D3.5 Report on Dynamic Data Assimilation Methodology

Deliverable number	**D3.5**
Deliverable title	**Report on Dynamic Data Assimilation Methodology**
Nature[1]	**R**
Dissemination Level [2]	**Public**
Author (email) Institution	**Fotios Barmpas (fotisb@auth.gr)** **George Tsegas (gtsegas@auth.gr)** **Eleftherios Chourdakis (chourdakis@auth.gr)**
Editor (email) Institution	**Andreas Papadopoulos (a.papadopoulos@cyric.eu)**
Leading partner	**AUTh**
Participating partners	**CYRIC, FTI**
Official submission date:	**30/9/2021**
Actual submission date:	**1/06/2022**

[1] **R**=Document, report; **DEM**=Demonstrator, pilot, prototype; **DEC**=website, patent fillings, videos, etc.; **OTHER**=other

[2] **PU**=Public, **CO**=Confidential, only for members of the consortium (including the Commission Services), **CI**=Classified, as referred to in Commission Decision 2001/844/EC

Modifications Index

Date	Version
01/04/2022	0.1 Table of content
04/04/2022	0.2 First version
06/05/2022	0.3 Second version
15/05/2022	0.4 Reviewed
01/06/2022	0.5 Final

This work is a part of the HYPERION project. HYPERION has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 821054.

ACRONYMS AND ABBREVIATIONS

3-D	Three Dimensional
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
AUTh	Aristotle University of Thessaloniki
CH	Cultural Heritage
DA	Data Assimilation
EZM	European Zooming Model
GeoTIFF	Geographic Tagged Image File Format
GPS	Global Positioning System
GRIB	GRIdded Binary
CHRAP	Cultural Heritage Risk Assessment Platform
IDW	Inverse Distance Weighted
JSON	JavaScript Object Notation
NASA	National Aeronautics and Space Administration
NetCDF	Network Common Data Form
NMSE	Normalised Mean Square Error
npy	Numerical Python
OMS	Operational Mesoscale System
PNG	Portable Graphics Format
TKE	Turbulent Kinetic Energy
UTC	Coordinated Universal Time
UTM	Universal Transverse Mercator

Table of Contents

Executive Summary 6
1. Introduction 7
1.1 Purpose and Scope 7
1.2 General Idea 7
2. Mesoscale Model Calculations 7
2.1 Overview 7
3. Data Assimilation Approach 14
3.1 General Idea 14
3.2 Methodological Background 15
3.3 Data Assimilation Module 16
3.4 Pilot Application Set-up 19
4. Results 21
4.1 Pilot Mesoscale Results 21
4.2 Validation of Pilot Mesoscale Results 24
4.3 Pilot Application of the Data Assimilation Approach 26
5. Conclusions 30
6. REFERENCES 32

List of Tables

Table 1 Specification and dimensions of OMS domains. **Error! Bookmark not defined.**

Table 2 Specification of the output of nowcasting and forecasting modes ..... **Error! Bookmark not defined.**

Table 3 Validation of the mesoscale results for the four pilot domains ..... **Error! Bookmark not defined.**

Executive Summary

This deliverable aims at describing a newly-developed methodology implemented by AUTh for the continuous update and refinement of numerical modelling results based on in-situ measurements. This method is based on a 3-D data assimilation protocol which supports fusion of data-streams both from micro weather stations installed at a number of selected demonstration pilots in Task 3.4 and from the networked unattended and low cost micro-climate stations developed within Task 3.5. More specifically, the data aggregated from the sensor network and delivered through the monitoring system are initially engaged with the simulated data and, as a second step, are analysed and fused (including aggregation, synchronization, calibration and assimilation) from a fully interoperable data management platform. In this way, local-scale atmospheric and climate stressing on the CH of the selected demonstrations pilots which are in principle not resolved are explicitly introduced into the high-resolution simulations.

1. Introduction

1.1 Purpose and Scope

This document is an outcome of Task 3.5 "Dynamic Data Assimilation of meteorological and climate data from sensors" towards the continuous update and refinement of numerical modelling results based on in-situ measurements. This deliverable focuses on the methods, approaches and techniques utilized for the production of revised meteorological and climate fields for a number of parameters of interest.

1.2 General Idea

The general idea for the present approach is based on the realization of a data assimilation layer used to assure for the proper transformation of real data into simulation parameters which feed back the modelling engine and contribute to more accurate and dynamically updated predictions. In this direction, AUTh has implemented, validated, calibrated and operationally utilized the application of the 3-D assimilation of data from micro-climate stations.

2. Mesoscale Model Calculations

2.1 Overview

An operational modelling module (Operational Modelling System - OMS) for the prediction and identification of extreme climate indicators and atmospheric stressors in connection to specific hazards has been developed by the Laboratory of Heat Transfer and Environmental Engineering (LHTEE) in the framework of WP3 Task 3.5. The module provides functionality for downscaling fields of temperature, relative humidity, precipitation, wind velocity and cloud cover through an operational forecast of main meteorological parameters. In this point it should be pointed out that OMS is designed to operate as an embedded module of the HYPERION CHRAP system (Cultural Heritage Risk Assessment Platform).

2.2 The MEMO model

The part of the OMS dealing with the calculation of meteorological parameters has an operative structure comparable to that followed by most systems performing on an constant basis. In greater detail, OMS encompasses two distinctive operation modes, nowcasting and forecasting, which are implemented based on a scheduler infrastructure controlling a set of grid meteorological model simulations. The mesoscale meteorological model MEMO (Moussiopoulos et al., 2012), part of the EZM (European Zooming Model) system (Moussiopoulos N., 1995), is used for this purpose. MEMO is a three-dimensional, non-hydrostatic, prognostic mesoscale model for the simulation of mesoscale air motion and inert pollutant dispersion at the local-to-regional scale, over complex terrain, allowing multiple nesting.

The MEMO model is capable of efficiently producing simulated datasets of hourly meteorological parameters, such as wind speed and direction, temperature, turbulent kinetic energy (TKE), incident solar radiation, cloud cover (diagnostically) and relative humidity, as well as turbulence parameters like surface roughness, Monin-Obukhov length and friction velocity for each grid point and at multiple heights above ground level in a 3-dimensional geographical grid, covering the target region. These high-resolution downcaled datasets offer flexibility in terms of data extraction. Another important feature of MEMO is that it is also capable of simulating local circulation systems, such as mountain-valley winds, sea/lake breezes, as well as the urban heat island. These mesoscale atmospheric processes also affect local-to-regional scale dispersion phenomena. Therefore, MEMO is considered as a suitable scientific tool to generate such comprehensive meteorological data for the needs of meteorological downscaling, as well as atmospheric pollutant dispersion.

For the initialisation of MEMO, a number of vertical profiles of the key meteorological variables originating from the ICON-EU (Zängl et al., 2015) operational numerical weather prediction model are used. Such profiles are assimilated into the model calculations on a 3-hour basis. The grid structure of ICON global model is based on an icosahedral (triangular) grid of the earth's sphere. The forecast data are also provided in standard packages on an icosahedral (triangular) grid. Regarding the ICON-EU calculations, there is a tightly coupled two-way interaction between the ICON-EU regional model and the global ICON.

The native model grid has a horizontal grid spacing of 6.5 km, the output grid a grid spacing of 0.0625 °(~ 7 km). In the vertical, ICON-EU relies on 60 levels up to a height of 22.5 km. Besides, the ICON-EU forecasts are available up to 120 hours from the four model runs at 00, 06, 12 and 18 UTC and up to 30 hours from the model runs at 03, 09, 15 and 21 UTC. The time interval for the forecast period up to 78 hours is one hour, while the forecast periods between +81 and +120 hours are covered by a 3-hourly time interval.

More specifically, the downloader module of the automated module undertakes the transfer of the ICON-EU data via the CHRAP system, processes them (the ICON-EU data files are initially decompressed and, as a second step, converted from grib2 format into ASCII), and stores them in a dynamic data pool, also part of the CHRAP infrastructure, which is kept updated at all times. The scheduler selects only the most recent dataset for input to the MEMO model. Each of these processes keeps a separate event log and diagnostic files accessible through the CHARP repository.

2.3 Configuration of MEMO model

The OMS has been configured for the four pilot areas in Spain (Granada), Greece (Rhodes), Italy (Venice) and Norway (Tonsberg). The configuration of the MEMO grid for all areas is shown in Table 1, while Figures 1-4 depict the domain extents and the topography layer providing a high-resolution representation of the topography for the model domain for all under consideration.

As regards the required high resolution topographical input data, they were derived from the satellite elevation datasets of NASA's Shuttle Radar Topography Mission - SRTM/90 m database (URL1). Besides, thematic layers of land use data were obtained from the Corine Land Cover 2006 (URL2) database, which includes 44 land use (LU) types, which were reclassified to the seven LU types used in MEMO applications, namely water, arid land, few vegetation, farmland, forests, suburban and urban.

Table 1: Specification and dimensions of OMS domains

Granada						
Grid	**Domain Center (UTM)**		**Domain size (km)**	**Domain size (cells)**	**Spatial Resolution (m)**	**Temporal Resolution (h)**
	x	**y**				
Coarse	446460	4085598	144	72	2000	1
Fine	446460	4115598	30	120	250	1
Rhodes						
Grid	**Domain Center (UTM)**		**Domain size (km)**	**Domain size (cells)**	**Spatial Resolution (m)**	**Temporal Resolution (h)**
	x	**y**				
Coarse	586151	4004227	120	120	1000	1
Fine	609698	4032194	30	120	250	1
Venice						
Grid	**Domain Center (UTM)**		**Domain size (km)**	**Domain size (cells)**	**Spatial Resolution (m)**	**Temporal Resolution (h)**
	x	**y**				
Fine	288629	5036430	50	200	250	1
Tonsberg						
Grid	**Domain Center (UTM)**		**Domain size (km)**	**Domain size (cells)**	**Spatial Resolution (m)**	**Temporal Resolution (h)**
	x	**y**				
Fine	580992	6566221	50	200	250	1

Figure 1: Domain extents and the topography layer for the pilot domain of Granada.

