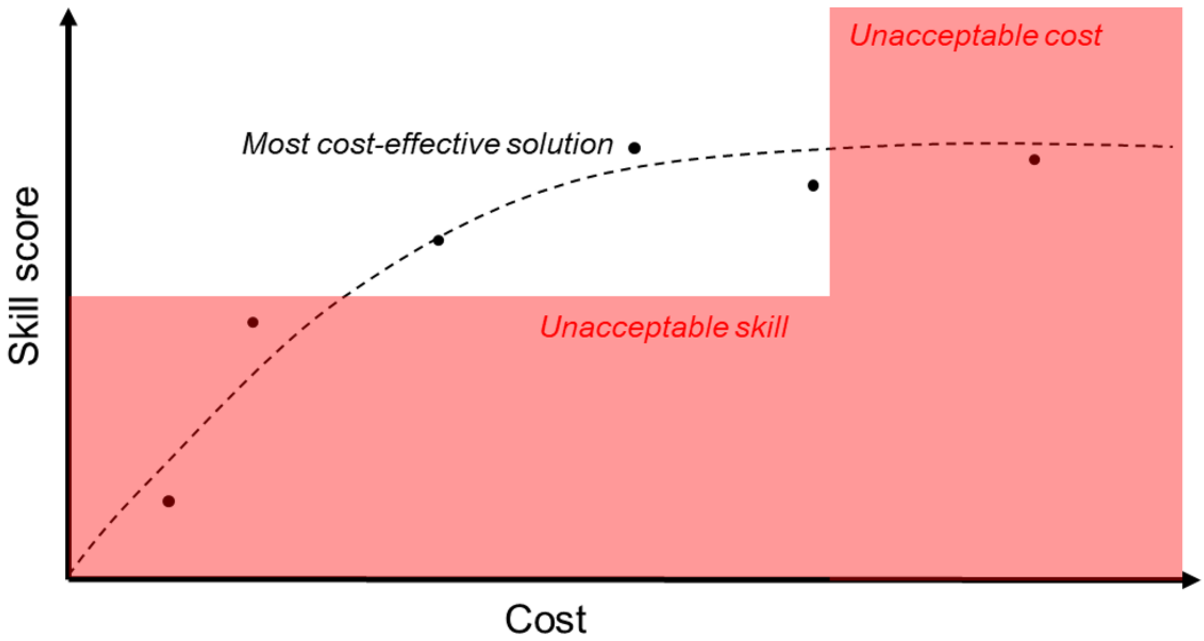
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of the Environment, Transport, Energy and Communications DETEC

Swiss Federal Office of Energy SFOE
Energy Research and Cleantech Division

Final report dated 19.12.2022

COMESI - Comparison metrics simulation challenge



Source: OST 2020

Date: 19.12.2022

Location: Rapperswil

Publisher:
Swiss Federal Office of Energy SFOE
Energy Research and Cleantech
CH-3003 Bern
www.bfe.admin.ch

Co-financing:
OST – Ostschweizer Fachhochschule
Oberseestrasse 10
8640 Rapperswil

Subsidy recipients:
OST – Ostschweizer Fachhochschule
Oberseestrasse 10
8640 Rapperswil

Authors:
Florian Hammer, OST, florian.hammer@ost.ch
Sarah Barber, OST, sarah.barber@ost.ch

SFOE project coordinators:
Katja Maus, katja.maus@bfe.admin.ch
Lionel Perret, lionel.perret@planair.ch

SFOE contract number: SI/501955-01

Zusammenfassung

In der Windenergiebranche hat die Genauigkeit der Schätzung der Windressourcen einen enormen Einfluss auf die erwartete Rendite eines Projekts. Aufgrund der komplexen Natur des Wetters und der Windströmung über der Erdoberfläche kann es sehr schwierig sein, die Windressource korrekt zu messen und zu modellieren. Für ein bestimmtes Projekt stehen Modelliererinnen und Modellierer vor einer schwierigen Auswahl aus einer breiten Palette von Simulationstools mit unterschiedlichen Genauigkeiten und Kosten. In diesem Projekt wird in Zusammenarbeit mit der IEA Wind Task 31 eine öffentliche "Simulationschallenge" für Windenergiestandorte in komplexem Gelände durchgeführt, bei der die Teilnehmerinnen und Teilnehmer ihre Simulationsdaten und -ergebnisse in einer vordefinierten Vorlage einreichen. Ziel ist es, mehrere Datensätze bezüglich der "Skill" und "Costs" von Simulationstools sowohl vor als auch nach der Durchführung der Simulationen zu sammeln, so dass Transferfunktionen für die genaue Vorhersage der "Skill" und "Costs" der Tools entwickelt werden können. Dies wird den Modelliererinnen und Modellierern helfen, für ein bestimmtes Windenergieprojekt das beste Modell für die Aufgabe auszuwählen.

