Report on the hierarchy of benchmarks
Deliverable D3.4

Edited by: Javier Sanz Rodrigo
Delivery date: 29 March 2016
Dissemination level: Public

With financial support from NEWA ERA-NET Plus, topic FP7-ENERGY.2013.10.1.2.

Author Information

Name	Organisation	E-mail
Javier Sanz Rodrigo	CENER	jsrodrigo@cener.com
Roberto Chávez	CENER	
Iván Moya	CENER	
Fidel González Rouco	UCM	
Sergio Fernández	UCM	
Andrea Hahmann	DTU	
Albert Soret	BSC	
Matias Avila	BSC	
Stefan Ivanell	UU	
Björn Witha	ForWind	
Julia Gottschall	IWES	
Chi-Yao Chang	IWES	
José L. Palma	UPorto	
Sibel Menteş	ITU	
Alexis Dutreux	ATM-Pro	
Arnaud Candaele	CENAERO	
Stefan Ivanell	UU	
Rory Donnelly	3E	
Johan Meyers	KUL	

Table of Contents

Executive Summary

This report describes the validation directed program that will guide the planning of experiments and model development activities of NEWA.

The program follows a formal verification and validation (V&V) framework originally developed by Sandia National Laboratory. It has two phases: the integrated planning and the experimental and modeling planning and execution. This report deals with the first phase where the aim is the identification of the relevant phenomena of the model-chain that will be necessary to meet the application objectives, and the hierarchy of validation exercises (also called benchmarks) that will be done in order to demonstrate how the model-chain actually integrates those phenomena.

The benchmarks hierarchy is defined in terms of three dimensions: the observational dataset from which the validation data is extracted, the part of the model-chain that is being addressed and the validation objectives. These objectives are related to the phenomena of interest for the given application, in this case, with the development of a mesoscale to microscale model-chain for wind conditions assessment.

The result is a planning instrument that will provide guidance to the data gathering activities in WP2, and help with the coordination of the work within WP3 to deliver a validated methodology for the production of the New European Wind Atlas and associated modeling tools. This initial planning is subject to change depending on the success of the observational data collection from the experiments and the call for wind data.

Introduction

The development of the New European Wind Atlas (NEWA) is based on the validation directed research program outlined in this document. The purpose of this program is to provide guidance on the development and execution of a highly integrated modeling and experimental research activity based on well-established verification and validation (V&V) practices adapted to the scope of the project and constrained by the available resources.

The formal V&V framework adopted here comes from Sandia National Laboratories. A recent review (Hills et al., 2015) has been published in the frame of the Atmosphere to Electrons (A2e) wind energy research program, based on existing V&V methodologies developed by various American organizations including DoE, NASA, AIAA and ASME[1] (AIAA, 1998; ASME, 2009; Oberkampf et al., 2007; Pitch et al., 2001; Trucano et al., 2002). The framework is also adopted in the frame of the IEA Task 31 Wakebench to establish a model evaluation protocol for wind farm flow models (Sanz Rodrigo and Moriarty, 2015). An overview of the essential aspects of this framework is provided here and applied to the NEWA application scope.

1. Wind Atlas Scope

A wind atlas is associated to the planning phase of wind energy development, which can last several years from strategic spatial planning, to site prospecting, to wind farm design and financing. Detailed and robust information about the relative size of the wind resource across an area is crucial for the commercial evaluation of a wind farm.

Today a number of well-established models and methodologies exist for estimating resources and design parameters. These can work well if good local data are available, but the wind energy community is still hampered by projects having large negative discrepancies between calculated and actual resources and design conditions.

A main objective of the NEWA project is to fundamentally change the state-of-the-art during the course of the project by developing and introducing a new methodology for the assessment of wind conditions based on a mesoscale to microscale model-chain approach. A generally approved method is highly needed so that data generated by the mesoscale model can be adapted and collated for use in various microscale models.

The New European Wind Atlas will provide a unified high-resolution and freely available dataset of wind resource in Europe. Wind statistics will cover onshore Europe and 100 km offshore plus the Baltic and the North Seas (Figure 1). The database will be based on at least 10 years of mesoscale simulations at 3 km resolution, with long-term corrections as well as subgrid microscale corrections to reduce the bias on the local mean wind resource.

---

[1] Department of Energy (DoE), National Aeronautics and Space Administration (NASA), American Institute of Aeronautics and Astronautics (AIAA), American Society of Mechanical Engineers (ASME)

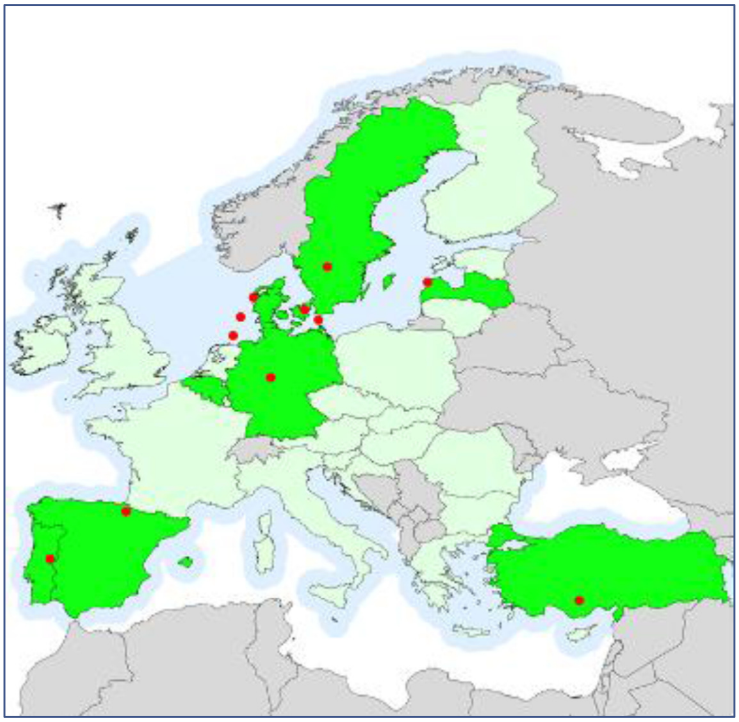

Figure 1: Initial extension of the European domain for the New European Wind Atlas and location of high fidelity experiments

In addition to wind resource information, the new wind atlas will provide information about site suitability conditions (turbulence intensity, wind shear, extreme wind speed), wind variability as well as wind power predictability from day-ahead to decadal. A probabilistic wind atlas methodology using mesoscale models is introduced in D3.1, and methodologies for predictability assessment are introduced in D3.2.

Besides variables of immediate use by resource planners, the wind atlas will provide means to feed boundary conditions on microscale models. This will allow not only to improve the wind atlas predictions at local level when better site data becomes available but also to allow a coherent integration with wind farm design tools. Hence, a generalized wind atlas, i.e. free of site effects, will be also part of the NEWA database. Downscaling methodologies with microscale models are introduced in D3.3.

Integral to the wind atlas methodology is the assessment of the associated uncertainties. The ultimate goal of the wind atlas is to reduce the uncertainties on the assessment of wind resource and the wind conditions that affect the design of wind turbines. To this end, the model-chain will be thoroughly validated across Europe with dedicated experiments and historical wind resource assessment campaigns from industry. Data collection from industry is organized within a Call for Wind Data (CfWD) during the first two years of the project.

