) for the various ongoing research lines, demonstrating the substantial progress made towards achieving the proposed goals.

Furthermore, this document serves as a **reference point** for the proposed research activities throughout the project's life cycle, providing a consistent guideline, a timeframe, and a thematic rationale to set the scope of planned research and the expected outcomes.

In the pursuit of scientific excellence, possible **bottlenecks** in software are acknowledged, and specify details regarding data transfer, reflecting on software and hardware requirements as defined in D2.1. **Scenarios** that will serve as a foundation for future developments and code improvements are also identified.

The deliverable outlines the **methodology** to be followed, encompassing the following key elements:

- Research coordination
- Benchmarking methodology and optimization
- High-Performance data analytics and AI
- Uncertainty quantification and ensemble methods

- Adoption of new technologies

To facilitate the monitoring of the achievement of proposed results, a series of **measures** and related areas of research have been defined. These are summarized in two tables for comparison:

- Table 1: **Expected results as stated in the Grant Agreement**, including use cases, key framework, key algorithms or models, current state in the project, objectives in HiDALGO2, and means to improve scalability.
- Table 2: **Expected research results**, including use cases, key framework, research goals, current status as of M10, and foreseen date of achievement.

More than **20 new research goals**[8] have been identified, representing a significant advancement beyond the current state of the art and science in various thematic areas developed within the project.

A comprehensive description of scientific research and technological development in each of the **pilots** is provided, encompassing:

- A brief summary of the **scientific baseline** (State of the art).
- A brief discussion of **requirements**.
- The **status of research** as of M10 in the project.
- A brief description of the foreseen developments **beyond state-of-the-art**, the research challenges, and, finally, a proposed roadmap to achieve the expected results.

Furthermore, research aspects in other secondary domains of interest are explored, specifically focusing on visualization and orchestration solutions.

Finally, the deliverable emphasizes the significance of sustaining and enhancing development through thematic **working groups**, particularly focusing on the following themes: WRF, coupling urban simulations, coupling forest fires and air pollution, and visualization. This approach facilitates coherent and consistent research activities while preventing duplication of efforts and fostering synergy among researchers.

Following figures, Figure 11 and Figure 12, show the main research activities in the HiDALGO2 Pilots and the main milestones in each of them.

[8] see Table 2. Expected Research Results, in section 3.3.

Figure 11. Main research activities and main milestones in the HiDALGO2 UAP (top) and UB (bottom) pilots. Milestones in blue refer to deliveries of partial functionality, deliveries that will make it possible to evaluate the progress of the activity and check the progress in functionality. Milestones in red refer to deliveries of the research lines that had been explained before in the corresponding chapter of this document.