Im ersten Jahr dieses Projekts wurde eine Vorlage für Vergleichsmetriken weiterentwickelt und im Vergleich zur ursprünglichen Version stark verbessert. Die Simulationschallenge wurde planmässig entworfen und gestartet. In der Folge wurden mehrere Publikationen und Kollaborationen lanciert.

Die Datenerfassung und –auswertung wurde mit einer Verzögerung von drei Monaten, aufgrund von einer geringen Anzahl von Einreichungen, abgeschlossen, weshalb sich die Planung von Stage 2 ebenso verzögert hat.

Im zweiten Jahr des Projekts wurde zunächst die Stage 1 abgeschlossen. Dabei haben wir uns bei den registrierten Teilnehmerinnen und Teilnehmern informiert, warum die Ergebnisse verspätet eingereicht wurden, und haben die Challenge den Umständen entsprechend angepasst. Der Evaluierungscode wurde zudem vervollständigt und die Analyse der Ergebnisse abgeschlossen. In einem Workshop wurden die Ergebnisse den Teilnehmerinnen und Teilnehmern präsentiert sowie diskutiert. Dabei konnten offene Fragen und Ungereimtheiten geklärt werden.

Die Ergebnisse von Stage 1 haben gezeigt, dass anspruchsvolle Simulationswerkzeuge wie Large Eddy Simulations (LES) nicht unbedingt zu höheren Genauigkeiten führen. Insbesondere für weniger komplexe Standorte ist es besser, einfachere Tools wie Reynolds Averaged Navier Stokes (RANS) oder WAsP zu verwenden, die mit einem Bruchteil der Kosten von LES-Simulationen ein hohes Mass an Genauigkeit erreichen. Insgesamt erzielte die RANS-Simulation mit der von Enercon entwickelten E-Wind-Software die besten und beständigsten Ergebnisse. Da der Teilnehmer auch der erfahrenste von allen war, könnte dies zu der Schlussfolgerung führen, dass die Fähigkeiten des Anwenders eine entscheidende Rolle für die Gesamtbewertung der Fähigkeiten spielen.

Für Stage 2 dieses Projekts wurde der manuelle Prozess von Stage 1 automatisiert. Das daraus resultierende Decision Tool ist in der Lage, Antworten auf Fragebögen automatisch in Kompetenz- und Kostenbewertungen umzuwandeln. Um die Algorithmen des Tools zu entwickeln, wurde eine neue Challenge veröffentlicht. In dieser Challenge wurden die Teilnehmer aufgefordert, Simulations- und Messergebnisse von beliebig verfügbaren Standorten hochzuladen. Auf der Grundlage dieser Ergebnisse sollten von der Standortkomplexität abhängige Funktionen und Gewichtungen entwickelt und abgestimmt werden. Aufgrund mangelnder Beteiligung konnte der Tuning-Teil jedoch nicht abgeschlossen werden. Die daraus resultierenden Methoden dienen jedoch als Ausgangspunkt und können leicht aktualisiert und erweitert werden, wenn weitere Daten hinzukommen. Wir stehen diesbezüglich bereits in Kontakt mit mehreren Unternehmen, darunter Enercon, UL und TÜV Süd. Außerdem wurden die Fragen des Fragebogens neu formuliert, um sie leichter zu beantworten und zu verstehen sowie objektiver zu machen.