An uncertainty map will calculate the confidence of the wind atlas and, therefore, the intensity to which in situ measurement must be employed before development of a wind farm.

2. Modelling Scope

Figure 2 shows schematically the wind assessment model-chain framework with typical scale ranges for each sub-model level and associated applications and flow modeling approaches of various physical fidelity levels (Sanz Rodrigo et al, 2016).

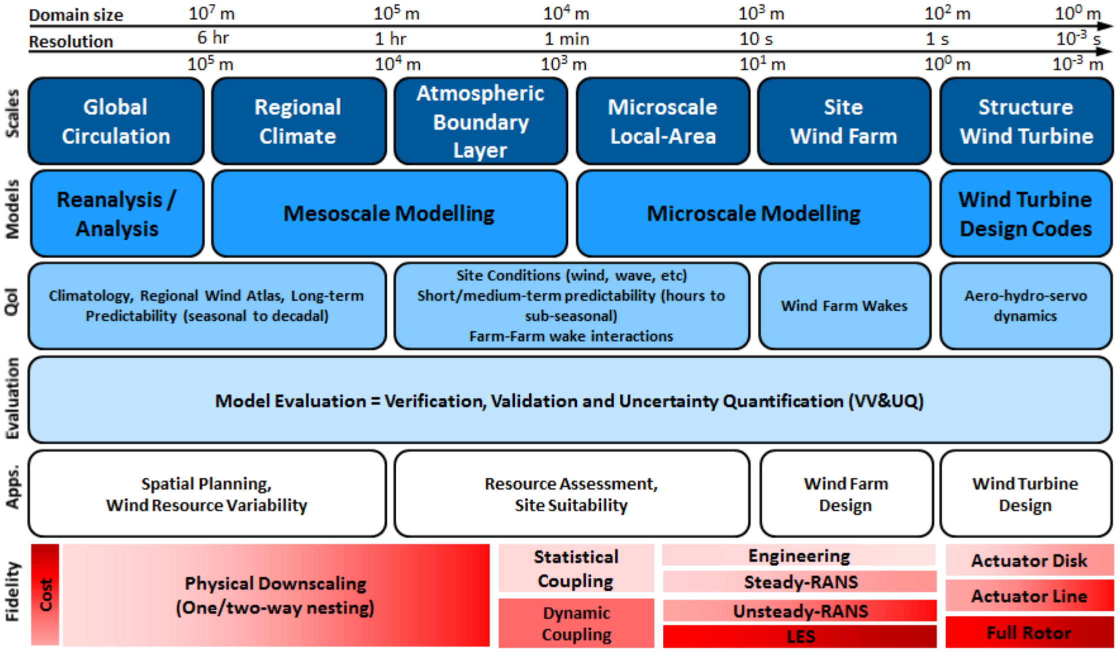

Figure 2: Wind assessment modelling framework indicating typical model scale ranges, relevant quantities of interest for different applications and high-level fidelity levels (the shading indicates the computational cost). (Sanz Rodrigo et al., 2016)

The NEWA methodology will be based on a mesoscale to microscale model-chain that can be used for wind resource assessment and site suitability. Hence, the scope will be in principle limited to the characterization of external wind conditions (Petersen and Troen, 2012), leaving wind turbine and wind farm model assessment out of the NEWA program.

3. V&V Framework for Wind Conditions Assessment

Conducting a complete full-system validation for wind energy is not possible due to the inherent complexity of the operational conditions of the physical environment. Due to the complexity of this multi-scale modeling system, the objective of the experiments is to support model validation of the physical phenomena that has the largest impact on improving the predictive capacity of wind energy design tools. Predictive capacity implies that the model will be typically used outside of the validation envelope. Then, a hierarchy of experiments and validation cases should be defined in order to test the phenomena for the application of interest.

A formal model validation directed program applied to wind conditions assessment has the objective of developing and executing a collaborative experimental and numerical research activity, which will lead to quantifying the predictive capability of state-of-the-art models for this application.

Validation is defined in the AIAA (1998) as the process of determining the degree to which the model is an accurate representation of the real world from the perspective of the intended uses of the model. Here accuracy is measured with respect to observational data, out of a dedicated experiment, with the objective of providing evidence of the model suitability.

In contrast, verification is the process of determining that the model implementation accurately represents the developer's conceptual description of the model and the solution of the model. Here accuracy is measured with respect to benchmark solutions of simplified model problems obtained from theory or from high-fidelity models.

The application determines the range of conditions for which the model is to be evaluated. For example, validation results should be particularized to quantities of interest (QoI) for wind resource assessment, at heights above the terrain of the order of 100 m, at well exposed sites and over the operating range of wind turbines, i.e. above 4 m/s and below 25 m/s. Other quantities and operating conditions may be useful to diagnose knowledge gaps, for example analyzing surface-layer turbulent fluxes to characterize boundary conditions.

The planning process is shown in the top panel of Figure 3, extracted from Hills et al. (2015). It is composed of four steps:

1. Identify the objectives of the model from the perspective of the intended use (application) in terms of quantities of interest and the impact on the application
2. Identify the phenomena of interest that the model should capture and prioritize the assessment based on the expected impact on the objectives
3. Define a validation hierarchy that will allow to assess model performance for the prioritized phenomena
4. Plan experiments to generate data for the validation hierarchy based on how the limited resources can be used most effectively

The planning document resulting from this analysis is this deliverable. The planning should be revisited along the NEWA project and adapted to include the outcome of each experiment and validation campaign.

The lower panel of Figure 3 shows the process of experiment design, execution and validation activities that lead to the model assessment. The credibility step in the end determines, by expert judgment, to what extent the verification and validation results will improve the predictive capacity in the operational conditions of the model.