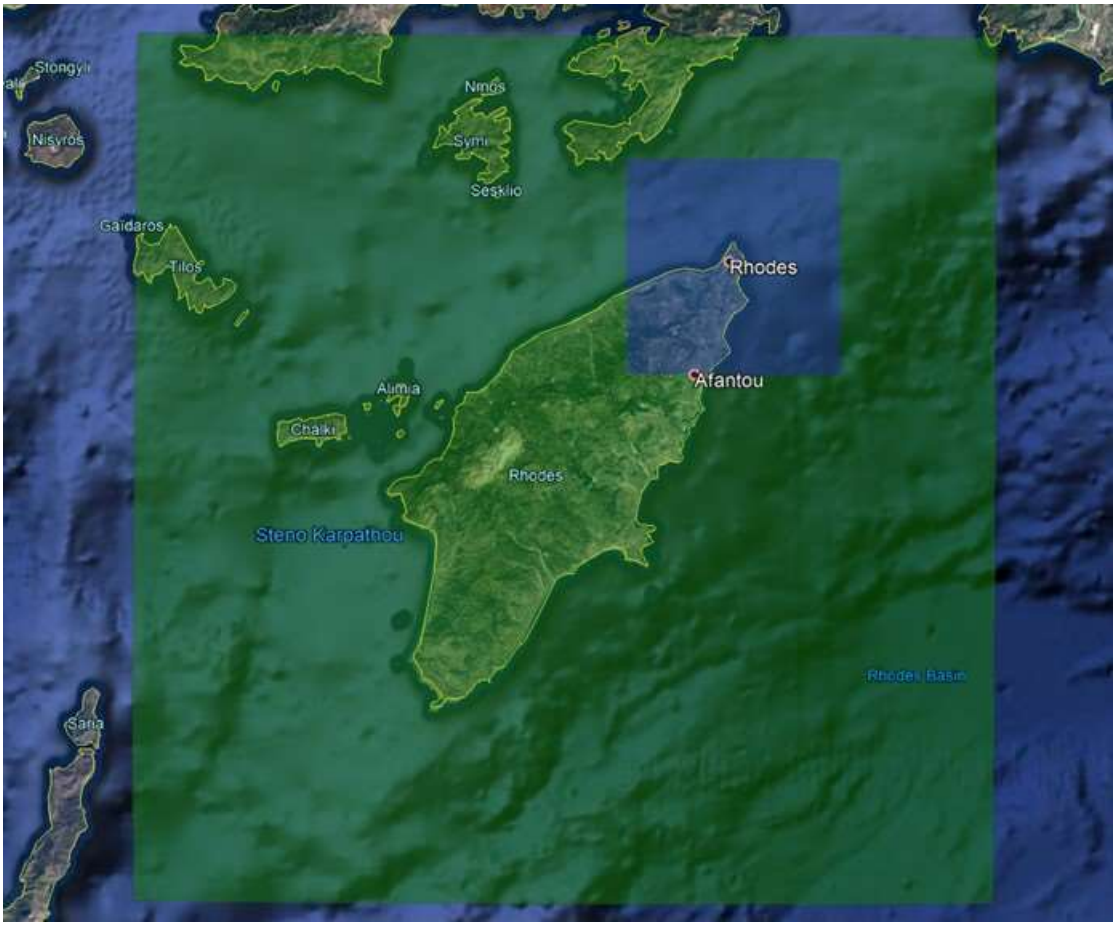

Figure 2: Domain extents and the topography layer for the of the model domain for the pilot domain of Rhodes.

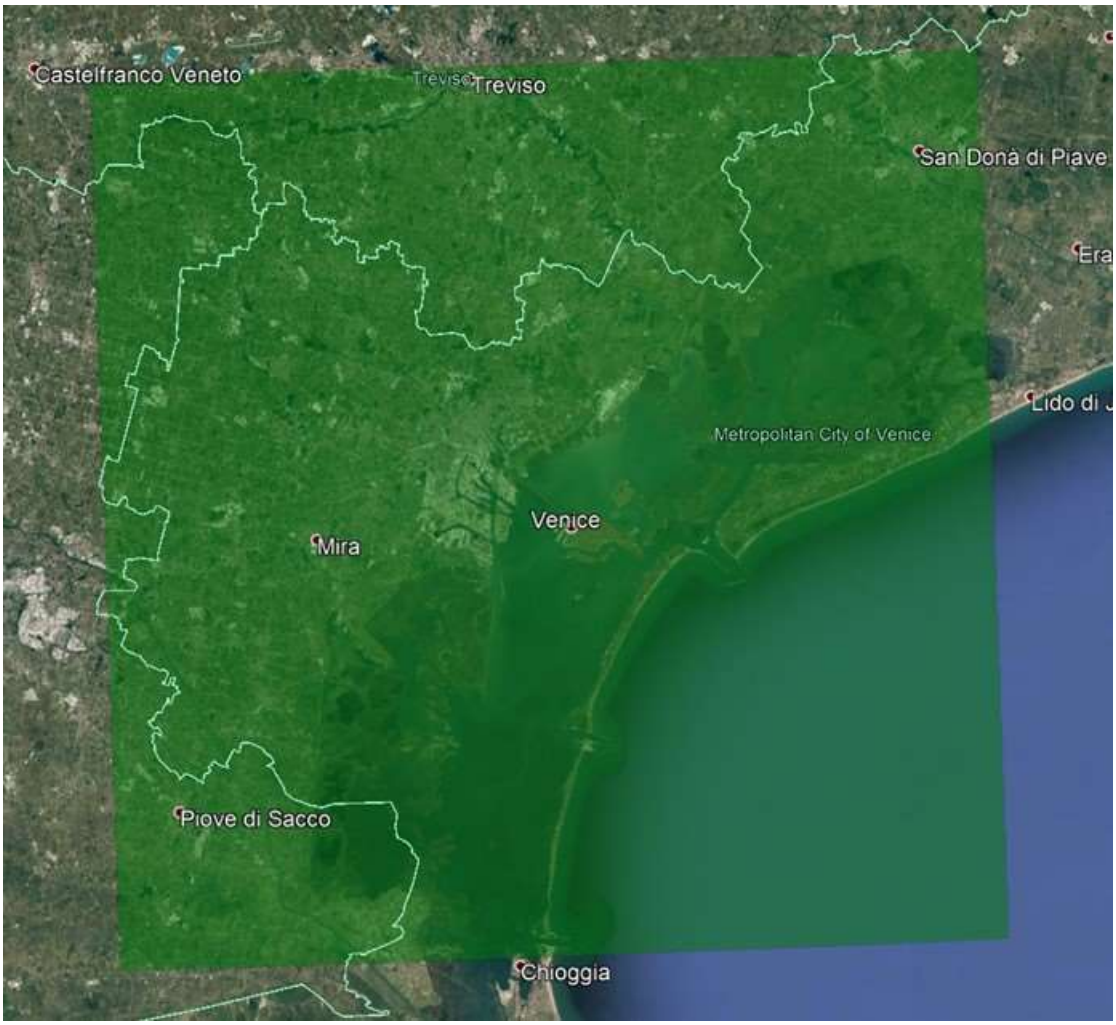

Figure 3: Domain extents and the topography layer for the of the model domain for the pilot domain of Venezia.

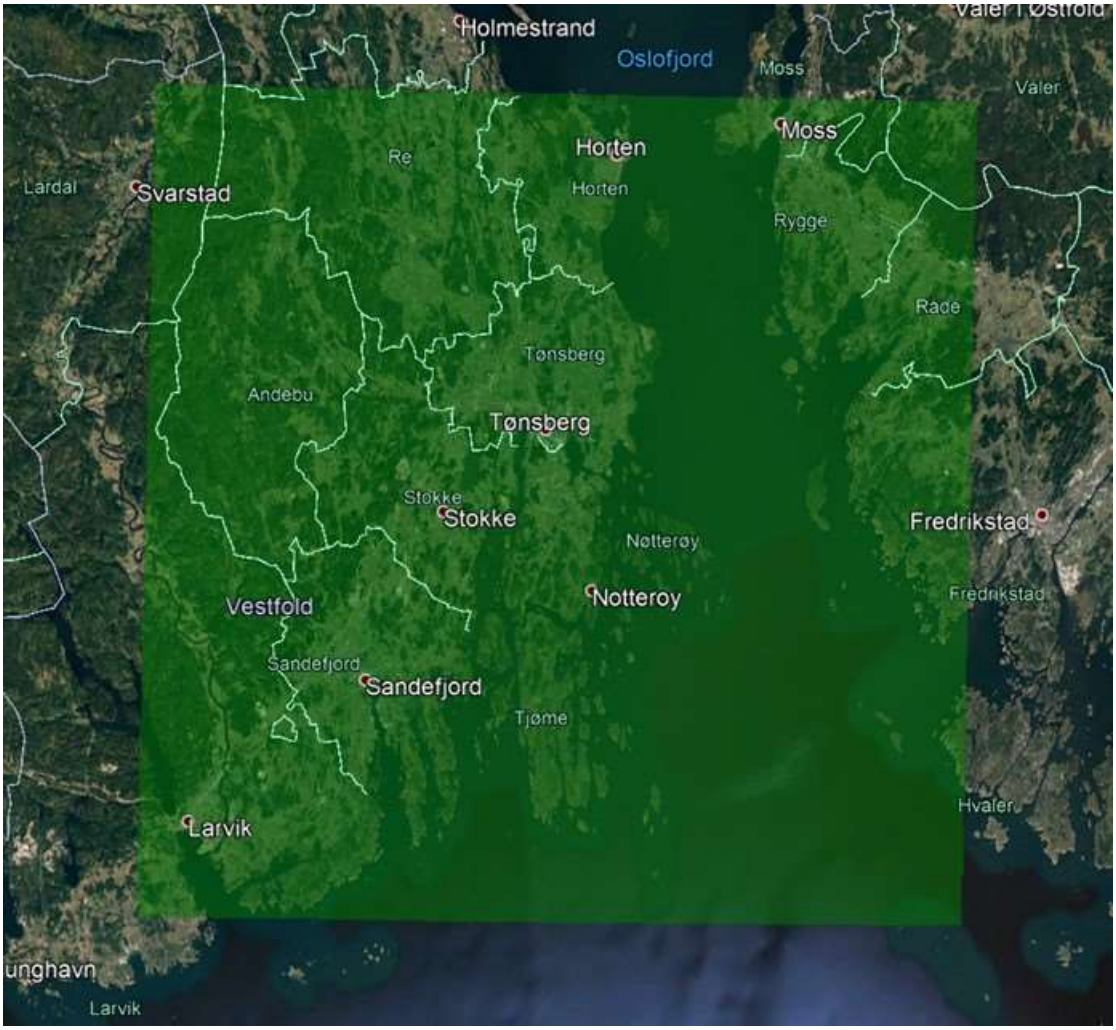

Figure 4: Domain extents and the topography layer for the of the model domain for the pilot domain of Tonsberg.

2.4 Initial Mesoscale Results

The meteorological mesoscale module provides functionality for downscaling several meteorological parameter fields which, amongst other, include temperature, humidity, incident solar radiation, relative humidity, precipitation, cloud cover and of course wind intensity and direction. At the nowcasting mode, model results are provided at an hourly basis for the current hour, as well as for the next three (3) hours as short term predictions. At the forecasting mode, results are provided at an hourly basis over the 24-hour period of the next calendar day (00:00 – 23:00). Model results are subsequently processed in an automated way by suitable post processing tools based on the object-oriented scripting language Python and a range of maps, timeseries graphs and statistical indices are produced for a number of predefined climate and atmospheric risk factors.