Der Einfluss der Komplexität des Standorts auf die Bewertung der Fähigkeiten im Vergleich zu den Kosten wurde kurz untersucht, indem drei verschiedene Modelle für vier Standorte mit zunehmender Komplexität verglichen wurden. Es zeigte sich, dass mit zunehmender Komplexität des Standorts die

"Before" Skill Scores abnahmen und die Abstände zwischen den Skill Scores der einzelnen Modelle in Abhängigkeit von ihrer Komplexität zunahm, d. h. LES übertraf die RANS- und WAsP-Simulationen. Bei weniger komplexen Standorten wiederum schnitten die RANS- und WAsP-Simulationen ähnlich gut ab wie der LES-Fall, wiesen aber deutlich bessere Kostenwerte auf. Die entwickelten Funktionen und Gewichte, mit denen diese Erkenntnisse erzielt werden konnten, wurden aufgrund der wenigen verfügbaren Datenpunkte von Hand abgestimmt. Mit mehr Daten können maschinelles Lernen und statistische Modelle diesen manuellen Prozess ersetzen, um allgemeinere und zuverlässigere Ergebnisse zu erhalten. Die derzeitigen Methoden und Funktionen dienen als Ausgangspunkt für die weitere Entwicklung.

Summary

In wind energy, the accuracy of the estimation of the wind resource has an enormous effect on the expected rate of return of a project. Due to the complex nature of the weather and of the wind flow over the earth's surface, it can be very challenging to measure and model the wind resource correctly. For a given project, the modeller is faced with a difficult choice of a wide range of simulation tools with varying accuracies and costs. In this project, a public "simulation challenge" for wind energy sites in complex terrain is being implemented in collaboration with IEA Wind Task 31, in which participants submit their simulation data and results in a pre-defined template. The goal is to collect hundreds of comparison metrics data regarding the "skill" and "costs" of simulation tools both before and after carrying out the simulations, enabling transfer functions for the accurate prediction of tool "skill" and "costs" to be developed. This will help modellers choose the best model for the job for a given wind energy project.

In the first year of this project, a submission template for comparison metrics was developed further and greatly improved compared to the initial version. The simulation challenge was designed and launched according to plan. Several publications and collaborations have been launched as a result of this. The data collection and evaluation was finished and a Python code for the analysis completed; however, we did not receive as many submission (five organisations with a total of 10 different submissions) as expected and the completion of Stage 1 had to be delayed by approximately three months. Due to this delay, the planning of Stage 2 was also delayed. However, the process of designing and evaluating Stage 1 allowed us to learn a great deal about simulation challenges, which served us well while designing Stage 2.

In the second year of the project, Stage 1 was first completed. This involved improving the submitted results by meeting and discussing the details directly with the participants, and then applying the data analysis code to the final results. The results of Stage 1 showed that sophisticated simulation tools such as Large Eddy Simulations (LES) do not necessarily lead to higher accuracies. Especially for less complex sites, one is better off using simpler tools such as Reynolds Average Navier Stokes (RANS) or WAsP, reaching high levels of accuracy with a fraction of the costs of LES simulations. Overall, the RANS simulation with the E-Wind software, developed by Enercon, achieved the best and most consistent scores. As the participant was also the most experienced amongst all, this leads to the conclusion that the user skill may play a crucial role for the overall skill score.

For Stage 2 of this project, the manual process of Stage 1 was automated. The resulting *Decision tool* is able to automatically convert answers of questionnaires into skill and cost scores. In order to develop the business logic of the tool, a new challenge was published. In this challenge participants were asked to upload simulation and measurement results of any available site. Based on these results site complexity dependent functions and score weightings were supposed to be developed and tuned. However, due to a lack of participation the tuning part could not be completed. However, the resulting methods serve as a starting point and can be easily updated and extended as more data comes in. We already have contacts with several companies including Enercon, UL and TÜV Süd

about this. Additionally, the questions of the questionnaire were reformulated to render them easier to answer and understand as well as more objective.

The influence of site complexity on the skill versus cost score plots was briefly explored by comparing three different models for four sites with increasing complexity. It was shown that for increasing site complexity the "before" skill scores decreased and the gaps between the skill scores of each model increased dependent on its sophistication, i.e. LES outperformed RANS and WAsP simulations. In turn, for less complex sites RANS and WAsP simulations performed similar to the LES case, but had significantly better cost scores. The developed functions and weights that were able to achieve these insights were tuned by hand due to the very few available data points available. With more data, machine learning and statistical models can replace this manual process in order to get more generalised and reliable results. The current methods and functions serve as a starting point for further development.