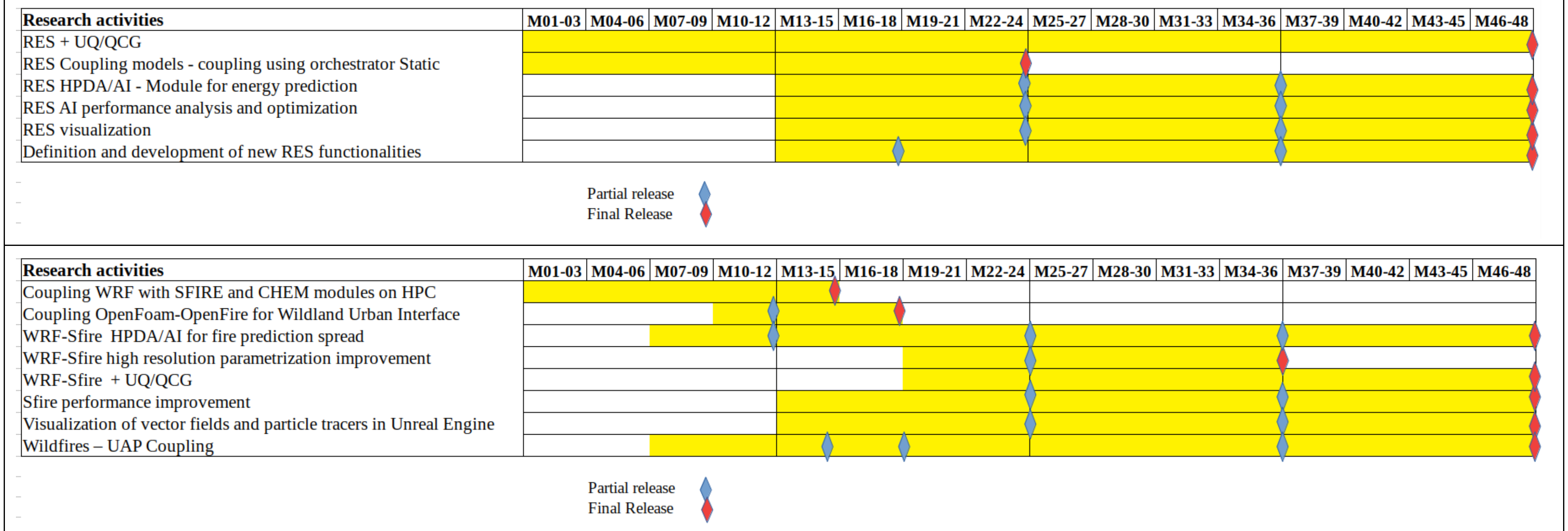

Figure 12. Main research activities and main milestones in the HiDALGO2 RES (top) and WF (bottom) pilots. Milestones in blue refer to deliveries of partial functionality, deliveries that will make it possible to evaluate the progress of the activity and check the progress in functionality. Milestones in red refer to deliveries of the research lines that had been explained before in the corresponding chapter of this document.

References

[1uap] (2023) 68% of the world population projected to live in urban areas by 2050, says UN https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html, [retrieved 2023-11-16].

[2uap] (2023) Air Pollution Data Portal, https://www.who.int/data/gho/data/themes/air-pollution, [retrieved 2023-11-16].

[3uap] (2023) Wind Assessment for Urban Planning and Architecture, https://rheologic.net/articles/urban-wind-assessment/, [retrieved 2023-11-16].

[4uap] Xiaoyue Xu, Zhi Gao, Mingjie Zhang, A review of simplified numerical approaches for fast urban airflow simulation, Building and Environment, Volume 234, 2023.

[5uap] J. Bakosi, Open-source complex-geometry 3D fluid dynamics for applications with unpredictable heterogeneous dynamic high-performance-computing loads, Computer Methods in Applied Mechanics and Engineering, Volume 418, Part B, 2024. https://doi.org/10.1016/j.cma.2023.116586.

[6uap] M. Lawenda et al., Innovative HPC Trends and the HiDALGO Benchmarks - revised version. 2021. Researchgate publication [link], DOI: 10.13140/RG.2.2.14687.79527.

[7uap] Saifon Chaturantabut & Danny C. Sorensen sorensen@rice.edu (2011) Application of POD and DEIM on dimension reduction of non-linear miscible viscous fingering in porous media, Mathematical and Computer Modelling of Dynamical Systems, 17:4, 337-353, DOI: 10.1080/13873954.2011.547660

[8uap] Martin Isoz, POD-DEIM based model order reduction for speed-up of flow parametric studies,

Ocean Engineering, Volume 186, 2019, 106083, https://doi.org/10.1016/j.oceaneng.2019.05.065.

[9uap] Moaad Khamlich, Giovanni Stabile, Gianluigi Rozza, László Környei, Zoltán Horváth, A physics-based reduced order model for urban air pollution prediction, Computer Methods in Applied Mechanics and Engineering, Volume 417, Part A, 2023, https://doi.org/10.1016/j.cma.2023.116416.

[1ub] (2020) Energy efficiency in buildings, https://ec.europa.eu/info/news/focus-energy-efficiency-buildings-2020-lut-17_en, [retrieved 2023-10-19]

[2ub] Rabani, M., Madessa, H. B., & Nord, N. (2021). Achieving zero-energy building performance with thermal and visual comfort enhancement through optimization of fenestration, envelope, shading device, and energy supply system. *Sustainable Energy* Technologies and Assessments, 44, 101020.

[3ub] Garreau, E., Berthou, T., Duplessis, B., Partenay, V., & Marchio, D. (2021). Solar shading and multi-zone thermal simulation: Parsimonious modelling at urban scale. *Energy and Buildings*, *249*, 111176.

[4ub] Schiefelbein, J., Rudnick, J., Scholl, A., Remmen, P., Fuchs, M., & Müller, D. (2019). Automated urban energy system modelling and thermal building simulation based on OpenStreetMap data sets. *Building and environment*, *149*, 630-639.

[5ub] Chen, Y., Hong, T., & Piette, M. A. (2017). Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis. *Applied Energy*, *205*, 323-335.

[6ub] Li, Z., Quan, S. J., & Yang, P. P. J. (2016). Energy performance simulation for planning a low carbon neighbourhood urban district: A case study in the city of Macau. *Habitat International*, *53*, 206-214.

[1res] Wright, D.W., Richardson, R.A., Edeling, W., Lakhlili, J., Sinclair, R.C., Jancauskas, V., Suleimenova, D., Bosak, B., Kulczewski, M., Piontek, T., Kopta, P., Chirca, I., Arabnejad, H., Luk, O.O., Hoenen, O., Węglarz, J., Crommelin, D., Groen, D. and Coveney, P.V. (2020), Building Confidence in

Simulation: Applications of EasyVVUQ. Adv. Theory Simul., 3: 1900246. https://doi.org/10.1002/adts.201900246.