More specifically, in the nowcasting mode, the system computes, on an hourly basis, high resolution downscaled meteorological fields for the domains of interest. Model results are automatically processed and a range of maps, timeseries graphs and statistical indices are produced for a number of predefined climate and atmospheric

risk factors. As regards the forecasting mode, next-day 24-hour meteorological model simulations are driven by the corresponding prognostic meteorological simulations of ICON-EU in order to produce daily sets of downscaled parameters graphs for all the designated locations of interest. The calculated datasets of both operational modes, are subsequently stored to the Hyperion system middleware or the CHRAP system. Table 2 depicts the specification of the output for both operational modes.

Table 2: Specification of the output of nowcasting and forecasting modes.

Parameter (unit)	Time reference	Spatial Resolution	Vertical Reference
Temperature (^{0}C)	1 h	250 m	2 m
Relative Humidity (%)			
Wind Speed (m/s)			
Wind direction (^{0})			
Precipitation (mm)			Surface
Cloud Cover (%)			-

3. Data Assimilation Approach

3.1 General Idea

It is a common practice among scientists to use ground based monitoring in order to provide information regarding meteorological and climate aspects, since it is expected to provide the best estimates. This may be suitable when a very limited area is to be assessed or in cases where monitoring data is representative of a large area. Nevertheless, local measurements have a limited spatial representativeness. This can be problematic in urban areas, since there can be significant variation in local meteorology due to the complex flow patterns caused by urban morphology (Denby et al, 2009).

A significant improvement in the spatial representativeness of the monitoring data can be achieved with the use of supplementary data sources, which are capable of providing a better spatial coverage than the monitoring data itself. Such supplementary information may include distances from major topographical features,

land use characteristics, satellite data, etc.. Though it is possible to use these data directly through a range of spatial statistical methods, it is the meteorological models that best describes the relevant physical processes and provide high spatial and temporal resolution data that can be used for improving the coverage of the monitoring information (Denby et al, 2009).

The major drawback of modeling is its level of uncertainty, which is usually significantly higher than that for monitoring. As a result, it would be beneficial to combine the monitoring and modeling data sources in an optimal way to produce spatio-temporal meteorological maps.

3.2 Methodological Background

There are a number of terms used to describe the combination of different data sources. Interpolation refers to methods that use monitoring as the primary dataset and, based on these data and possibly other supplementary data, predict values of meteorological parameters at any arbitrary point in space (e.g. Beelen, 2009).

Spatial interpolation techniques can be divided into two main categories, those which are based on deterministic approaches and those which are perfomed with the use geostatistical methods. As regards the first category, deterministic methods do not attempt to capture the spatial structure in the data, but to utilize predefined mathematical equations to predict values at various unsampled locations by weighing the attribute values of samples with known location (Aspexit, 2019).

On the contrary, geostatistical approaches intend to fit a spatial model to the data. This approach enables the generation of prediction values at unsampled locations and to provide users with an estimate of the accuracy of this prediction. Regarding the deterministic methods, the TIN, IDW and Trend surface analysis techniques are the most commonly used, while geostatistical approaches include kriging and its variants (Isaaks and Srivastava, 1989). In the frame of the present application, the kriging method is utilized.

In addition, methods combining a set of potential data sources contain data fusion and data integration approaches. They take any number of datasets and combine these in a range of ways, either through geometric means or based on statistical optimization methods. Data fusion and interpolation methods are generally not concerned with any physical or chemical constraints but are mainly subject to statistical constraints (Denby et al, 2009).

On the other hand, data assimilation refers to a modeling technique which incorporates monitoring data directly into the model simulations during the modeling process itself. It is the measured data which play of major role in the guidance of the model towards an optimal solution, and one that is consistent with the physical description provided by the model. The most common type of data assimilation applied are the variational methods (Elbern et al., 1999), which are extensively used

in meteorological forecast, but other methods such as Ensemble Kalman filters (van Loon et al., 2000) may also been applied.

In the frame of the present approach, a combination of data fusion and data assimilation methods has been employed to incorporate observational data in the OMS. The modelling tools incorporate and complement the sensor data on meteorological parameters with predictions of high temporal and spatial resolution, covering the areas of interest, to underline possible hazardous or stressful conditions.

Observational data are integrated on an hourly frequency and continuously assimilated into model calculations with the aid of a spatial interpolation method (kriging – Goovaerts, 1997). In more detail, an automated downloader module downloads and quality-assures the sensor data, which are then fed into the appropriate format into the interpolation module. A sophisticated spatial and temporal analysis algorithm is then applied, whereby outlier measurements related to accumulated error are distinguished from actual local extremes and either discarded or corrected.

Kriging is the most commonly used geostatistical approach for spatial interpolation and is based on a multistep process. More specifically, it includes exploratory statistical analysis of the data, variogram modeling (Webster and Oliver, 1993), creating the surface, and exploring a variance surface. In spatial statistics the theoretical variogram is a function describing the degree of spatial dependence of a spatial random field or stochastic process.

A variogram is a half of the variance sum of the increment that is the regionalized variables Z(x) at the x and x + h. The common theoretical variogram fits the function model: spherical model, exponential model, power function model, and logarithmic function model. Regarding its applicability, kriging could be the most appropriate approach in cases that there is a spatially correlated distance or directional bias in the data.

In order to further ensure that data flows from the observational network are appropriately incorporated into consistent fields, a continuous on-line quality-testing and statistical evaluation process is performed on the basis of measurements from traditional monitoring devices. Model results and measurements are further processed off-line in order to assess the validity of the simulations, as well as to detect diurnal, seasonal or spatial patterns, which could lead to additional correction factors.

3.3 Data Assimilation Module

Core of the Methodology

The utilization of measurement data can be achieved through the implementation of a system of measurement fusion (assimilation) and analysis of comptutational data (reanalysis), based on the use of data from both measurements and archived hourly forecasts of the OMS for the areas of interest.

With respect to the above, a newly developed function has been enhanced in the system, which is capable of reconstructing the meteorological fields for the entire computational area of interest, for any period for which there are available data both from measurements and MEMO model calculations. For the quantification and estimation of the total uncertainties introduced in the estimated meteorological fields, uncertainties and representativeness errors in the primary measurements, as well as the various forms of uncertainty that come up with the model results are taken into account.

This process can be conducted manually, giving the testing operator the capability to select the starting and ending dates of the application period, or applied during operational nowcasts and forecasts. Regarding the operational functioning of this approach, its computational infrastructure has been incorporated in the OMS core, in order to produce real time corrections for the simulated meteorological fields. In this context, the reanalysis function has been added both in the operational schedule of the system, as well as to the list of eligible system functions which can be used either manually.

The computational part of this function is coordinated by a scheduler infrastructure which undertakes at the initial level the collection of all available data, derived from both measurements and model results, for a predefined time period regarding the four pilot areas under investigation. This task is undertaken by an efficient downloader module which conducts operational observational data retrievals from micro weather stations installed at the selected demonstration pilots.

Data Retrievals

The data assimilation procedure is applied to increase the accuracy of the OMS calculations. The procedures data (in-situ measurements) are collected through the data retrieval process. The data are retrieved from weather stations via Middleware API. In three pilot cases (Granada, Rhodes, and Venice) one weather station measures different meteorological parameters, while in Tonsberg, the observations are collected from two weather stations.

The values of the following measured meteorological parameters are used as input for the data assimilation process:

- Air temperature (^{o}C)
- Air relative humidity (%)
- Wind speed (m/s)
- Wind direction (^{o})

The data retrieval process uses the Middleware API to collect the data of the above-mentioned variables in real-time or near real-time. The whole process is applied every hour to collect the data of the past 60 minutes. The locations of weather stations are presented in Figure 5.

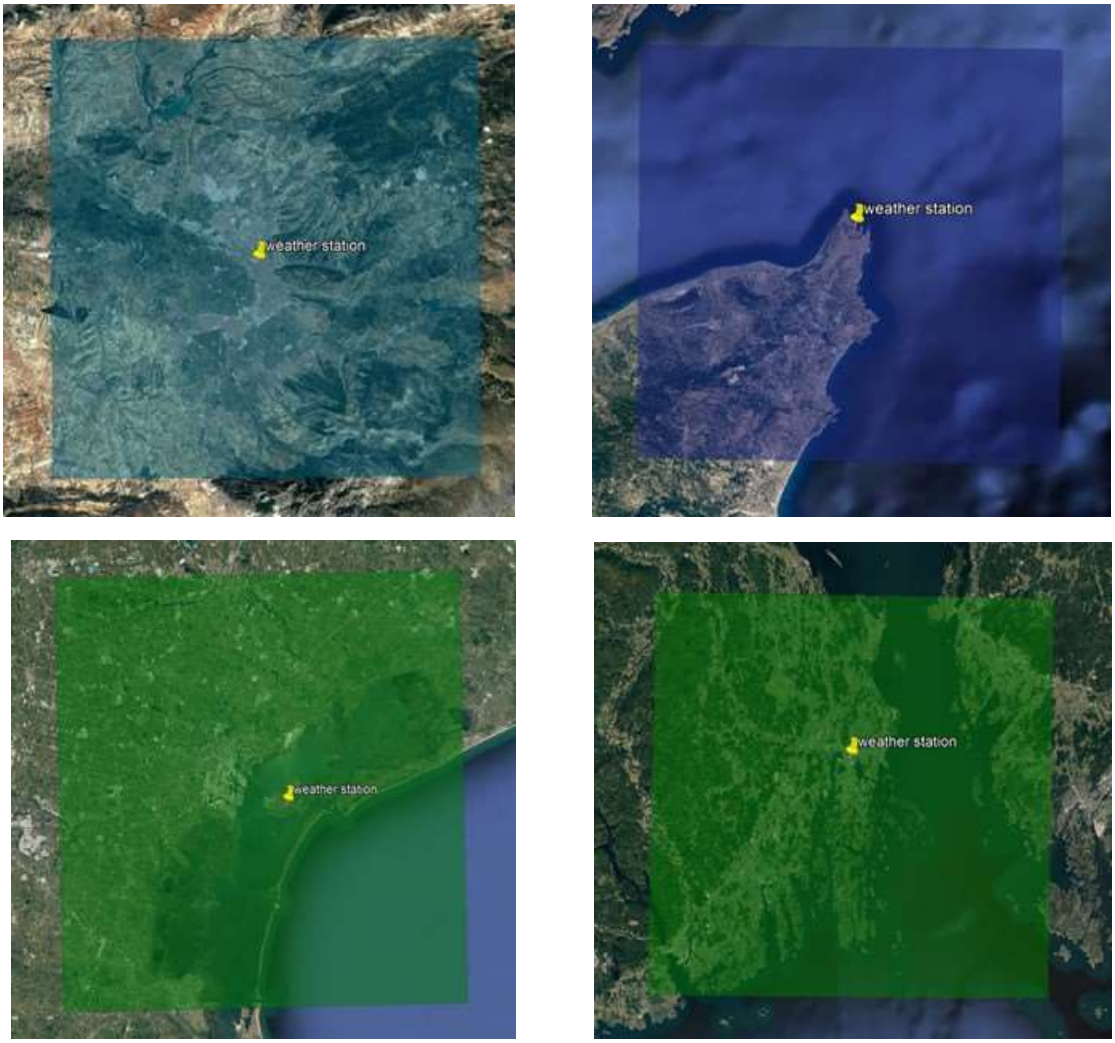

Figure 5: Weather stations used for validation and assimilation purposes in the case study of Granada (upper left), Rhodes (upper right), Venice (lower left) and Tonsberg (lower right).