Summary

Dans le domaine de l'énergie éolienne, la précision de l'estimation de la ressource éolienne a un effet énorme sur le taux de rendement attendu d'un projet. A cause de la complexité des conditions météorologiques et des vents à la surface de la Terre, il peut être très compliqué de mesurer et de modéliser correctement la ressource éolienne. Pour un projet donné, le modélisateur ou la modélisatrice est confronté à un choix difficile entre plusieurs outils de simulations avec des précisions et des coûts différents. Dans ce projet, un "challenge de simulation" public a été élaboré pour plusieurs sites avec un terrain complexe, en collaboration avec le IEA Wind Task 31, dont les participants soumettent leurs données et résultats de simulation en suivant une structure pré-définie. Le but est de recueillir des centaines de données métriques de comparaison suivant la "technicité" et le "coût" des outils de simulations, avant et après la réalisation des simulations, ce qui permet de développer des fonctions de transfert pour la prédiction précise de la "technicité" et des "coûts" des outils. Cela aidera les modélisateurs et modélisatrices à choisir le meilleur modèle pour le projet éolien donné.

Lors de la première année de ce projet, un modèle de soumission pour les métriques de comparaison a été développé et grandement amélioré par rapport à la version initiale. Le challenge de simulation a été conçu et lancé selon le plan. Plusieurs publications et collaborations ont été lancées à la suite de ce challenge. La collection des données et leur évaluation ont été achevées ainsi que la réalisation d'un code Python pour l'analyse des données ; cependant, nous n'avons pas reçu autant de soumissions que prévu (5 organisations avec un total de 10 soumissions distinctes) et la fin de la première phase du projet a dû être retardée d'environ trois mois. A cause de ce retard, la planification de la deuxième phase a elle aussi été retardée. Cependant, la conception et l'évaluation de la première phase nous a permis d'apprendre beaucoup sur les challenges de simulation, ce qui nous a servi pour mieux préparer la deuxième phase.

Lors de la seconde année de ce projet, la phase une a été achevée. Ceci inclut l'amélioration des résultats soumis en rencontrant et en discutant des détails directement avec les participants, et l'utilisation du code d'analyse des données aux résultats finaux. Les résultats de la première étape ont montré que les outils de simulation complexes, comme les simulations des grands tourbillons (LES), n'impliquent pas forcément une meilleure précision. En particulier pour les sites moins complexes, il est préférable d'utiliser des outils plus simples, comme les méthodes RANS (Reynolds Average Navier Stokes) ou WAsP, qui permettent d'atteindre des bons niveaux de précision en une fraction du coût des simulation LES. Globalement, la simulation RANS avec le logiciel E-Wind, développé par Enercon, a obtenu les meilleurs résultats et les plus cohérents. Comme le participant était aussi le plus expérimenté de tous, ceci permet de conclure que l'expérience de l'utilisateur joue sans doute un rôle crucial dans le score générale de "technicité".

Pour la deuxième phase du projet, le processus manuel de la première phase a été automatisé. L'outil de décision qui en résulte est capable de convertir automatiquement les réponses du questionnaire en scores de technicité et de coût. Dans le but de développer la logique commerciale de l'outil, un nouveau challenge a été publié. Dans ce challenge, il était demandé aux participants de téléverser leurs résultats de simulation et de mesure de n'importe quel site disponible. A partir de ces résultats, des fonctions dépendant de la complexité du site et la pondération des scores devaient être développées et ajustées. Cependant, à cause du manque de participation, l'ajustement n'a pu être achevé. Cependant, les méthodes qui en résultent servent comme point de départ et peuvent être facilement mises à jour et étendues au fur et à mesure de l'arrivée de nouvelles données. Nous avons déjà des contacts avec plusieurs entreprises dont Enercon, UL et TÜV Süd. De plus, les questions du questionnaire ont étés reformulées pour les rendre plus facile à répondre et à comprendre ainsi que plus objectives.

L'influence de la complexité du terrain sur le score de la technicité par rapport au score du coût a été brièvement étudiée en comparant trois modèles différents pour quatre sites ayant une complexité croissante. Il a été démontré que pour une augmentation de la complexité du terrain, le score a-priori de la technicité décroit et les écarts entre les scores de technicité entre chaque modèle augmentent en fonction de la sophistication du modèle, c'est-à-dire que les simulations LES surpassent les simulations RANS et WAsP. En revanche, pour les sites moins complexes, les simulations RANS et WAsP ont des résultats similaires aux simulations LES, mais ont obtenu des scores bien meilleurs en termes de coût. Les fonctions et pondérations développées qui ont permis d'obtenir ces conclusions ont été ajustées à la main à cause du trop faible taux de données disponible. Avec plus de données, l'apprentissage automatique et les modèles statistiques peuvent remplacer ce processus manuel afin de mieux généraliser et rendre les résultats plus fiables. Les méthodes et fonctions actuelles servent comme point de départ pour un prochain développement.