[2res] D. Groen , H. Arabnejad , V. Jancauskas , W. N. Edeling , F. Jansson , R. A. Richardson , J. Lakhlili , L. Veen , B. Bosak , P. Kopta , D. W. Wright , N. Monnier , P. Karlshoefer , D. Suleimenova , R. Sinclair , M. Vassaux , A. Nikishova , M. Bieniek , Onnie O. Luk , M. Kulczewski , E. Raffin , D. Crommelin , O. Hoenen , D. P. Coster , T. Piontek and P. V. Coveney, VECMAtk: a scalable verification, validation and uncertainty quantification toolkit for scientific simulations, Phil. Trans. R. Soc. A.3792020022120200221, https://doi.org/10.1098/rsta.2020.0221.

[3res] B. Bosak, T. Piontek, P. Karlshoefer, E. Raffin, J. Lakhlili, P. Kopta, “Verification, Validation and Uncertainty Quantification of Large-Scale Applications with QCG-PilotJob”, ICCS 2021, Lecture Notes in Computer Science, 12746 (2021), DOI: 10.1007/978-3-030-77977-1_39

[1wf] Finney,M.A.(2004)FARSITE: Fire Area Simulator—Model Development and Evaluation. USDA FS Research Paper RMRS-RP-4 Revised

[2wf] Patton, E. G. and Coen, J. L. (2004) WRF-Fire: A Coupled Atmosphere-Fire Module for WRF, in:5Preprints of Joint MM5/Weather Research and Forecasting Model Users’ Workshop, Boulder,CO, 22–25 June, NCAR, 221–223, 499,514

[3wf] Mandel, J., Beezley,J.D., Kochanski,a.k. (2011). Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011.Geosci. Model Dev., 4, 591–610, 2011

[4wf] Coen, J. L., Cameron, M., Michalakes, J., Patton, E. G., Riggan, P. J., and Yedinak, K. (2013) WRF-Fire: Coupled Weather-Wildland Fire Modeling with the Weather Research and Forecasting Model, J. Appl. Meteor. Climatol., 52, 16–38, doi:10.1175/JAMC-D-12-023.1

[5wf] Kochanski,A.K., Jenkins,M.A., Mandel,J., Beezley,J.D.,Clements,C.B., Krueger,S. (2913) Evaluation of WRF-SFIRE performance with field observations from the FireFlux experiment. Geosci. Model Dev., 6, 1109–1126.

[6wf] Mandel,J., Amram,S., Beezle,J.D., Kelman, G., Kochanski,A.K., Kondratenko,V.Y.,Lynn,B.H.,Regev,B., Vejmelka,M. (2014) New features in WRF-SFIRE and the wildfire forecasting and danger system in Israel. Nat. Hazards Earth Syst. Sci. Discuss., 2, 1759–1797, 2014 www.nat-hazards-earth-syst-sci-discuss.net/2/1759/2014/ doi:10.5194/nhessd-2-1759-2014.

[7wf] Mell, W.,Jenkins,M.A., Gould,J.,Cheney,P. (2007) A Physics-Based Approach to Modeling Grassland Fires.10.1071/WF06002_AC©IAWF 2007 Accessory Publication: International Journal of Wildland Fire, 2007, 16(1), 1–22.

[8wf] Hoffman CM, Ziegler J, Canfield J, Linn RR, Mell W, Sieg CH, Pimont F (2016) Eval uating crown fire rate of spread predictions from physics-based models. Fire Technol 52:221–237. https://doi.org/10.1007/s10694-015-0500-3.

[9wf] Derek V. Mallia, Adam K. Kochanski (2023) A Review of Modeling Approaches Used to Simulate Smoke Transport and Dispersion. Landscape Fire, Smoke, and Health: Linking Biomass Burning Emissions to Human Well-Being.

[10wf] Firefighter Training VR by Chronos VR https://www.firefightervr.de/?locale=en.

[11wf] Xu,Z., Lu,X.Z.,Guan,H.,Chen,C.,Ren,A.Z.,(2014) A virtual reality based fire training simulator with smoke hazard assessment capacity. Advances in Engineering Software Volume 68, February 2014, Pages 1-8.

[12wf] Hazmat Virtual VR First Responder Training https://hazmatvirtual.com/.

[13wf] VRFire trainer https://www.vrfiretrainer.com/.

[14wf] Graziani, A.; Meerpoel-Pietri, K.; Tihay-Felicelli, V.; Santoni, P.-A.; Morandini, F.; Perez-Ramirez, Y.; Mell, W. Modelling the Burning of an Ornamental Vegetation with WFDS: From Laboratory to Field Scale. Environ. Sci. Proc. 2022, 17, 118.