Computational Part

The next step in the application of the methodology is related to the calculation of the hourly differences between the measured and calculated values of each meteorological parameter of interest at the points where the measuring stations are located. These differences are then fed into a suitable computing infrastructure, which, using spatial interpolation techniques, produces a map of differences relating to the pilot area under investigation. Spatial interpolation is the process of using points with known values to estimate values at other unknown points. In this point, it should be noted that the spatial resolution of these meteorological fields is of the order of 250 m.

This method assumes that the distance and direction between the sampling points reflect a spatial correlation, which can be used to explain the spatial change in the surface. More specifically, a mathematical function is applied to a certain number of points, or all points within a detection radius, to calculate the value at each position.

In addition to producing a predictable surface, it has also the ability to produce estimates of the prediction error of each variable of interest at each sampling position.

The final part of the application is the correction of the initial meteorological fields of the model simulations using the difference fields produced in the previous step. In particular, the difference maps are added incrementally to the original fields for each hour of the defined period of interest in order to correct the initial calculations. Then, the improved fields are used by the system to create numerous meteorological maps in various formats (GeoTIFF, PNG, NetCDF, npy) which are kept in specific subfolders in the servers of AUTh and, periodically, uploaded to the Middleware.

In conclusion, this assimilation and analysis subsystem provides the ability to accurately assess meteorological status for previous operating periods (reanalysis), based on measurement data and model results (if any), but also to introduce real time corrections on the operationally produced meteorological fields of OMS. This information can be provided for any point or area of application domains, in higher resolution and with greater accuracy than would be possible using only time series measurements.

3.4 Pilot Application Set-up

With regard to the OMS operational application, in conjunction with assimilated data streams from a sensor network, a functional setup has been under way for the four areas under investigation, with a number of technical aims which are described followingly.

First of all, the joint functioning of the mesoscale model and the data assimilation approach during the operational deployment of the OMS can be described as a sequence of steps conducted in an operational manner, which are repeated on an hourly basis, or whenever a new pack of sensor data becomes available. Throughout this period of time, iterations of the last four steps of the aforementioned progression can be performed internally, aiming at the improvement of the accuracy of the last available data.

As regards the goal of each step, the first one deals with the retrievals of the observational data. More specifically, this procedure controls all necessary low-level operations for the communication with the station monitoring software, the confirmation of the availability of updated data, as well as the conversions regarding the formats and the management of errors. Data received includes time-stamps and location/GPS data (in case of mobile sensors).

The second step handles the rejection of all outdated data. In this step, the time-stamps of the received data are examined with the use of a proper algorithm and out-of-order data are discarded. At the same time, mobile sensor location data are checked, as well, and invalid positionings (e.g. outside the fine computational grid) are also eliminated.

In the next step in the sequence, a set of sophisticated testing procedures is perfomed aiming at the rejection of outliers. More specifically, measurement data are checked

against predefined (per-species) hard bounds. Data from stations with unjustifiably high spatial variability are selectively rejected.

During step four measurement data are "normalised" in order to remove as much of the systematic/drift error as possible, on the basis of "offset corrections" calculated at each assimilation cycle (step 7), as residuals between calculations and observations from nearby stations. A machine-learning algorithm is responsible for updating the corrective factors in the offset calculation.

Followingly, a calculation of numerical tendencies takes place. In this direction,a set of difference terms indicating the required adjustments to the estimated values are calculated to be applied in one or more computational cells of MEMO after an appropriate spatial and temporal "spreading" (next step).

During the sixth step, the dynamical equations are time-stepped by the numerical solver as usual, but including the additional corrective terms in order to result in the incorporation of tendencies in the dynamical terms and step integration.

Finally, the extraction of normalisation terms/data is conducted. During this step, the last available measurements are compared to the corresponding simulations and a potential correction factor is calculated for the observations using a multilinear empirical equation. The equation factors are updated by a machine-learning algorithm using as convergence criteria (1) the minimization of total prediction error over all the sensors and over a specific period of time (2) the minimization of monotonic components (drift) in the measurements over the course of 24- or 48-hour periods.

4. Results

4.1 Pilot Mesoscale Results

As already described in detail, the OMS has been configured for four pilot areas in Spain,Greece, Italy and Norway (Granada, Rhodes, Venice and Tonsberg, respectively). Figures 6-9 depict visualizations of the spatial distribution of temperature at the surface level for a typical spring day as regards the four domains under investigation, respectively, originating from the initial MEMO simulations and after the application of the data assimilation process. Besides, Figures 10-13 depict visualizations of the spatial distribution of wind speed at the surface level for a typical winter day as regards the four domains of interest originating from the initial MEMO simulations.

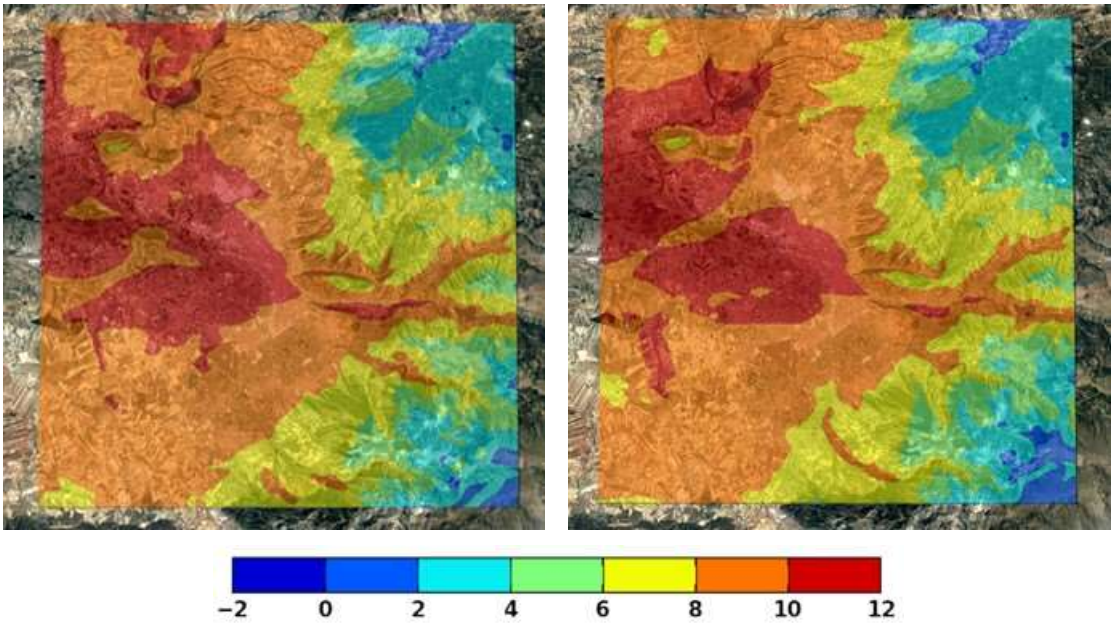

Figure 6: Temperature fields (^{0}C) at 2 m height above ground as regards the Granada fine domain originating from the initial MEMO calculations.

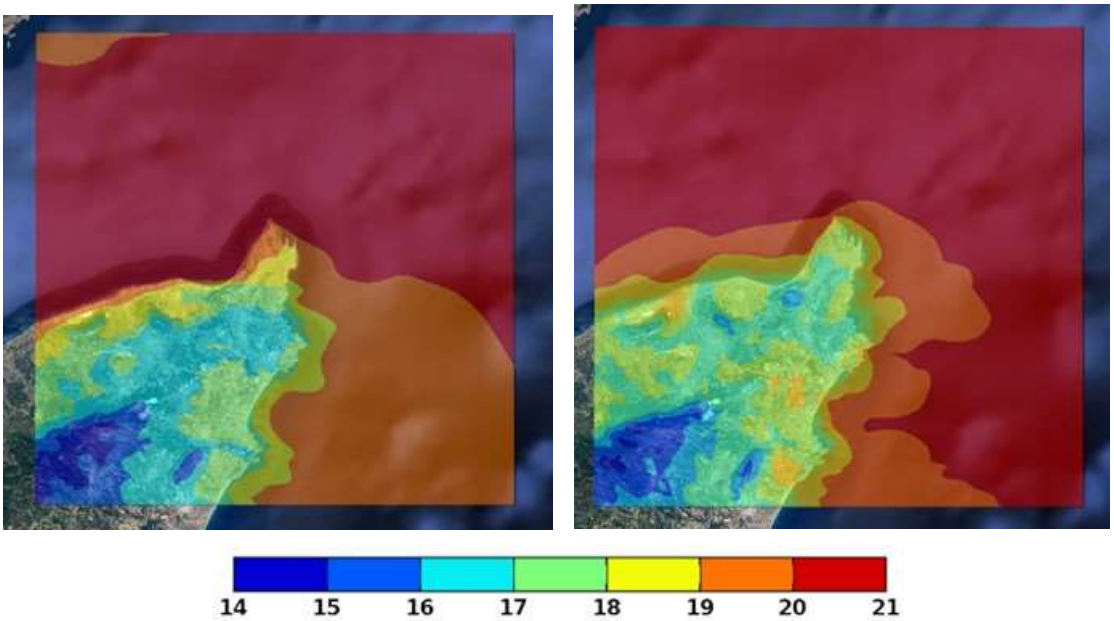

Figure 7: Temperature fields (^{0}C) at 2 m height above ground as regards the Rhodes fine domain originating from the initial MEMO calculations.