Main findings

Stage 1:

1. A submission template for comparison metrics was developed further and greatly improved compared to the initial version.
2. A Python library for the analysis was developed and will be made public.
3. A simulation challenge was published and five organisations took part, leading to a total of ten different submissions.
4. The results showed that sophisticated simulation tools such as LES do not necessarily lead to higher accuracies.
5. The RANS simulations with the E-Wind software, developed by Enercon, achieved the best and most consistent scores for all locations at the Perdigao site.

Stage 2:

1. The manual process of Stage 1 was automated. The resulting *Decision tool* is able to automatically convert answers of questionnaires into skill and cost scores.
2. A new simulation challenge was published and one organisation took part, leading to a total of four different submissions. We have established contact with two other global organisations who are interested in contributing in the future.
3. The questions of the questionnaire were reformulated to render them easier to answer and understand as well as more objective.
4. It was shown that for increasing site complexity the "before" skill scores decreased and the gaps between the skill scores of each model increased dependent on its sophistication, i.e. LES outperformed RANS and WAsP simulations. In turn, for less complex sites RANS and WAsP simulations performed similar to the LES case, but had significantly better cost scores.

Contents

Zusammenfassung **3**

Summary **4**

Summary **5**

Main findings **7**

Contents **8**

1 Introduction **10**

1.1 Background information and current situation 10
1.2 Purpose of the project 11
1.3 Objectives 11

2 Procedures and methodology **12**

3 Challenge Stage 1 ..... **13**

3.1 Design 13
3.1.1 Comparison metrics process 13
3.1.2 Results templates 14
3.1.3 Publishing 16
3.1.4 Results evaluation 19
3.2 Results and discussion 22
3.2.1 Data Submission 22
3.2.2 Data overview 23
3.2.3 Site complexity score 24
3.2.4 Before score metrics 25
3.2.5 After score metrics 26
3.2.6 Skill versus costs 36
3.3 Summary of Stage 1 40

4 Challenge Stage 2 ..... **42**

4.1 Challenge design 42
4.1.1 Decision Tool design 43
4.2 Results and discussion 52
4.2.1 Data Submission 52
4.2.2 Data overview 52
4.2.3 Site complexity score 53
4.2.4 Before score metrics 53
4.2.5 After score metrics 60
4.2.6 Skill versus costs 60
4.2.7 Summary of Stage 2 63

5 Conclusions **65**

6 Outlook and next steps 66

7 National and international cooperation 66

8 Communication..67

9 Publications 67

10 References...68

11 Appendix..69

11.1 Simulation setups for Stage 1 challenge...69

1 Introduction

1.1 Background information and current situation

This project is an extension of the recently completed SFOE project "A new process for the pragmatic choice of wind models in complex terrain", which involved developing a new decision process for the optimal choice of wind modelling tool in complex terrain. This will ultimately help wind resource assessors improve their processes and reduce the LCOE[1] of wind energy. The project involved applying various simulation tools to five test sites of varying terrain complexity and defining comparison metrics related to the "skill" (or accuracy) of the model as well as those related to its "cost" (or complexity). A resulting plot similar to the one in Figure 1(a) allows the most effective solution to be chosen.