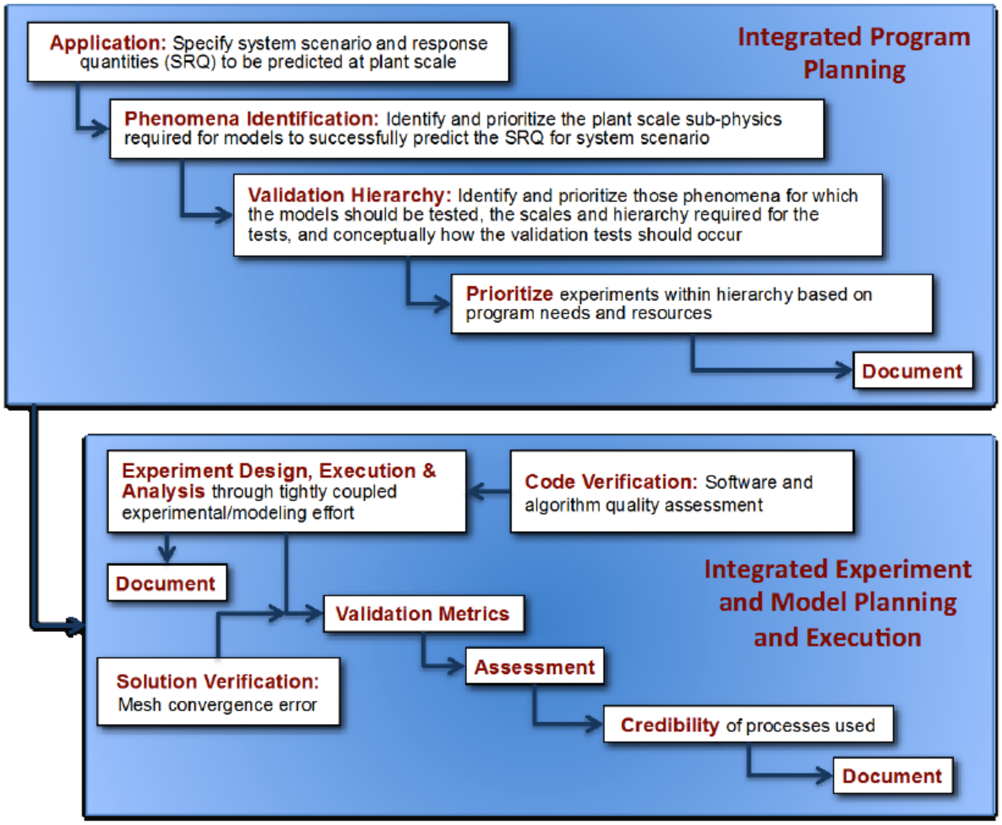

Figure 3: Validated directed program planning and execution (Hills et al., 2015)

4. Quantities of Interest

Quantities of interest during the wind energy planning phase are mainly: long-term annual energy production (AEP); air density; long-term mean wind speed at hub-height; 50-year recurrent 10-min-averaged extreme wind speed (V_{ref}); ambient turbulence intensity at hub-height; effective (wake added) turbulence intensity at hub-height; inflow angle (angle off the horizontal plane at which the mean wind flow comes into the rotor) at hub-height; wind shear (vertical velocity gradient) and wind veer (vertical wind direction gradient) across the rotor span. These variables are required for energy yield assessment and site suitability studies following the standards of the International Electrotechnical Commission (IEC) regarding wind turbine design aspects (IEC, 2005).

The operational phase covers the lifetime of the wind farm which is typically assumed to extend for 20 years. Relevant variables in this phase are mainly the weather conditions and the wind farm power forecasts at different look-ahead times in connection to wind farm and transmission system operation and maintenance and energy trading. When these variables are characterized during the planning phase, in order to anticipate the associated operational and financial costs, we talk about wind predictability assessment.

Project financing will assess the risk associated to the investment uncertainty, defined in terms of percentiles (p_{50}, p_{95} are often used) or exceedance probabilities of the wind farm's AEP. The financial cost due to uncertainty can be quite important. Hale (2015) provides a couple of examples: for a 200 MW project, a 3% difference on the AEP p_{50} (bias) means $17MM difference in the net project value; a 1.5% difference on p_{95} (uncertainty) results in $1.5MM difference on the net project value. These deviations are representative of the variability on AEP estimates in industry and show how sensitive project financing is to relatively small changes on wind resource assessment.

5. Phenomena Identification and Ranking Table

Once defined the needs for the application of interest, the next step towards improving model credibility is to identify the phenomena of interest that needs to be captured by the model to meet these needs. The Phenomena Identification and Ranking Table (PIRT) is a well-established tool to:

- rank these physical and other related phenomena for the intended use,
- characterize the adequacy of the existing model, experiment and validation data and
- provide a gap analysis as to what are the issues associated to the modeling of these phenomena and how they can be addressed.

Through expert elicitation, it is determined if a model has sufficient evidence to be used for the intended application and, if not, how to efficiently prioritize phenomena of interest that are expected to maximally improve model credibility within the available resources.

5.1 Phenomena of Interest

As a planning tool, the PIRT table will support the decision making process by providing a structured method to link end-user objectives and priorities for model development and experiment design. Expert elicitation and gap analysis are methods to identify needs and determine which phenomena of interest should be prioritized conditioned by the research program focus and resources.

Hills et al. (2015) provide a table with examples of phenomena of interest that could be considered for inclusion in PIRT.

Table 1 is a reproduction of this table adapted to the wind assessment context (the research program focus). Examples are provided in terms of some questions that highlight gaps in the wind assessment process. The phenomena identification is accompanied by the associated issues (what the problem is) and the potential responses, i.e. what actions need to be taken to remove the issues and move the phenomenon down the list of priorities.

Table 1: Examples of phenomena for inclusion in PIRT in the NEWA context (adapted from Hills at al., 2015)

Type	Issues	Potential Responses	Example
Physics	Important physics inadequately represented or missing	Model development or experimental characterization to better represent the phenomena Model validation to assess the uncertainty associated with the lack of physics	Is steady state model sufficient for AEP assessment? How important is to assume a dry atmosphere?
	Not clear if important phenomena, or interactions between phenomena, are adequately represented by model	Model validation to incorporate the effect of the phenomena	I have included a potential temperature equation in my microscale model and buoyancy terms in momentum and turbulence. Is it sufficient to characterize non-neutral conditions?
	Ranking of phenomena not clear	Sensitivity analysis to rank importance for the quantities of interest	I have validated my non-neutral ABL solver. What is the impact of these new phenomena in AEP assessment?
Model and Geometric Fidelity	Sub-components poorly represented	Sensitivity analysis of subsystem level with higher fidelity model to assess impact of underrepresented components	I have both LES and RANS models implemented in my ABL solver. Is RANS a good enough turbulence model for site suitability so I can save some computing resources?
	Geometric fidelity and/or grid resolution insufficient to capture behavior	Sensitivity analysis of subsystem level with higher fidelity model to assess impact of under-resolved geometry Grid studies (solution verification) to characterize uncertainty due to grid dependencies	Is the actuator-disk rotor model good enough for array efficiency assessment? How fine is fine enough when meshing complex terrain? How far and detailed should I model upstream terrain?
Characteri-zation	Inadequate inputs (inflow, boundary conditions, site) characterization	Refine characterization to the required fidelity using experimental techniques or other techniques	Is the resolution of the topographic database good enough? What is the bias of the forcing from my global/mesoscale model?
	Inadequate parameter characterization	Characterize based on literature or experimental data	Can I use the same set of turbulence constants for any simulation or should I characterize site specific ones based on measurements?
Uncertainty Quantifi-cation	Uncertainty in model prediction not adequately characterized due to large number of runs	Approximate methods such as surrogate model or other advanced UQ methods to reduce the number of runs	Can I use ad-hoc UQ engineering methods based on standard practices or should I do a formal UQ assessment for each site?

5.2 Gap Analysis

Gap analysis is done to determine the adequacy level of the model to be able to represent the phenomena of interest. A traffic light color code is used to indicate three levels of adequacy: red (low), yellow (medium) and green (high) (Table 2).