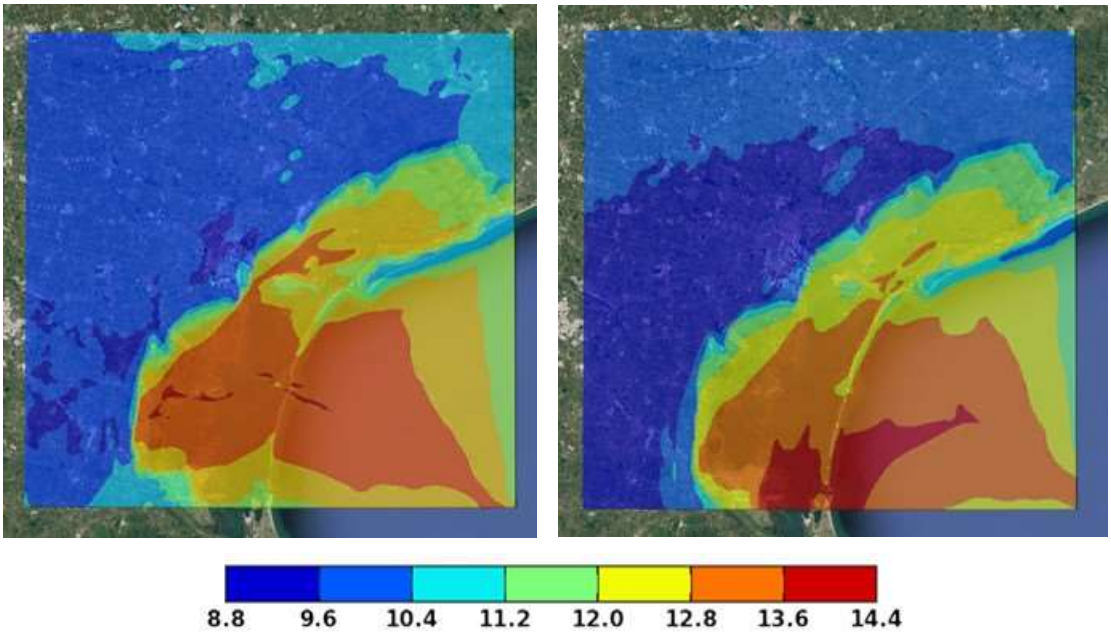

Figure 8: Temperature fields (^{0}C) at 2 m height above ground as regards the Venice fine domain originating from the initial MEMO calculations.

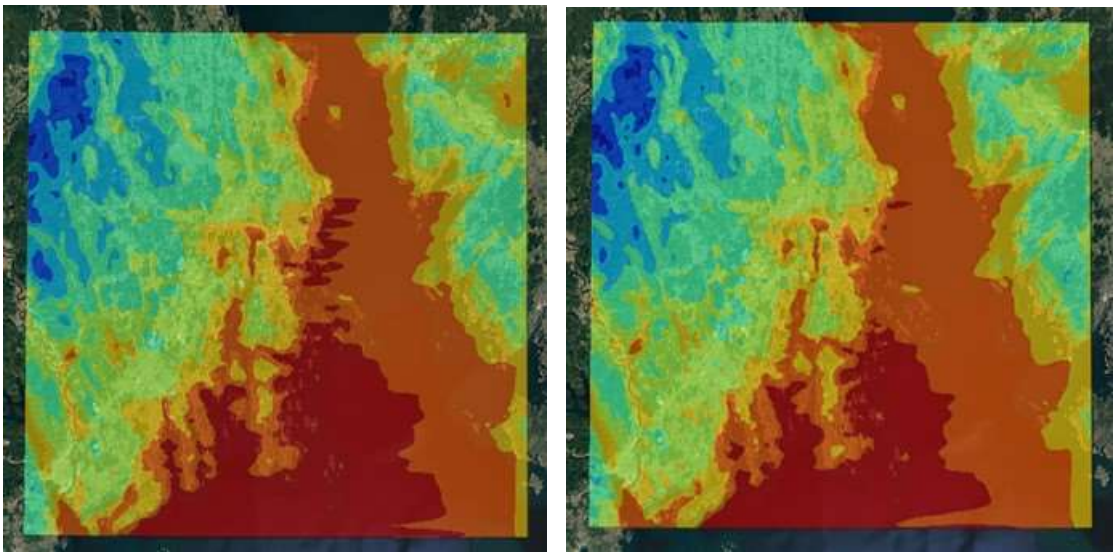

Figure 9: Temperature fields (^{0}C) at 2 m height above ground as regards the Tonsberg fine domain originating from the initial MEMO calculations.

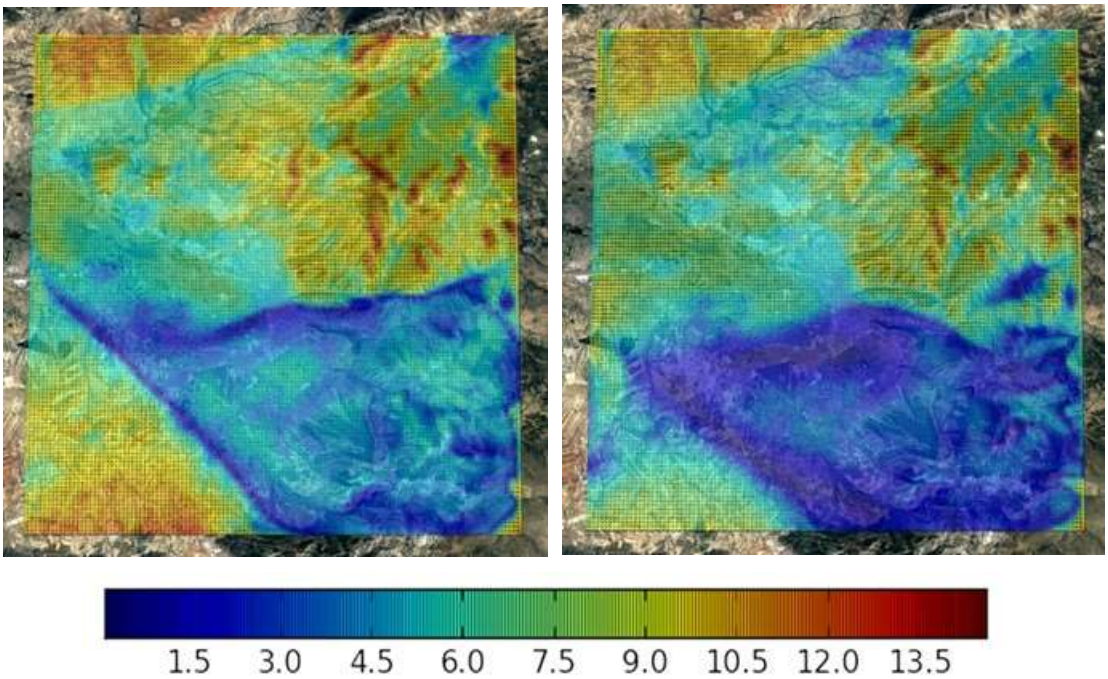

Figure 10: Wind fields (m/s) at approximately 10 m height above ground as regards the Granada fine domain originating from the initial MEMO calculations.

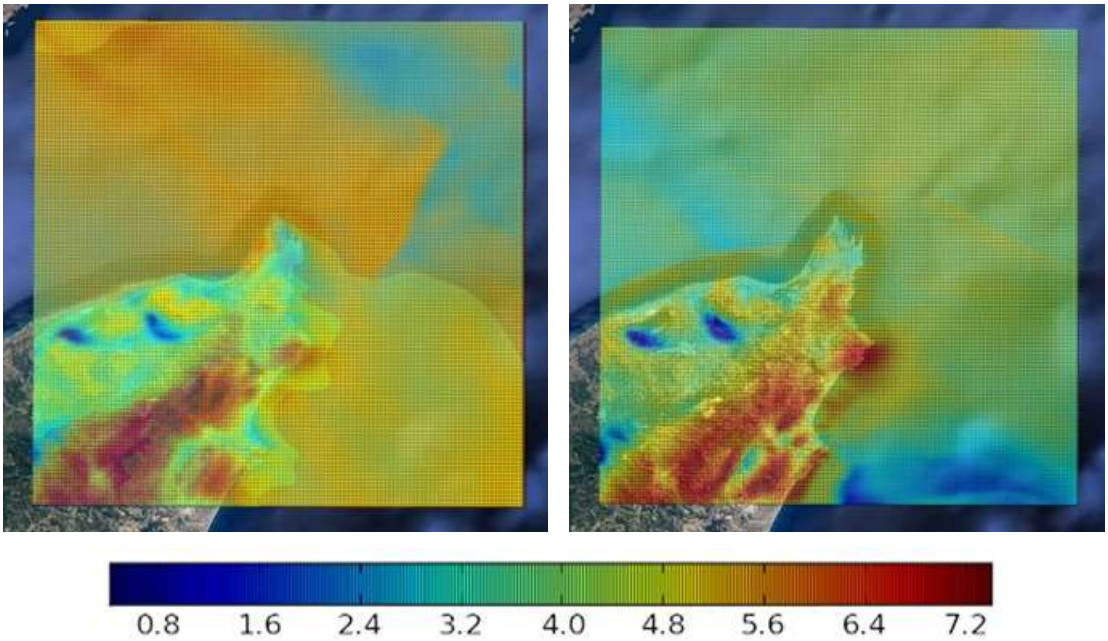

Figure 11: Wind fields (m/s) at approximately 10 m height above ground as regards the Rhodes fine domain originating from the initial MEMO calculations.

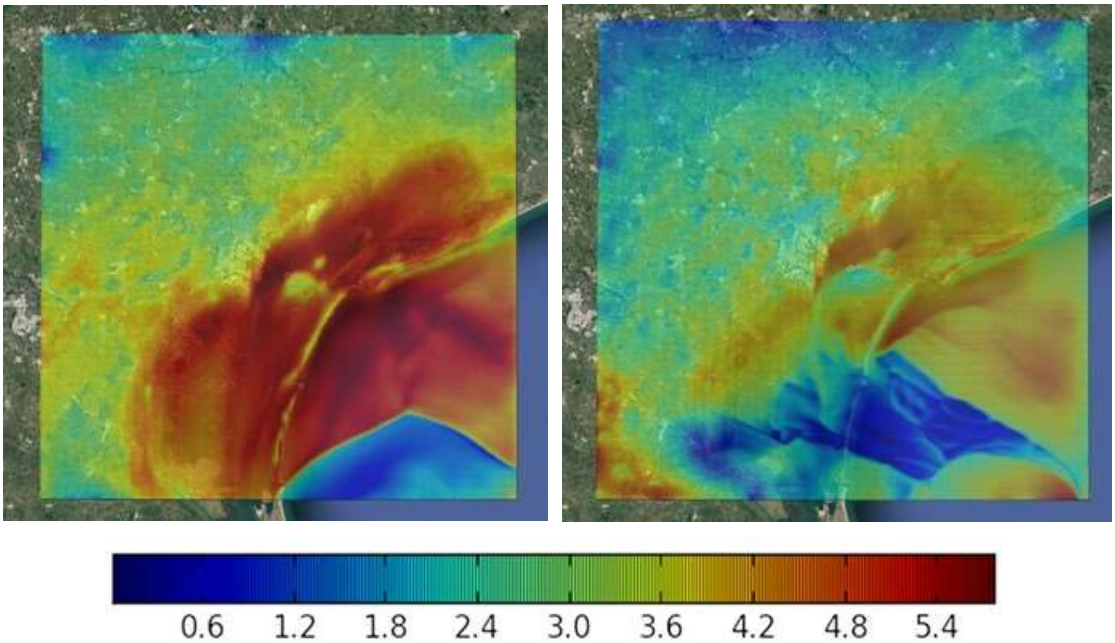

Figure 12: Wind fields (m/s) at approximately 10 m height above ground as regards the Venice fine domain originating from the initial MEMO calculations.