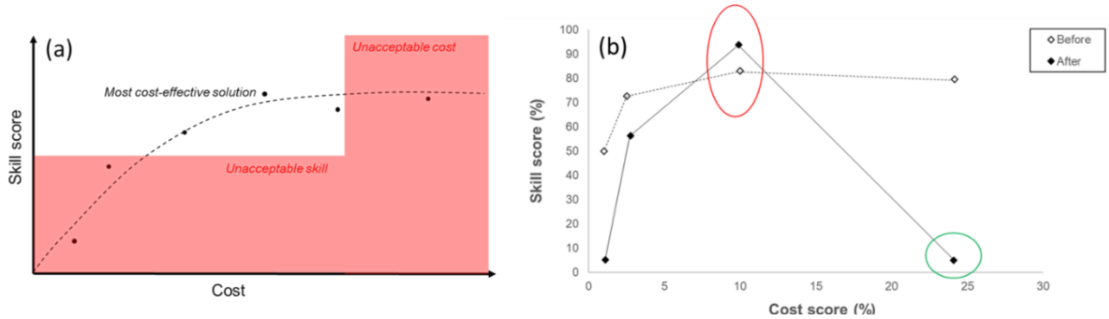


Figure 1. (a) Theoretical cost vs- skill score; (b) Results of initial study

In order to achieve this goal, a large number of simulations with different wind modelling tools and WRA workflows were carried out at four complex terrain wind energy sites, followed by a detailed analysis of the results. The project was carried out by OST and Meteotest in collaboration with the Hochschule Esslingen, who implemented a separate German-funded project with the same goal at two mutual wind energy sites (Stötten and Enercon). This collaboration enabled the development of modelling and WRA workflow methods and the comparison of results with a wider range of different tools than would otherwise have been possible. The wind modelling tools applied in this project by the research partners included *WindPro*, *WindSim* and *PALM* (Meteotest), *Fluent* and *Palabos LBM* (OST), as well as *CFX* (HSE).

The project was split into two main parts: (1) Design of the decision process; (2) Demonstration and validation of the decision process.

For part (1), the test site Stötten in Germany was focused on, because data was available to all project partners including Hochschule Esslingen. In this part, WRAs were carried out using different combinations of wind modelling tools and WRA workflows, and the results compared and analysed in detail. Based on this, automated WRA processes were developed for the used wind modelling tools. Following this, a set of parameters (called "Comparison Metrics") for deciding the most optimal wind modelling tool and WRA workflow were developed. This resulted in several publications as well as a new publicly-available template for estimating the Comparison Metrics.

In part (2), three other sites (ewz, Enercon and ADEV) were used to demonstrate and validate the Comparison Metrics parameters and method. This was done by firstly applying the new automated WRA processes to each of these sites and analysing the results. Finally, a comparison and evaluation of all the results at all four sites enabled the new decision tool to be designed, as well as several improvements to the WRA process and to the Comparison Metrics method to be suggested.

[1] Levelized Cost of Electricity

We are carrying out the present project with the goal to further expand and exploit these ideas to a new level on the international wind energy stage. A close collaboration with IEA Wind Task 31 aims to ensure maximum efficiency of the process and visibility of the results.

1.2 Purpose of the project

In wind energy, the accuracy of the estimation of the wind resource has an enormous effect on the expected rate of return of a project. Due to the complex nature of the weather and of the wind flow over the earth's surface, it can be very challenging to measure and model the wind resource correctly. For a given project, the modeller is faced with a difficult choice of a wide range of simulation tools with varying accuracies and costs. Additionally, different tools have different functionalities - some calculate the entire wind climate (all wind directions) and the energy production, whereas some have to be manually set up to extract this information. Some include mesoscale nesting or forcing, whereas others focus only on microscale features. If the choice of model is made incorrectly, either many resources are wasted in needlessly high accuracy simulations, or the rate of return is inaccurate and investors risk losing large amounts of money. As there are currently no guidelines or tools available to the modeller to help with this choice, it is usually left to gut feeling - and this can be catastrophic for investors or acquirers of wind farms.

1.3 Objectives

In this project, a public "simulation challenge" for wind energy sites in complex terrain was planned in collaboration with IEA Wind Task 31, in which participants submit their simulation data and results in a pre-defined template. The goal was to collect hundreds of comparison metrics data regarding the "skill" and "costs" of simulation tools both before and after carrying out the simulations, enabling transfer functions for the accurate prediction of tool "skill" and "costs" to be developed. This aims to help modellers choose the best model for the job for a given wind energy project.

The targets and expected results are shown in Table 1**Fehler! Verweisquelle konnte nicht gefunden werden.**.

Table 1. Project targets and expected results.

Target	Expected results
• Prepare the results template in tabular form.	• Template completed.
• Publish the simulation challenge Stage 1 (well-defined open data complex terrain site).	• Challenge published.
• Collect and evaluate the results of the simulation challenge Stage 1.	• Hundreds of data points received. • Results evaluated (see Section 2.2 for expected results).
• Publish the simulation challenge where participants can use their site of choice (Stage 2).	• Challenge published.
• Collect and evaluate the results of the simulation challenge Stage 2.	• Hundreds of data points received. • Results evaluated (see Section 2.2 for expected results).
• Integrate the work into IEA Wind Task 31.	• Improvement of process efficiency by sharing experiences. • Generation of more interest and visibility of project and thus more data.