Table 2: Guidelines to determining adequacy level during gap analysis (adapted from Hills et al., 2015)

	Physics	Code	Validation
High	Mature model that can represent the phenomenon over the full operational range	Intended model implemented with code and solution verification conducted and documented on relevant benchmarks	Comprehensive validation evidence relevant to the application
Medium	Medium-fidelity model in general captures relevant phenomenon	Intended model implemented but verification is not complete or not dealing with relevant benchmarks	Partial validation, not covering the operational range in sufficient extent
Low	Reduced-order model with poor representation of phenomenon	Intended model not yet implemented with incomplete verification showing potential bugs or issues that prevent from usage	Insufficient validation or not relevant for the application

Each expert involved in the elicitation process will provide an assessment of the adequacy levels for each phenomena of interest. Depending on the ensemble opinions, each phenomenon will be provided an importance ranking:

Table 3: Guidelines to determining the importance ranking level (adapted from Hills et al., 2015)

	Assessment	Meaning
High	Model, code and validation adequacy should be at the high level.	Critical phenomena with large impact
Medium	Model, code and validation adequacy should be at least at the medium level	Second-order importance
Low	High consensus with supporting evidence that these physics are not relevant for the application (validation is at least medium)	The model does not need to reproduce this phenomenon
Uncertain	Not yet assessed due to lack of supporting evidence (validation is low)	Potentially important phenomena to be analyzed through sensitivity analysis or validation experiments to provide a ranking

5.3 Expert Elicitation for the Wind Conditions Assessment PIRT

The PIRT is largely based on subject matter expert (SME) elicitation to decide priorities and phenomena of interest by consensus. The PIRT development team should comprise end-users, modelers, experimentalists and V&V specialists.

An “expert” in the NEWA context is considered someone with good background knowledge on wind resource assessment and site suitability applications, broad knowledge on the experimental and numerical characterization of wind conditions and advanced knowledge on some modeling and/or experimental techniques. This specialization has to be clearly stated in order to be able to weight opinions in the elicitation process. This will allow to arrive to a consensus based on an interdisciplinary team of experts with different types of expert knowledge.

Following an iterative process, as in the Delphi method (Helmer, 1968), experts are first provided with the task of identifying the most important phenomena of interest that should be considered to meet the wind assessment objectives stated in section 1 of this document to target the quantities of interest of section 4.

Table 1 can be used to guide the selection of appropriate phenomena. To facilitate this process, the phenomena should be classified in terms of the dominant physical scale as per Figure 2: global, mesoscale (regional), meso-micro, microscale (site level). The interaction between wind conditions and wind turbines is left out of the scope of NEWA.

Consistent with the philosophy of NEWA, experts should not limit their opinions to their area of specialization but also address the meso-micro range, as core development area of the project. Hence, mesoscalers should consider how to improve subgrid microscale processes and microscalers should consider how to improve the representation of larger-scale processes. Experimentalists are especially welcomed to identify ABL phenomena that can be particularly difficult to characterize experimentally. During this initial phase of the PIRT we shall not create links between the phenomena of interest with planned NEWA experiments. Instead, identifying which objectives can be met with which experiments should be the consequence of the PIRT process.

Besides identifying phenomena, each expert should provide an assessment of the model adequacy in terms of physics, code and validation as per Table 2. Here “model” should not be regarded as own model but as the NEWA model-chain considering the state-of-the-art of the consortium. As soon as a first version of the model-chain is produced, we could be more specific as to the particular needs of such a collaborative model.

After a first round of responses from NEWA experts, a consensus meeting is set up to inform about the identified phenomena, adequacy levels and importance ranking (Table 3). The coordinator of the expert elicitation tries to homogenize results to come up with a first draft of the PIRT table. After this initial review, experts are asked to reexamine their initial assessment of model adequacies, this time providing marks to the consolidated list of phenomena on the draft PIRT.

Initially, the PIRT table addresses high-level issues and, as more data is added it will become more detailed, for example, dividing one phenomenon in two or more issues. Nevertheless, the experts can always ask to include more detail in the existing description or add new phenomena that may deserve its own place in the list during any iteration. The next rounds of expert elicitation will rank these new phenomena and so on.

It is expected that this iterative process will naturally remove outliers and lead to a reasonable level of consensus across the project. This will typically require two to four iterations.

To avoid a biased assessment due to possible lack of representativeness of the NEWA experts group, the PIRT table will be shared with the IEA-Wakebench and IEC 61400-15 groups. This will increase the size of the experts group and widen the scope of the PIRT process by incorporating other areas that are not developed in NEWA. For instance, the Wakebench forum will contribute with wind farm modeling aspects, notably through interaction with researchers from DoE’s A2e research program. Feedback from the IEC 61400-15 group, busy with the drafting of an IEC standard for wind resource assessment, will reinforce the end-user orientation and probably put more emphasis on uncertainty quantification.

5.4 Identifying Planning Priorities

The PIRT table is an initial instrument to list our needs for model development and evaluation. The bottom-up approach followed in the expert elicitation is not constrained by budgetary, time or other resource constraints. When a research project is associated to the PIRT process, one needs to examine which actions identified in the PIRT are more impactful considering the constraints of the project (funding, timing, experimental capabilities, computational resources, etc).

In the case of NEWA, the sites that will host the dedicated experiments were already selected during the proposal stage, each one with particular logistic and budgetary constraints besides their differences in the geophysical characteristics. Of course this site selection was not done arbitrarily and the sites share a fair amount of complementarities to allow a good assessment of the mesoscale to microscale model-chain. The research program in NEWA will determine how to best link these sites with the PIRT to efficiently produce the most impactful experimental database and validation hierarchy.

Again, sharing these plans with other international groups will be sought to benefit from synergies with other parallel projects of similar orientation.

6. PIRT for Wind Conditions Assessment

Table 4 shows the PIRT table that resulted from the initial survey to NEWA partners. The scope is limited to addressing external atmospheric wind conditions (not wakes) relevant for the EU wind atlas context.

For each phenomenon the level of detail in the survey responses was sufficient to identify the issue that the model needs to address and to define specific responses the NEWA consortium can implement. This research roadmap is subject to change throughout the project as more evidence is acquired from the model development and evaluation activities.

In total, 17 phenomena have been identified. Many are interrelated at this stage. As better assessment of each of them is obtained from the validation activities, it will be possible to be more specific and produce a more targeted strategy.