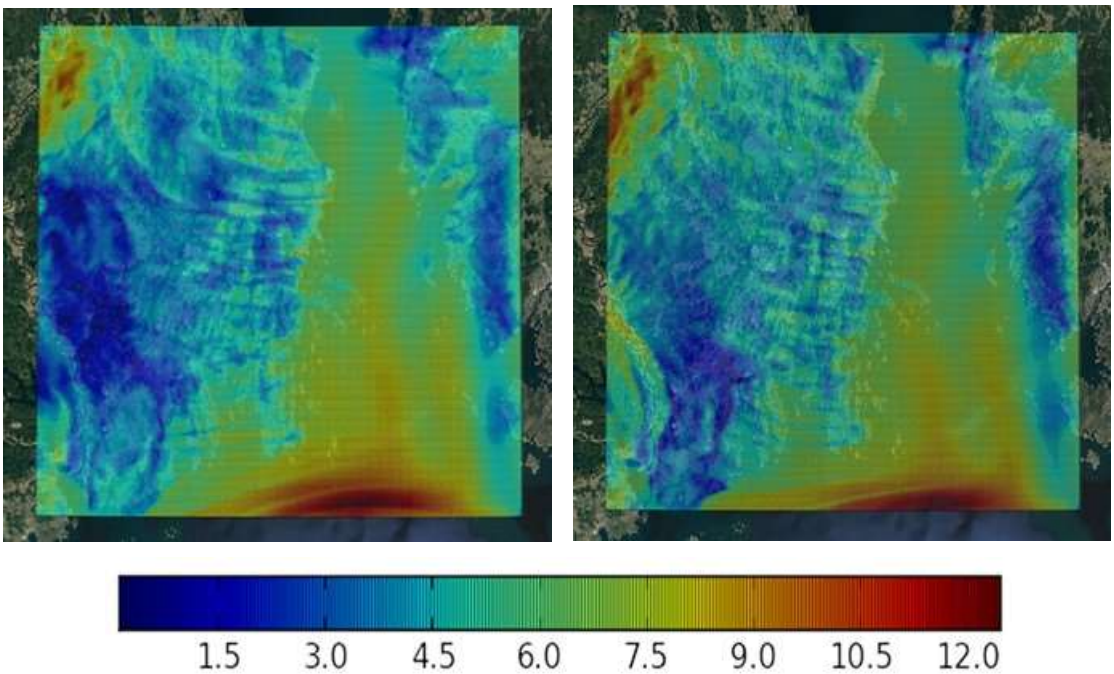

Figure 13: Wind fields (m/s) at approximately 10 m height above ground as regards the Tonsberg fine domain originating from the initial MEMO calculations.

4.2 Validation of Pilot Mesoscale Results

A pilot validation process was conducted for the four areas of interest. Regarding the availability of model and measurement data, the validation process refers to a two month period for the case studies of Venice and Tonsberg and for a two week period for the case studies of Granada and Rhodes. Table 3 presents the outcome of this evaluation regarding Granada, Rhodes, Venice and Tonsberg, respectively. More specifically, the indicators used in this application comprise of the Bias (BIAS), the

Normalized Mean Square Error (NMSE) and the Index of Agreement (IOA). Model Bias is the mean error that is defined as the predicted value (P_i) less than the observed value (O_i). It is given by:

$$BIAS = \frac{1}{N}\sum_i (P_i - O_i)$$

Regarding the NMSE, this indicator emphasizes on the scatter in the entire data set. The normalization by the product $P_i * O_i$ assures that the NMSE is not biased towards models that over predict or under predict. Smaller values of NMSE denote better model performance. The expression for the NMSE is given by:

$$NMSE = \frac{1}{N}\sum_i \frac{(P_i - O_i)^2}{\overline{PO}}$$

The Index of Agreement (IoA) developed by Willmott (1981) as a standardized measure of the degree of model prediction error and varies between 0 and 1. A value of 1 indicates a perfect match, and 0 indicates no agreement at all (Willmott, 1981). The index of agreement can detect additive and proportional differences in the observed and simulated means and variances; however, it is overly sensitive to extreme values due to the squared differences (Legates and McCabe, 1999).

$$IoA = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}[|(O_i - \overline{P})| + |(P_i - \overline{O})|]^2}$$

Table 3: Validation of the mesoscale results for the four pilot domains

Venice			
Statistical Indicator	**Bias**	**NMSE**	**IOA**
Temperature	-0,22	0,05	0,85
Relative Humidity	-25,78	0,15	0,71
Wind Speed	-2,34	0,43	0,94
Rhodes			
Statistical Indicator	**Bias**	**NMSE**	**IOA**
Temperature	-0,67	0,03	0,55
Relative Humidity	-10,66	0,01	0,92
Wind Speed	2,00	0,89	0,22
Tonsberg			
Statistical Indicator	**Bias**	**NMSE**	**IOA**
Temperature	1,76	0,15	0,94
Relative Humidity	-22,21	0,25	0,72
Wind Speed	-3,57	0,20	0,96
Granada			
Statistical Indicator	**Bias**	**NMSE**	**IOA**
Temperature	-0,94	0,04	0,86
Relative Humidity	-22,96	0,13	0,82
Wind Speed	-1,14	0,04	0,98

As depicted from Table 3, the results of the mesoscale model for temperature are in a fairly good agreement with the corresponding measurements in most cases, especially with respect to the Index of Agreement indicator. On the other hand, it is clear that the model systematically underestimates relative humidity. Besides, some differences between model results and observational data are also observed regarding wind speed. This is something, more or less, expected due to the fact that the spatial resolution of 250 meters used in the model results is relatively low with respect to micro-climate effects occurring in the particular sites under investigation, due to the continuous and abrupt changes of topography and land use.

Based on the above, the introduction of the data assimilation approach in the core of the OMS is expected to provide a notable upgrade in the outcome of the computational procedure.

4.3 Pilot Application of the Data Assimilation Approach

A pilot application of ten days was conducted for the four areas of interest in order to depict the potential corrections introduced in the OMS's results by the incorporation of the data fusion and assimilation methodologies in its computational core. Due to the fact that kriging techniques preserve the values of the initial samples in the interpolated maps and taking into consideration that in this phase of the project there are available data for only one weather station for each city, we could not perform a validation process similar to that presented in the previous section.

More specifically, a subset of the available meaurements were used to feed the data assimilation computational infrastructure, which, subsequently, based on the optimal interpolation approach, performed the corresponding calculations.

Figures 13-20 depict that there is a clear improvement in the corrected simulations (DA) compared to the initial mesoscale calculations. This is something, more or less, expected due to the fact that the spatial resolution of 250 meters used in the initial model results is relatively low with respect to micro-climate effects occurring in tareas with complex terrains as those under investigation.

On the other hand, the introduction of observational data in the computational procedure of the OMS via the optimal interpolation approach gives the system the ability to perform simulations of high accuracy not only in the mesoscale, but also in the microscale.

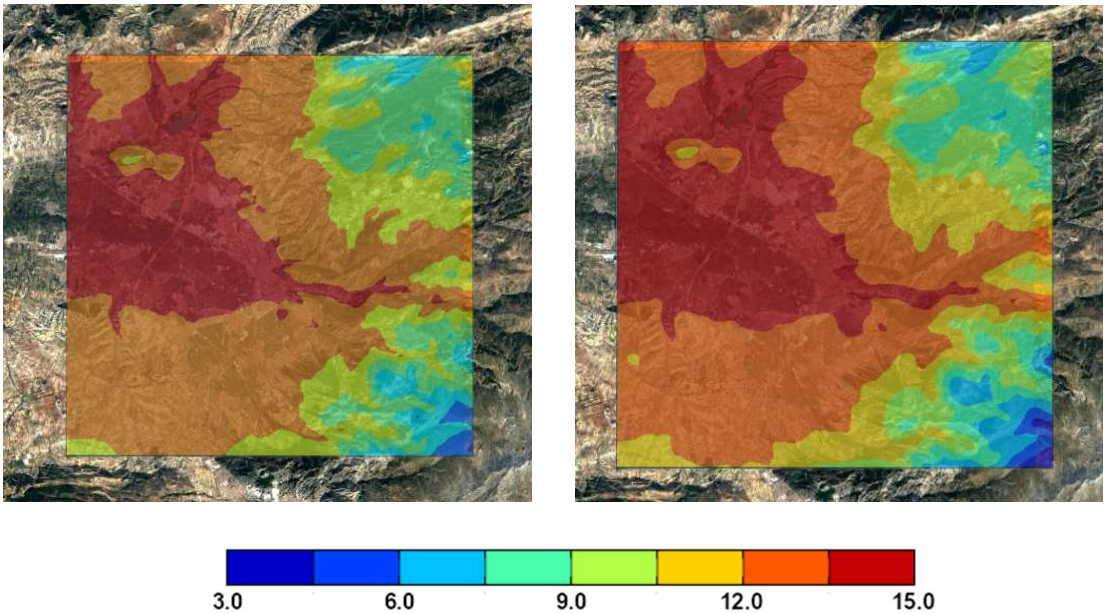

Figure 13: Visualization of the average temperature field (^{0}C) at 2 m height above ground regarding the period of the pilot application for the Granada fine domain originating from the initial MEMO simulations (left) and after the application of the data assimilation process (right).