2 Procedures and methodology

The following two challenge stages were planned for this project:

Stage 1: Open data complex terrain

In this stage, the participants submit their results in a pre-prepared table allowing us to calculate weighted parameters related to the skill scores and costs both before and after carrying out the simulations. The simulation case is clearly defined to allow all the results to be compared with each other. All the results are plotted on one graph and may look something like Figure 2(a) below, where each point represents one tool. The clusters are expected due to different category of tool (e.g. linear model, RANS-CFD, LES-CFD). As can be seen, there will be a discrepancy between the metrics predicted beforehand and those determined using the results of the simulations. Transfer functions to better predict the skill scores and costs are developed based on these results

Stage 2: Test cases of choice

In this stage, the participants submit their results in a pre-prepared table allowing us to calculate weighted parameters related to the skill scores and costs both before and after carrying out the simulations. However, in this case, no particular test case is pre-defined, allowing us to collect a much wider range of different sites and external conditions. The results are expected to be clustered according to different categories of input conditions and may look something like Figure 2(b). In this figure, only the best-fit lines through the data are shown, and an example for only two different categories is shown for simplicity. Transfer functions to better predict the skill scores and costs are developed based on these results.

Stage 1 is discussed in Chapter 3, and Stage 2 in Chapter 4.

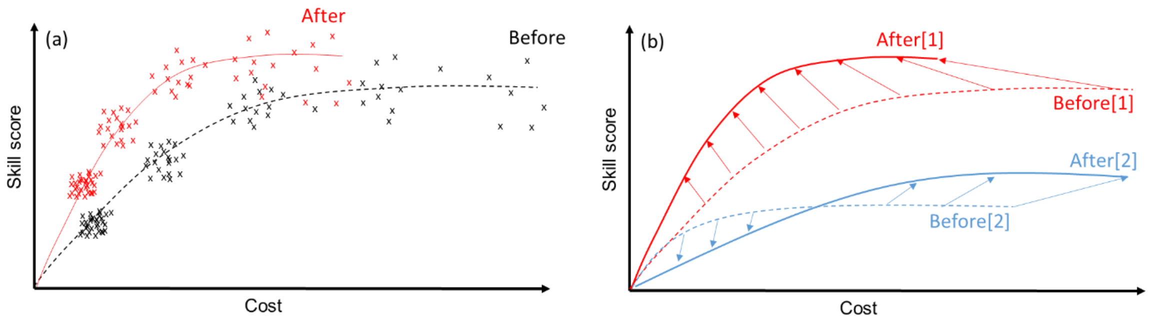


Figure 2. (a) Simplified possible result of Stage 1; (b) Simplified possible result of Stage 2.

3 Challenge Stage 1

In this chapter, the design of Stage 1 is first discussed, followed by the results and then a summary of the learnings used for the design of Stage 2.

3.1 Design

3.1.1 Comparison metrics process

Based on the results of the initial Bolund Hill study and the previous project [1], the Comparison Metrics were studied further. Additionally, some useful inputs were provided by the wind energy community at the IEA Wind Task 31 annual meeting at Amherst, USA in October 2019. The results of these activities were (and are described in [5] and [6] in more detail):

- Definition of a **new Comparison Metrics method** for estimating the skill and costs scores before and after carrying out the simulations, shown in Figure 3 below.
- A **final definition of the parameters used to estimate the skill and cost scores** for both wind speed and Annual Energy Production (AEP).
    - For the **skill scores**, a lot of thought has gone into the parameter definition in order to be able to include all the steps in the WRA process in the scores. This includes parameters related to the wind model, the input data quality, the calibration and validation methods, the AEP calculation method, the skill of the user and the robustness of the model.
    - For the **cost scores**, the so-called "Actual Total Costs" are estimated by splitting the costs up into the categories described in the table and adding up the totals. The software costs are estimated by dividing the total license and support costs by the number years of usage, and dividing this by the number of projects carried out per year. The time to learn and training costs are estimated by adding the staff costs for the time taken to learn how to use the tool to any training costs, and dividing this by the number of projects carried out per year and the estimated number of years of usage. The simulation set-up effort costs are estimated by recording the number of hours required to set up the simulations and multiplying this by the hourly staff rate (for all calculated wind directions). The simulation run time costs are estimated by recording the run-time of the simulations and the number of cores that they were run on, and multiplying this with the computational cost per core per hour (for all calculated wind directions). The post-processing effort costs are estimated by recording the number of hours required to post-process the results and multiplying this by the hourly staff rate (for all calculated wind directions). In order to compare the results with each other fairly, the following parameters should be norrnalised: Number of years of usage of the software; Number of projects per year; Staff hourly rate; Computational cost per core per hour = \$0.04/hour/core; Number of cores; Processor clock speed = 2 GHz.