Table 4: Phenomena Identification and Ranking Table for the NEWA Model-Chain

Phenomenon		Importance	Model Adequacy			Plan. Priority	Issue	Response
ID	Description		Physics	Code	Validation			
Global to Mesoscale								
1	Mesoscale **probabilistic wind atlas** approach not defined. Determine the value of a probabilistic versus deterministic approach in wind assessment applications	H	L	L	L	H	The uncertainty of the wind resource using a meso-micro modeling methodology requires the introduction of a probabilistic approach at mesoscale level. This will define the input uncertainties that will be propagated through the microscale models.	Build on ensemble forecasting techniques to produce a probabilistic wind atlas based on high-res simulations. Extend historical coverage with analog method if necessary. Evaluate uncertainty at experimental sites using probabilistic metrics. Categorize uncertainties depending on climate and site characteristics
2	Set-up of **mesoscale** model, **sensitivity analysis**	M	H	H	L	H	Determine an "optimal" configuration of the WRF model for wind assessment inputs, grid, PBL scheme, spin-up, nesting, nudging, etc)	Sensitivity analysis to determine the most influential settings of WRF on wind assessment. Definition of a WRF reference model for wind energy
3	Incorporation of "other" environmental factors that may affect AEP, in particular **icing**	U	L	L	L	L	There is good experimental evidence of the potentially large impact of icing on wind farm performance but there is no conventional method on how to assess these losses using model simulations	Experiment to characterize ice build-up with environmental conditions and correlate with wind farm performance. Validation of WRF simulations predicting ice build-up
Meso to Microscale								
4	Characterization of mesoscale forcing free of microscale effects consistently at various terrain complexities. On the definition of the **generalized wind climate** to define appropriate inputs for microscale models	H	L	M	L	H	Physical downscaling from meso to micro using statistical coupling requires the characterization of adequate mesoscale forcing free of microscale effects. This requires the characterization of the effect of aggregated terrain drag on mesoscale/geostrophic forcing (upscaling to a generalized wind climate)	Definition of a generalized wind climate methodology applied to mesoscale outputs to produce adequate inputs for microscale models. Assess consistency of the method across various mesoscale resolutions and CFD domain treatment. Assess impact in terms of bias reduction on downscaled results at experimental sites

5	**When to switch** from mesoscale to microscale. What is the optimum cut-off scale (mesoscale final resolution)?	M	L	L	L	H	It is not completely clear the scales (spatial and temporal) and under which climate/orographic characteristics when the terra-incognita is reached by the models,. How does a coupled meso-micro model-chain compares to a high-resolution multi-scale model?	Continue with a systematic “hierarchy of complexity” V&V procedure, where the experiments can be categorized in accordance to these scales. Comprehensive sensitivity analysis of domain’s extents from both mesoscale and microscale models.
6	Dynamical or statistical **coupling?**	L	M	M	L	L	Multi-scale (WRF) dynamical modeling is limited by limitations in steep terrain due to terrain-following coordinate system. Statistical coupling to a CFD solver can help solving complex terrain flows but it is not clear how many atmospheric "classes" are needed to characterize the wind climate.	WRF immersed boundary method (under implementation) should allow more flexibility in complex terrain. Optimization of microscale runs is required to handle the large dimensionality of inputs in microscale models. Perform a cost/benefit analysis of dynamical vs statistical coupling methods.
7	**Extreme events** from mesoscale and other transients affected by microscale	M	M	L	L	H	How to compare mesoscale data (spatial averaged data) with single-point measurements? How can information on extreme events be derived from spatially averaged information? How important is short-term variability (which is not accounted for in mesoscale models and steady state or statistical microscale models) for AEP assessment?	Application of the analog ensemble method to derive local time series from spatially-averaged mesoscale data. Comparison with measurements and turbine data. Development of models for the derivation of information on extreme events.
8	**Model evaluation across scales**. How model and observations can be more fairly compared? Downscaling simulations and/or upscaling measurements? How do we quantify the value of added physics/resolution?	H	L	L	L	H	To evaluate the realism of the simulated wind field, the simulated dataset needs to be comparable with the observational one by using spatio-temporal masking or other "upscaling" methods	Ensure model and observations are interpreted similarly before comparing them. The degree of upscaling of the observations required to reach the mesoscale level determines the relative importance of local effects that need to be modeled by the microscale model. Sensitivity analysis with varying physical complexity and grid resolution.

Microscale								
9	Introduction of **atmospheric stability** in ABL microscale modeling. Characterization from measurements and mesoscale modeling to define adequate inflow and boundary conditions at microscale.	H	M	M	L	H	Inadequate inflow and boundary conditions for microscale runs in non-neutral conditions. Homogeneous (idealized) or mesoscale (realistic) inflow? Steady or unsteady?	Generation of appropriate inflow for transient CFD simulations of varying atmospheric stability. Validation in non-neutral and transient conditions
10	Characterization of **surface conditions** across scales. Building consistency between land-cover information used at mesoscale and high-resolution scans at microscale.	M	L	L	L	H	Consistent treatment of surface conditions across scales is lacking to propose a robust methodology on the selection of appropriate input data from terrestrial databases. Validation of surface conditions requires measurements of momentum and energy fluxes in the surface layer	Define a method for the definition of surface boundary conditions (terrain height, roughness length, canopy drag, surface temperature/heat flux) in terms of mesoscale outputs and terrestrial databases. Evaluate at experimental sites in terms of surface momentum and heat fluxes at different scales.
11	Assess the impact of **grid sensitivity** in unresolved physics, especially in **complex terrain**	M	M	M	L	H	Geometric fidelity and/or grid resolution insufficient to capture behavior. Determine a robust methodology for grid generation in complex terrain and ways of quantifying grid quality/dependency	Sensitivity analysis of subsystem level with higher fidelity model to assess impact of under-resolved geometry. Grid studies (solution verification) to characterize uncertainty due to grid dependencies
12	Simulation of **land-sea transitions** and their impact on near offshore wind conditions	H	L	M	L	H	Streaks of low wind speed have been observed in mesoscale simulations of the land-sea transition (Dörenkämper et al., 2015). These streaks are likely to be caused by roughness changes over land and extend several tens of kilometers offshore.	Simulations of the land-sea transition with turbulence resolving models (LES). Experimental investigation (LiDAR, SAR). If necessary, modifications of mesoscale model physics.
13	Interaction of **forest canopies** with ABL in simple and complex terrain under different atmospheric conditions	H	M	M	L	H	Consistency between canopy characteristics and turbulence model in mesoscale and microscale models is required. This should include also the effect of atmospheric stability and the interaction with terrain.	Characterization of the wind profile within a forest canopy at various stabilities in simple and complex terrain. Determine the consistency of canopy characteristics between meso/microscale inputs and turbulence

								parameterizations.
14	Characterization of **vertical profile** under different atmospheric forcing conditions (surface roughness, stability, in-homogeneous conditions, etc)	H	M	M	L	H	How do we characterize "non-conventional" wind profiles originating from realistic conditions (heterogeneous, transient). Beyond the power-law exponent, are there more physically meaningful ways of defining a wind profile in terms of the underlying forcing? How relevant are these different forcing mechanisms in terms of turbine and wind farm performance?	Analysis of measured wind profiles for different stabilities and mesoscale conditions and comparison with turbine and wind farm performance data. Assessment of wind profile climatology in terms of various forcing mechanisms to determine the most influential ones in different terrain conditions. Measure up to ABL height to characterize geostrophic conditions. Complement observations with mesoscale simulations to characterize mesoscale tendencies.