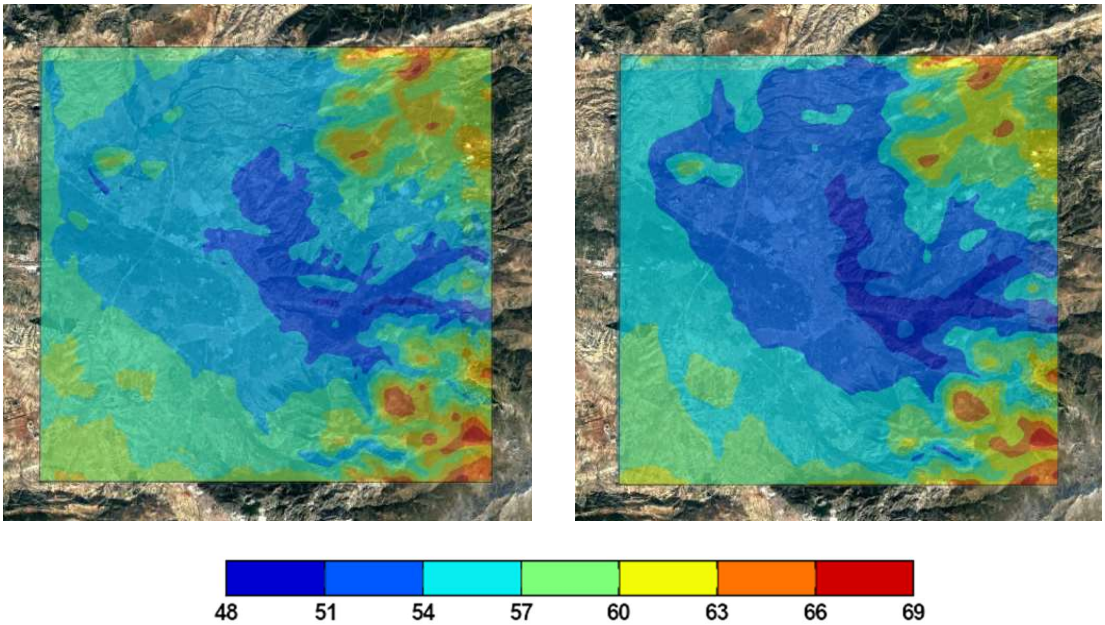

Figure 14: Visualization of the average relative humidity field ($^{\%}$) at 2 m height above ground regarding the period of the pilot application for the Granada fine domain originating from the initial MEMO simulations (left) and after the application of the data assimilation process (right).

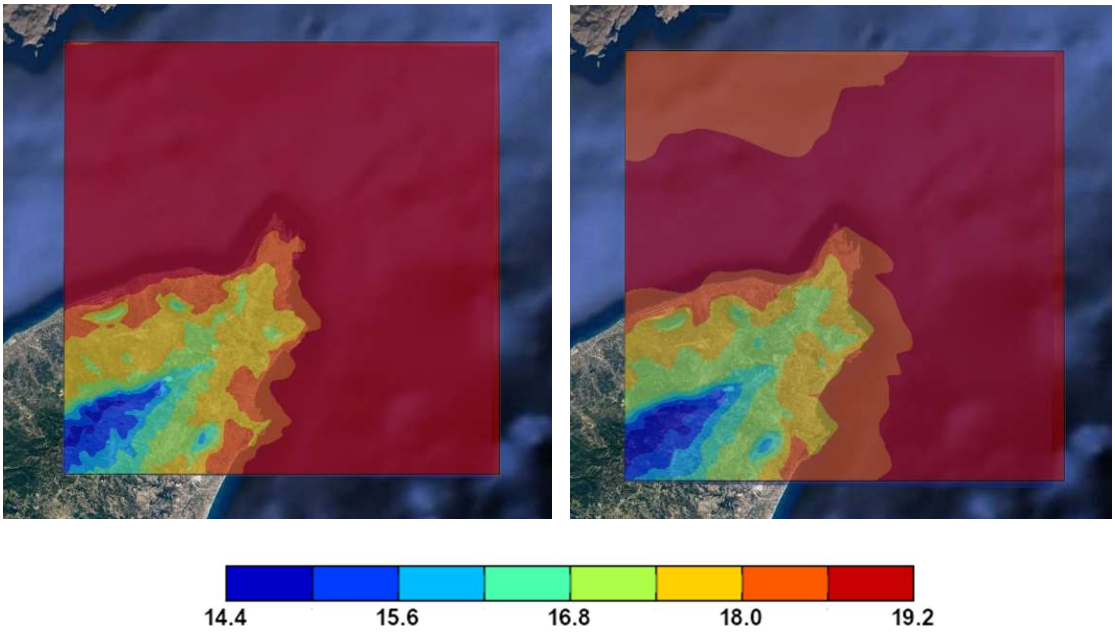

Figure 15: Visualization of the average temperature field (^{0}C) at 2 m height above ground regarding the period of the pilot application for the Rhodes fine domain originating from the initial MEMO simulations (left) and after the application of the data assimilation process (right).

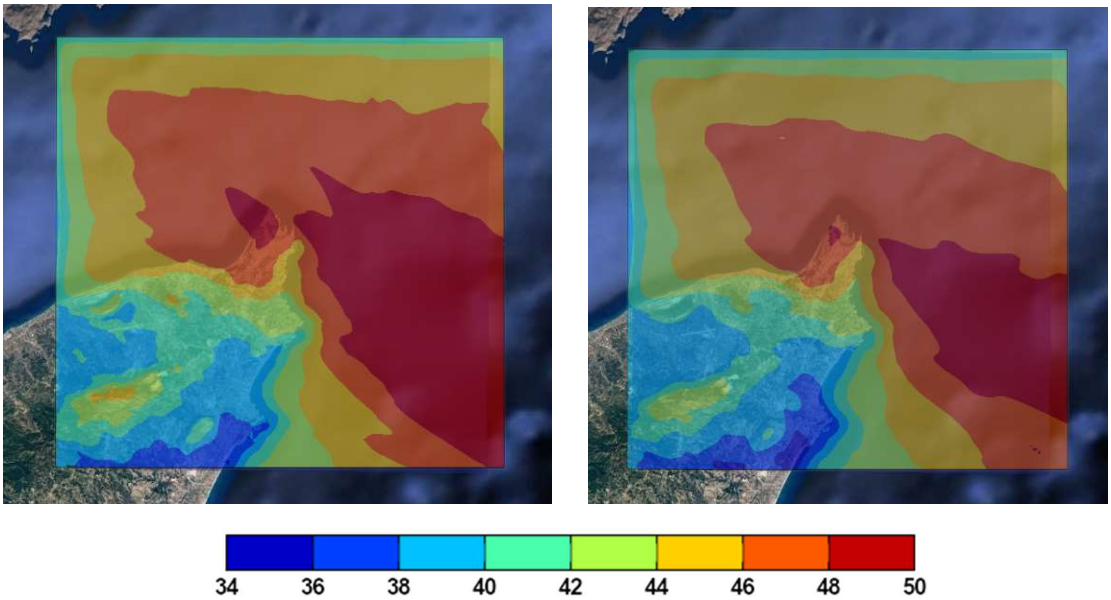

Figure 16: Visualization of the average relative humidity field ($^{\%}$) at 2 m height above ground regarding the period of the pilot application for the Rhodes fine domain originating from the initial MEMO simulations (left) and after the application of the data assimilation process (right).

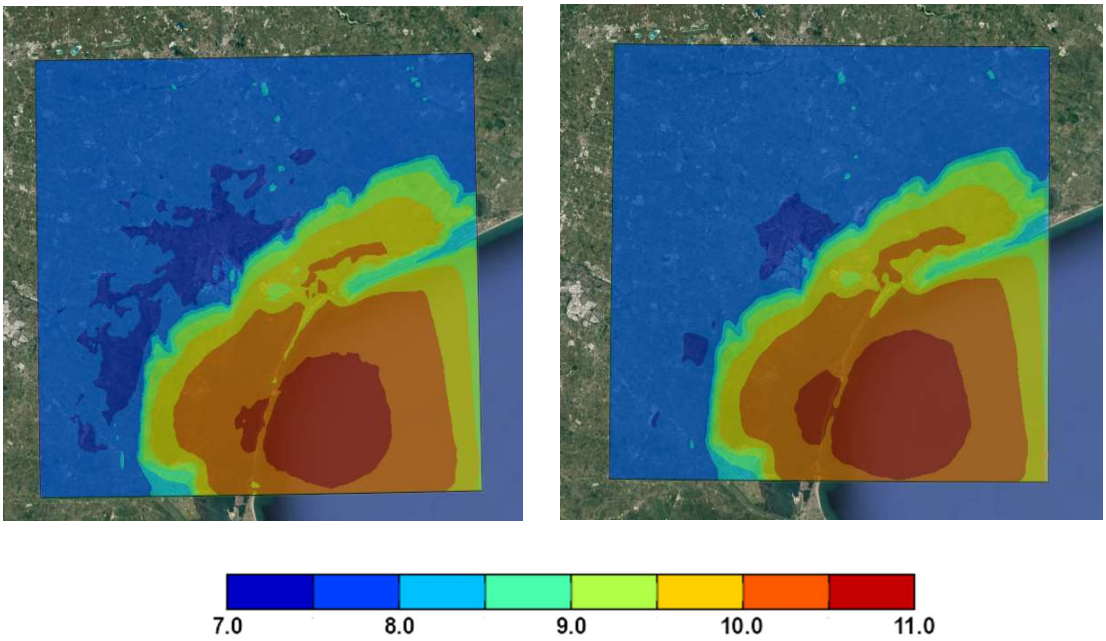

Figure 17: Visualization of the average temperature field (^{0}C) at 2 m height above ground regarding the period of the pilot application for the Venice domain originating from the initial MEMO simulations (left) and after the application of the data assimilation process (right).

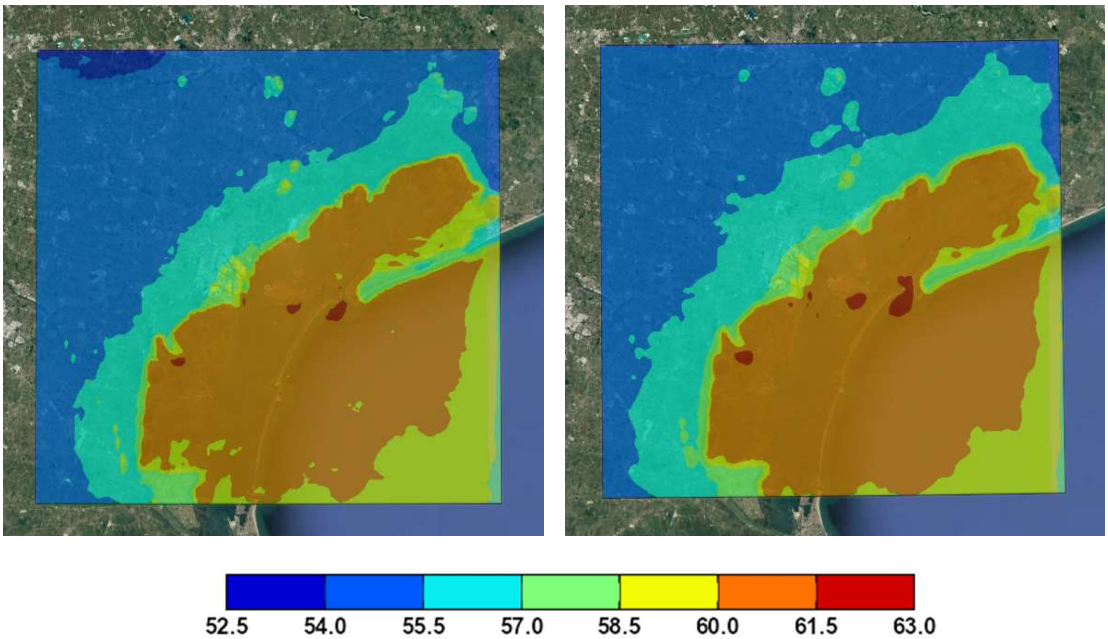

Figure 18: Visualization of the average relative humidity field (%) at 2 m height above ground regarding the period of the pilot application for the Venice domain originating from the initial MEMO simulations (left) and after the application of the data assimilation process (right).