Phenomenon		**Importance**	**Model Adequacy**			**Plan. Priority**	**Issue**	**Response**
ID	**Description**		**Physics**	**Code**	**Validation**			
Uncertainty Quantification								
15	Characterization of **parameter uncertainty**	M	L	L	L	M	Inadequate characterization of model parameter uncertainties	Characterize from experimental data, data provided in literature, or from new experiments
16	Characterization of **uncertainty on experimental data** for various data sources (synoptic surface stations, tall met masts, remote sensing)	H	M	M	L	H	Insufficient or low-quality observational data will reduce the success of model evaluation	Need better quantification of uncertainty of experimental data
17	Uncertainty characterization with **low sample size**	H	L	L	L	H	Bayesian-related probabilistic methods, like ensembles, require a potentially large number of simulations. Uncertainty in model prediction not adequately characterized due to unfeasible large number of runs required	Approximate methods like polynomial chaos, initial and boundary data uncertainty, parametric uncertainty (how accurate are model parameters)

7. NEWA Observational Datasets

The observational datasets are gathered in WP2. Data of different quality is gathered for different purposes and have different accessibility constraints. The following categories of identified data are sorted in ascending order of quality:

- Public meteorological database of surface observations (NEWA-Synop)
    - NEWA access, will be open-access
    - Large spatio-temporal coverage
    - To evaluate regional climatology and areas without “wind energy” measurements
    - 1000+ sites
- Restricted tall mast data from external contributors (NEWA-Tall-Blind)
    - Private access
    - To run blind tests based on standardized evaluation methodology
    - Performance in terms of statistics, with aggregated results categorized by wind climate and site conditions
    - Groups of 100+ sites
- Tall mast data available from NEWA partners or publically available (NEWA-Tall-Open)
    - NEWA-access, some will be open-access
    - High quality profile data in several wind climate and site conditions
    - To increase the validation range of experiments on meso-micro methodologies
    - Performance evaluated on reference sites with data access that can be representative of the categories identified in the aggregated results of the blind tests
    - 20+ sites
- Experiments:
    - NEWA access, will be open-access
    - High fidelity experiments targeting specific modeling objectives, typically in terms of flow cases that are specifically meaningful for the phenomenon under investigation
    - At least one year, including an intensive campaign with WindScanner and other non-conventional systems
    - 5+ sites

8. NEWA Benchmarking Strategy

Figure 4 shows a block diagram that summarizes the verification, validation and uncertainty quantification (VV&UQ) strategy.

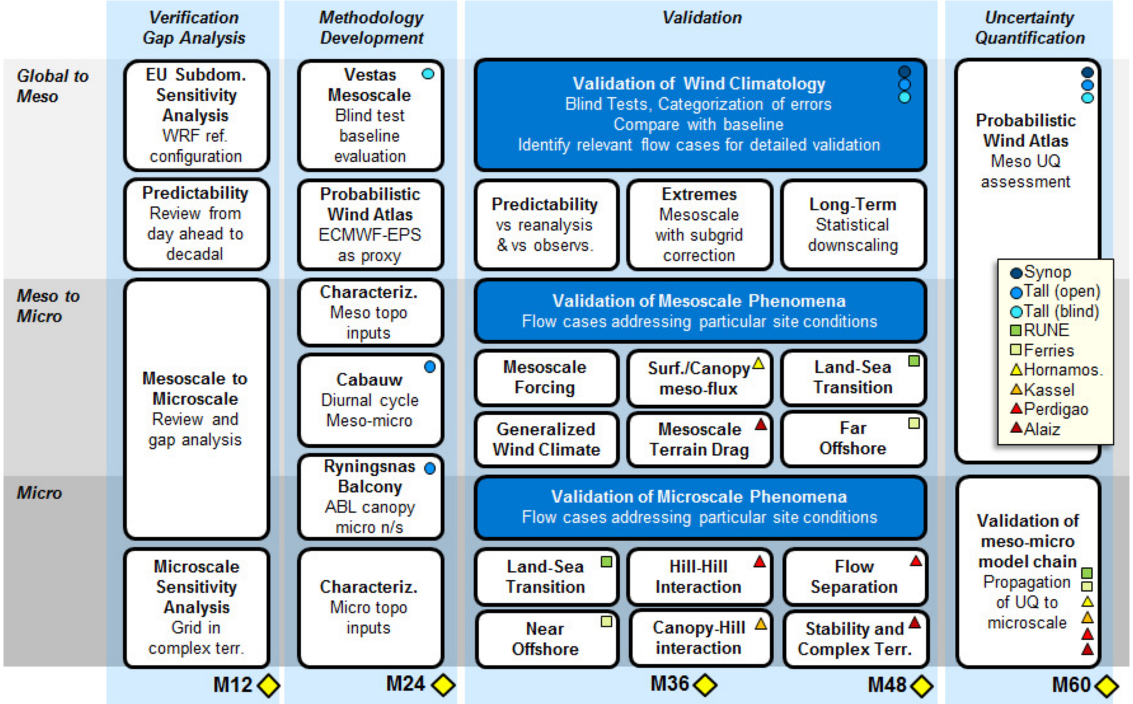

Figure 4: Schematic of the NEWA VV&UQ strategy. Symbols are used to identify a validation building block with the most relevant datasets

The process is based on four phases:

1. Verification and Gap Analysis: The first year of the project is dedicated to coordinating the different teams of modelers to "speak the same language". Model evaluation is not so much directed to comparing with observations but to compare different models available in the project and their numerical sensitivities. A gap analysis (this document) is done in order to identify modeling needs and seek state-of-the-art solutions.
2. Methodology Development: the second year will be devoted to developing the modeling capabilities required in the model chain. Some validation activities are conducted in order to establish a baseline performance of the model-chain based on a set of metrics focused on the quantities of interest. An initial blind test with data from Vestas will be used to design these metrics so they can be used later on in connection to other datasets from industry. At microscale, the activity will be focused on developing ABL models and their coupling with mesoscale outputs.
3. Validation: As the CfWD and the experiments produce data, there will be a number of validation studies to test and improve the model-chain. Based on the CfWD there will be a number of blind tests that will use the same set of metrics of the baseline. By analyzing the aggregated results in terms of wind climate and site characteristics it will be possible to identify the main shortcomings of the model-chain that can be subject to more detailed investigation with the experimental datasets. This will result in a number of validation exercises that will target specific modeling objectives in terms of well-defined flow cases.

These can be divided into two large groups: 1) those related to specific site conditions at mesoscale and the interfacing between mesoscale and microscale; and 2) those related to specific site conditions at microscale.

4. Uncertainty Quantification: The last year of the project will be devoted to the production of the wind atlas and its associated uncertainties. From a meso-micro perspective this has to be with the assessment of the mesoscale uncertainty and how this is propagated to microscale by the downscaling methodology.

Annex I presents the hierarchy of model evaluation activities of the project defined in terms of the test case (dataset) and associated benchmarks, each one with an objective and expected results. The persons in charge of each test case and benchmark are identified as well as the institutions that intend to participate in each one.

9. Conclusions

The NEWA model evaluation strategy is presented following the V&V framework adopted also in the A2e research program. A research roadmap is defined in terms of a PIRT table to identify the relevant phenomena that the NEWA model-chain should incorporate in the development and evaluation process. From this table, a hierarchy of validation cases has been defined considering the expected datasets of the project.

References

AIAA (1998), Guide for the Verification and Validation of Computational Fluid Dynamics Simulations. AIAA-G-077-1998, American Institute of Aeronautics and Astronautics, Reston, VA, p. 3.

ASME (2009), Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer. ASME V&V 20-2009, The American Society of Mechanical Engineers, New York, NY.