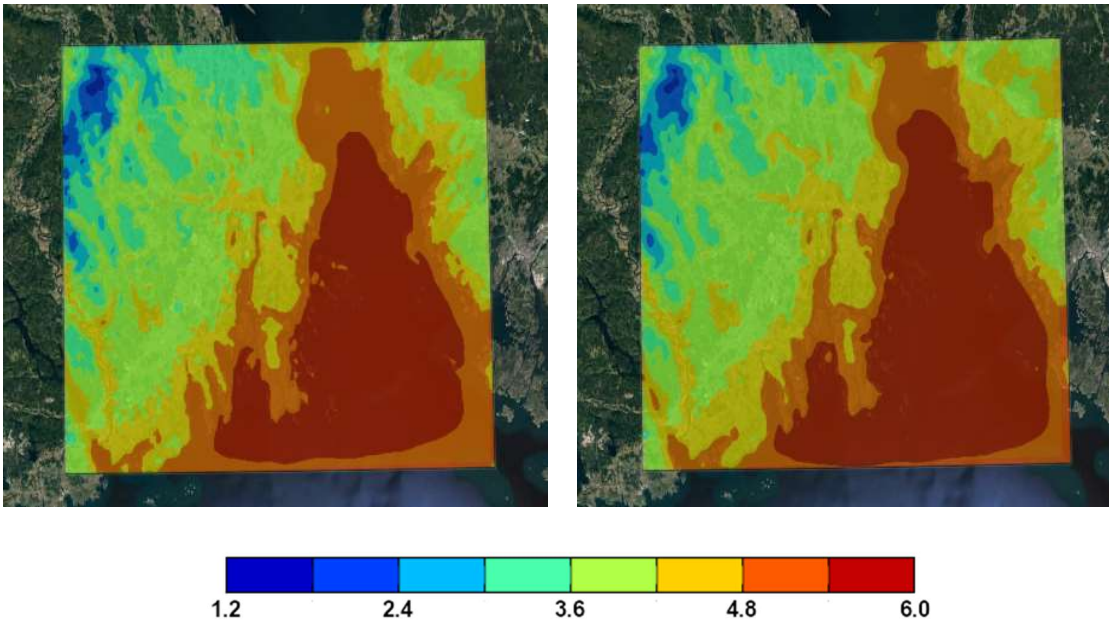

Figure 19: Visualization of the average temperature field (^{0}C) at 2 m height above ground regarding the period of the pilot application for the Tonsberg domain originating from the initial MEMO simulations (left) and after the application of the data assimilation process (right).

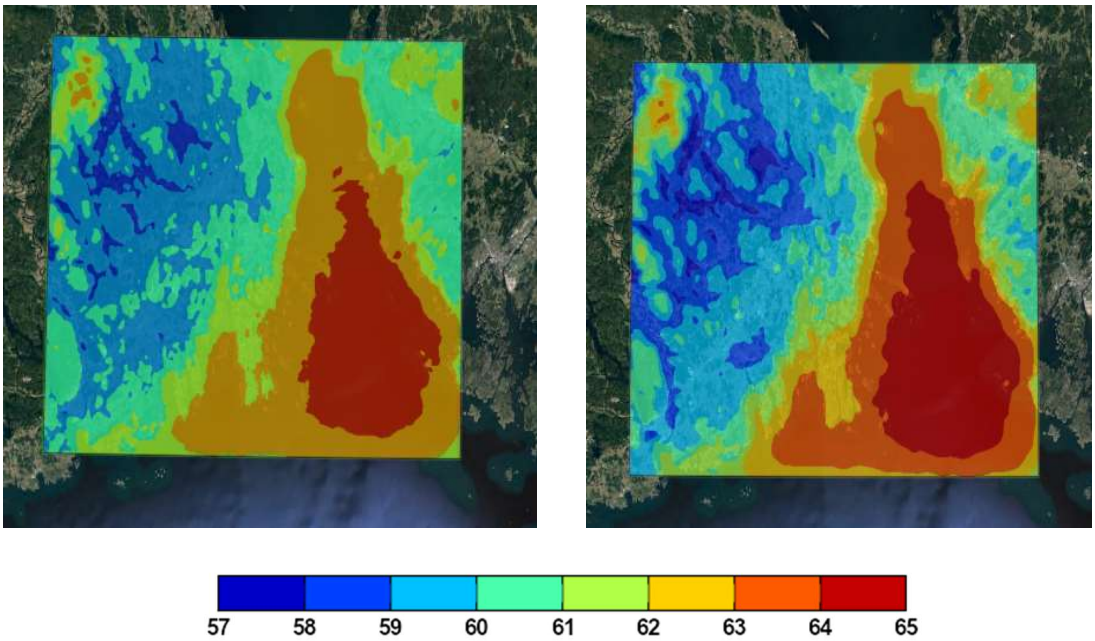

Figure 20: Visualization of the average relative humidity field ($^{\%}$) at 2 m height above ground regarding the period of the pilot application for the Tonsberg domain originating from the initial MEMO simulations (left) and after the application of the data assimilation process (right).

5. Conclusions

The purpose of this Task lies in employing an innovative computational procedure, capable of producing revised meteorological maps as regards a number of parameters

of interest. This set of parameters contains the ambient temperature, the relative humidity, wind velocity (wind speed and direction) and precipitation for HYPERION pilot areas in Spain (Granada), Greece (Rhodes), Italy (Venice) and Norway (Tonsberg) using data fusion and assimilation methods.

The core of the methodology is based on a 3-D data assimilation protocol which supports fusion of data-streams originating from micro weather stations installed at the selected demonstration pilots in Task 3.4. Within the present task, a data assimilation layer is realized to assure the proper transformation of real data into simulation parameters feeding back the modelling engine and contributing to more accurate and dynamically updated predictions.

Moreover, aiming at the evaluation of the system's performance, a structural modification on the core of the OMS has been performed. As a result, tools for validating the meteorological calculations have been implemented, including the automatic calculation and visualisation of a number of statistical indicators for the locations of the micro-climate stations in the computational domains. In this direction, suitable statistical indicators are estimated at the end of each day. Indicators can be calculated both for the initially simulated by MEMO model fields, as well as the corrected fields by the application of the data assimilation methodology, against measurement data series if there is a sufficient number of point measurements.

An automated procedure operating on an around-the-clock basis undertakes the downloading of the available observation data and the subsequent processing and storage in the main OMS data base. At the same time, model results calculated for the same locations are registered in a parallel data pool and are archived for later inspection by the user. At the end of each day, statistical indicators are calculated according to the guidelines set in COST728 (2008), for the station locations and pollutants of interest, and numerous charts are produced for visually assessing the accuracy of the simulations in both nowcasting and forecasting mode.

In order to assess the improvement introduced in the OMS's results by the integration of the data assimilation methodology in its computational core, a pilot application was conducted for a period of ten days. The results of this application are expected to introduce a clear upgrading in the accuracy of the produces simulations, which can be directly attributed to the fact that this newly developed function of the OMS enhances its performance in the microscale.

6. REFERENCES

Beelen R., G. Hoek, E. Pebesma, D. Vienneau, K. de Hoogh, D. J. Briggs (2009) Mapping of background air pollution at a fine spatial scale across the European Union Science of The Total Environment, Volume 407, Issue 6, Pages 1852-1867.

COST728, 2008: Overview of Tools and Methods for Meteorological and Air Pollution Mesoscale Model Evaluation and User Training.

Denby B., M. Schaap, A. Segers, P. Builtjes and J. Horálek (2008). Comparison of two data assimilation methods for assessing PM10 exceedances on the European scale. Atmos. Environ. 42, 7122-7134.

Denby, B., V. Garcia, D. M. Holland, and C. Hogrefe. Integration of Air Quality Modeling and Monitoring Data for Enhanced Health Exposure Assessment. EM: AIR AND WASTE MANAGEMENT ASSOCIATION'S MAGAZINE FOR ENVIRONMENTAL MANAGERS. Air & Waste Management Association, Pittsburgh, PA, (10/2009):46-49, (2009).

Elbern, H. and H. Schmidt (1999): A four-dimensional variational chemistry data assimilation scheme for eulerian chemistry transport modeling. JGR, 104, 18583-18598.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Applied Geostatistics Series. xiv + 483 pp. New York, Oxford: Oxford University Press.

Isaaks, EH, and RM Srivastava. 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York.

Moussiopoulos N., Douros I., Tsegas G., Kleanthous S. and Chourdakis E., 2012. An air quality management system for policy support in Cyprus, Hindawi Publishing Corporation, Advances in Meteorology 2012, doi:10.1155/2012/959280.

N. Moussiopoulos, “The EUMAC Zooming Model, a tool for local-to-regional air quality studies,” *Meteorology and Atmospheric Physics*, vol. 57, no. 1–4, pp. 115–133, 1995.

Sahu, S., Yip, S., and Holland, D. M. (2009). Improved space-time forecasting of next day ozone concentrations in the eastern U.S.. Atmospheric Environment, 43, 494–501.

van Loon, M, P.J.H. Builtjes, A. Segers, 2000. Data assimilation of ozone in the atmospheric transport chemistry model LOTOS. Environmental Modeling and Software 15, 603-609.

Webster, R, and MA Oliver. 1993. How Large a Sample Is Needed to Estimate the Regional Variogram Adequately?. Geostatistics Troia '92, ed. A Soares, Vol 1, pp. 155-66. Kluwer Academic Publishers, Dordrecht.

Willmott, C. J. 1981. On the validation of models. Physical Geography, 2, 184–194

Zängl, G., D. Reinert, P. Ripodas, and M. Baldauf, (2015): The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI M: Description of the non-hydrostatic dynamical core.Q.J.R.Meteorol Soc.,141, pp. 563.

URL1: http://www2.jpl.nasa.gov/srtm/

URL2: http://www.eea.europa.eu/data-and-maps/data/corine-land-cover