Hale E. The Uncertainty of Uncertainty. 2015 Wind Energy Systems Engineering Workshop, University of Colorado Boulder, Colorado, January 2015. http://www.nrel.gov/wind/systems_engineering/pdfs/session_2_hale_uncertainty_of_uncertainty.pdf, accessed November 2015

Helmer O (1967) Analysis of the Future. The Delphi Method. RAND Corporation, January 1967, http://www.rand.org/pubs/papers/P3558.html, accessed November 2015

Hills R.G., Maniaci D.C., Naughton J.W. V&V Framework. SANDIA Report SAND2015-7455, September 2015, http://prod.sandia.gov/techlib/access-control.cgi/2015/157455.pdf, accessed November 2015

IEC (2005) IEC 61400-1 Wind Turbines Part1: Design Requirements, 3rd edition. International Electrotechnical Commission IEC 61400-1:2005(E), pp 85

Oberkampf WL, Pilch M, Trucano TG (2007). Predictive Capability Maturity Model for Computation Modeling and Simulation, Sandia National Laboratories Report SAND2007-5948, October 2007, https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/Oberkampf-Pilch-Trucano-SAND2007-5948.pdf , accessed November 2015

Petersen EL, Troen I (2012) Wind conditions and resource assessment. WIREs Energy Environ, 1: 206–217 doi: 10.1002/wene.4

Pilch MT, Trucano TG, Moya J, Froehlich G, Hodges A, Peercy D (2001). Guidelines for Sandia ASCI Verification and Validation Plans: Content and Format: Version 2.0. Tech. Report SAND2000-3101, Sandia National Laboratories, Albuquerque, New Mexico 87185 and Livermore, California 94550, January 2001, http://prod.sandia.gov/techlib/access-control.cgi/2000/003101.pdf, accessed November 2015

Sanz Rodrigo J, Moriarty P (2015) WAKEBENCH Model Evaluation Protocol for Wind Farm Flow Models, Edition 1. IEA Task 31 Report to the IEA-Wind Executive Committee. May 2015

Sanz Rodrigo J, Chávez Arroyo R.A., Moriarty P, Churchfield M, Clifton A, Kosovic B, Réthoré P-E, Hansen KS, Hahmann A, Mann J, Mirocha JD, Rife D (2016) Mesoscale to Microscale Wind Farm Modelling and Evaluation. WIREs Energy Environ, accepted for publication

Trucano TG, Pilch M, Oberkampf WL (2002) General Concepts for Experimental Validation of ASCI Code Applications. Sandia Report: SAND2002-0341, Sandia National Laboratories, Albuquerque, NM, http://prod.sandia.gov/techlib/access-control.cgi/2002/020341.pdf, accessed November 2015

Annex I: NEWA Validation hierarchy as of March 2016

Test Case (dataset)	TC Manager	Benchmarking objectives	PIRT	Benchmark Manager	Expected results	Participants	Start	Due
NEWA-Synop	Fidel González Rouco (UCM)	Support for blind tests for wind climate and site categorization	1,2,4, 5,8	Elena García (CIEMAT)	Extend categorization of errors to a wider diversity of site and wind climate conditions, especially in regions without tall mast measurements	UCM, CIEMAT, LU (Latvia only)	M12	M24, M36, M48, M60
		Analysis of long-term representativeness of wind atlas using stations with 20+ years (combine with Tall mast data where available)	1,2,4, 5,8	Gerald Steinfeld (ForWind)	MCP technicques, assessment of the integration period for the NEWA mesoscale runs	ForWind, CIEMAT, UCM, ATM-PRO	M12	M24
		Predictability at various scales. Model results will be evaluated against observational data sets; re-analyses and observational data sets wherever available.		Albert Soret (BSC)	Validation of predictability atlas	BSC, CENER	M12	M24, M36, M48, M60
		Extreme winds using stations with 20+ years (combine with Tall mast data where available)	4,7	Xiaoli G. Larsen (DTU)	Validation of Vref atlas	DTU, CIEMAT, UCM	M12	M24
		Uncertainty quantification. ECMWF-EPS dataset will be used as proxy for developing the methodology for probabilistic wind assessment.	1,8,1 5,16, 17	Sergio Fernández (UCM)	Characterization of uncertainty in probabilistic models by an optimal spread index.	UCM	M8	M24

Test Case (dataset)	TC Manager	Benchmarking objectives	PIRT	Benchmark Manager	Expected results	Participants	Start	Due
NEWA-Tall-Blind	Andrea Hahmann (DTU)	Blind test using tall mast data from industry (Vestas) to assess skill of different meso-micro methodologies. The set of evaluation metrics will be used in connection to other datasets as a result of the CfWD	4,8	Bjarke Tobias Olsen (DTU)	Validation of wind assessment methodologies and categorization of errors in terms of site and climate complexities. Baseline validation results for first release of NEWA model-chain	DTU, Vestas, ATM-PRO?	M15	M24
		Other blind tests resulting from CfWD at various stages of the project, using the same metrics developed with the Vestas case	4,8	TBD	Depends on the outcome of the CfWD, due in M24. Improved performance as project evolves. Improved interpretation of metrics including uncertainty quantification	TBD	TBD	M36, M48, M60
NEWA-Tall-Open	Andrea Hahmann (DTU)	Define WRF set-up methodology. Sensitivity analysis at various EU subdomains. Preliminary analysis around reference tall masts	2	Andrea Hahmann (DTU)	A reference WRF set-up will be defined as baseline for optimization in the validation process	DTU, CIEMAT, UCM, LU, ITU, ForWind, Wtech, DNV-GL	M8	M15
		Detailed analysis of meso-micro methodologies at reference (open) sites	1,2,4, 5,8	Bjarke Tobias Olsen (DTU)	Validation leading to fine-tunning of models	ITU, CIEMAT, UCM, ATM-PRO	M15	M24, M36, M48, M60

Test Case (dataset)	TC Manager	Benchmarking objectives	PIRT	Benchmark Manager	Expected results	Participants	Start	Due
NEWA-Cabauw	Javier Sanz Rodrigo (CENER)	Based on Cabauw data, evaluate meso-micro methodologies reproducing a mean diurnal cycle over horizontally homogeneous terrain. Characterization of wind profile under a mesoscale forcing climatology. Propagation of mesoscale uncertainty to microscale	1,4,6, 8,9,1 4,17	Javier Sanz Rodrigo (CENER)	Follow-up of GABLS3 using ensemble-averaged diurnal profiles. Characterization of mesoscale forcing for microscale and definition of vertical profile metrics relevant for wind energy	CENER, ForWind, CENAERO, CIEMAT, UCM, DTU, KUL	M16	M24
NEWA-Ryningsnas	Stefan Ivanell (UU)	Blind test for the generation of wind profiles over an heterogeneous forested canopy in nearly-flat terrain. Near and stable conditions for three wind direction sectors	10,13 ,14	Stefan Ivanell (UU)	Baseline validation study for ABL microscale models above a forest canopy	CENER, ForWind, DTU, Cenaero, VESTAS?	M12	M16, M19
NEWA-Hornamossen	Stefan Ivanell (UU)	Experiments with a 180 m met mast in combination with remote sensing instruments along a line of 14 km. (in total about 12 remote sensing instruments) The site has heterogeneous forested canopy in moderate complex terrain. The benchmark will be used for model validation for both micro and mesoscale models.	2,4,8, 9,11, 11,13	Stefan Ivanell (UU)	Improved validation of existing models and posibilities to improve model parameters from experimental data.	UU, WTech	M24	M36

Test Case (dataset)	TC Manager	Benchmarking objectives	PIRT	Benchmark Manager	Expected results	Participants	Start	Due
NEWA-Mut	Sibel Menteş (ITU)	Verification of simplified microscale model. Grid dependency study in complex terrain	11	Sibel Menteş (ITU)	Quantification of numerical errors at different grid resolutions. Definition of methodology for microscale model set-up	ITU	M18	M36
NEWA-RUNE	Alfredo Peña (DTU)	Simulation of the development of a stable internal boundary layer from land to sea (warm air advected over cold water)	9,10, 11,12 ,14	Björn Witha (ForWind)	Baseline validation study for LES and ABL microscale models for shallow marine boundary layers	ForWind, DTU, IWES, 3E, KUL	M12	M18 (with Task 2.14)
NEWA-Balcony	Ebba Dellwik (DTU)	Simulation of heterogeneous forested site - validation with lidar horizontal scans	4,10, 13,14	Dalibor Cavar (DTU)	Characterize heterogeneous canopy flow based on lidar scans and simulations	DTU, U.Porto?	M24	M36
NEWA-LIDARFerryLines	Julia Gottschall (IWES)	Mesoscale validation study for near offshore wind fields under different stability conditions (depends on available data)	12.14	Björn Witha (ForWind)	Knowledge on WRF's ability to capture the marine boundary layer in near shore regions	ForWind	M27	M36
		Mesoscale validation study for far offshore wind fields under different stability conditions (depends on available data)	12.14	Björn Witha (ForWind)	Knowledge on WRF's ability to capture the marine boundary layer in far shore regions	ForWind	M27	M36
		Microscale validation study: Land-sea transition, verify streaks with LES and ABL microscale models (compare with LiDAR and WRF results and satellite data)	9,10, 11,12 ,14	Björn Witha (ForWind)	Baseline validation study for LES and ABL microscale models for land sea transition with offshore wind speed streaks	ForWind	M33	M42

Test Case (dataset)	TC Manager	Benchmarking objectives	PIRT	Benchmark Manager	Expected results	Participants	Start	Due
	Andrea Hahmann (DTU)	Mesoscale validation study: occurrence of low-level jets in obs and models and their influence in the coupling between meso- and micro	4,8,9, 14	Patrick Volker (DTU)	Methodology for the generalization of wind climatologies when low-level jets exist and they impact the wind climate	DTU	M12	M42
NEWA-Kassel	Doron Callies (IWES)	Verification of simplified microscale model. Grid dependency study in complex terrain	11	Chi-Yao Chang (IWES)	Quantification of numerical errors at different grid resolutions. Definition of methodology for microscale model set-up	IWES	M8	M12
		Characterization of canopy and testing of different canopy and surface roughness models	10,13	Chi-Yao Chang (IWES)	Development of numerically and physically consistent model regarding turbulent transports	IWES, ForWind, CENER, CENAERO, BSC	M16	M24
		Validation of mean flow for the SSW wind direction under different thermal stratification classes based on mesh sensitivity study and measure campaign	9,10, 11,13	Chi-Yao Chang (IWES)	Evaluation of the impact of the wake from Rödeser Berg upstream on the 200m met mast location		M40	M48
		Validation of NEWA model-chain over the experimental campaign period	1,4, 9,10, 11,13	Chi-Yao Chang (IWES)	Full-validation of model-chain including uncertainty quantification	IWES, ATM-PRO	M40	M60

Test Case (dataset)	TC Manager	Benchmarking objectives	PIRT	Benchmark Manager	Expected results	Participants	Start	Due
NEWA-Perdigao	José L. Palma (UPorto)	Verification of simplified microscale model. Grid dependency study in complex terrain	11	Andreas Bechmann (DTU)	Quantification of numerical errors at different grid resolutions. Definition of methodology for microscale model set-up	DTU, BSC	M8	M12
		Validation of NW sector perpendicular to the eastern ridge (46 azimuth)	9,10, 11,14	José L. Palma (Uporto)	Evaluation of flow pattern over the ridges	ITU, CENER, ForWind, CENAERO, BSC, ATM-PRO	M15	M36
		Validation of SW sector perpendicular to the western ridge (231 azimuth)	9,10, 11,15	José L. Palma (Uporto)	Evaluation of flow pattern over the ridges		M15	M36

Test Case (dataset)	TC Manager	Benchmarking objectives	PIRT	Benchmark Manager	Expected results	Participants	Start	Due
NEWA-Alaiz	Elena Cantero (CENER)	Verification of simplified microscale model. Grid dependency study in complex terrain	11	Roberto A. Chávez (CENER)	Quantification of numerical errors at different grid resolutions. Definition of methodology for microscale	CENER, BSC	M8	M12
		Characterization of canopy and testing of different canopy and surface roughness models. Assessment of large scale forcing	4,10	Roberto A. Chávez (CENER)	Verified topographic inputs and development of a climatology of mesoscale forcing for meso-micro modeling	CENER, CIEMAT	M12	M24
		Validation of mean flow for the northerly wind sector under various stabilities (cycle of varying stability) based on intensive campaign	4,9,10,11, 13	Javier Sanz Rodrigo (CENER)	Evaluation of the impact of the wake from Tajonar hill upstream on the Alaiz wind conditions	CENER, ITU, BSC	M40	M48
		Validation of mean flow for the southerly wind sector under various stabilities (cycle of varying stability) based on intensive campaign	4,9,10, 11,13	Javier Sanz Rodrigo (CENER)	Evaluation of wind conditions at the lee of the Alaiz mountain ridge	CENER, ITU, BSC	M40	M48
		Validation of NEWA model-chain over the experimental campaign period	1,4,5, 8,15, 16,17	Javier Sanz Rodrigo (CENER)	Full-validation of model-chain including uncertainty quantification	CENER, ITU, CIEMAT, UCM, ATM-	M40	M60
		Validate the methodology developed for the ECMWF EPS by probabilistic simulations carried out with WRF mesoscale model. Experimental data of Alaiz field campaign will be used for the validation.	1,2,3, 5,8,15,16, 17	Sergio Fernández (UCM)	Quantification of numerical errors in a mesoscale probabilistic wind forecast. Icing risk could also be estimated.	UCM	M18	M36

Acknowledgements

The NEWA project counts with the support from an ERA-Net Plus consortium composed by the European Commission and 9 funding agencies from 8 member states:

- Public Service of Wallonia, Department of Energy and Sustainable Building (Belgium)
- Department of Economy, Science and Innovation Flemish Government (Belgium)
- Danish Energy Authority (Denmark)
- Federal Ministry for the Economic Affairs and Energy, on the basis of the decision by the German Bundestag (Germany)
- Latvijas Zinatnu Akademija (Latvia)
- Fundação para a Ciência e a Tecnologia (Portugal)
- Ministerio de Economía y Competitividad (Spain)
- The Swedish Energy Agency (Sweden)
- The Scientific and Technological Research Council of Turkey (Turkey)