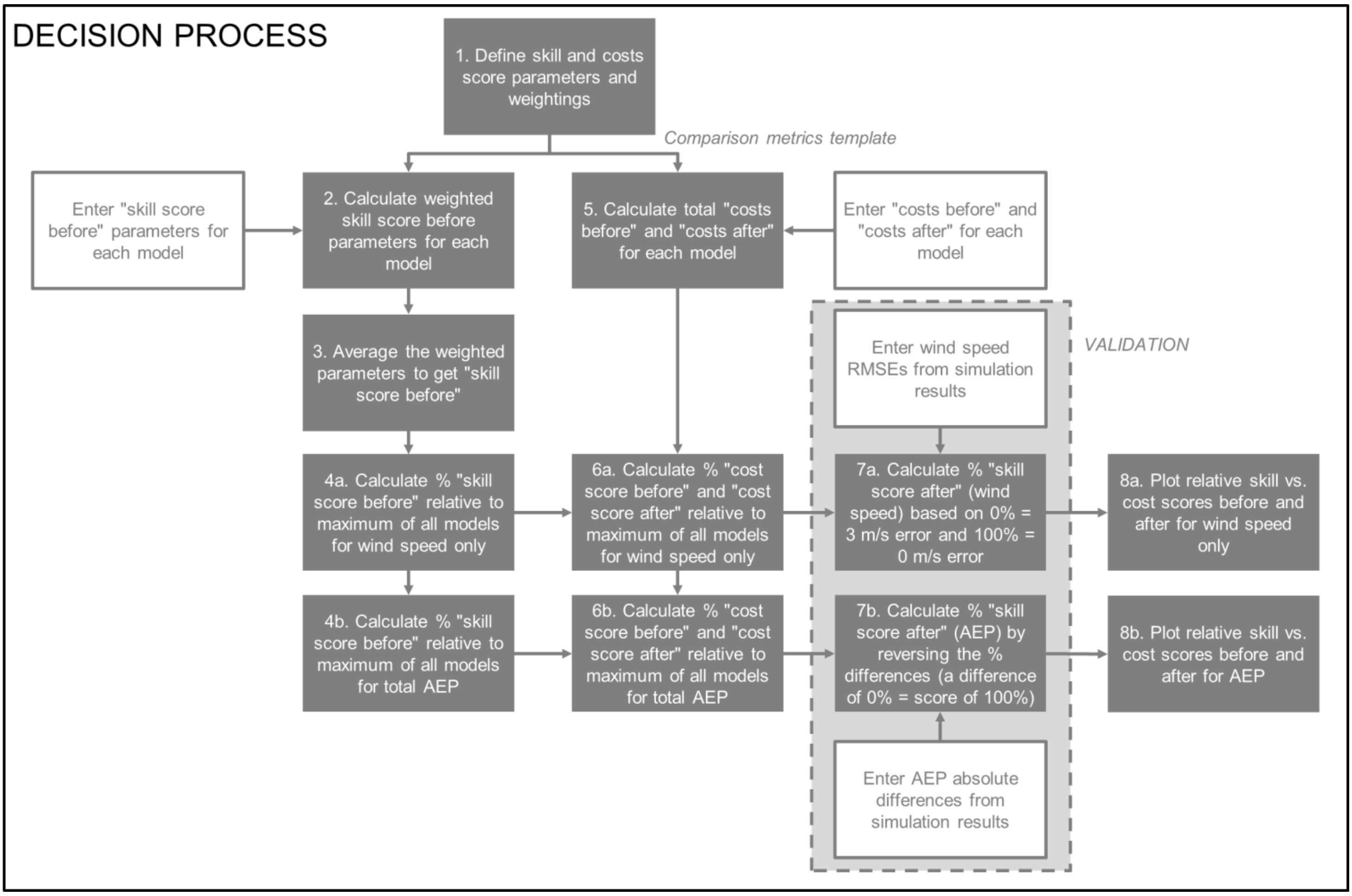


Figure 3. Comparison Metrics method

3.1.2 Results templates

In the first three months of the project, the Excel template that had been used for collecting comparison metrics in the previous project [1] was published and made available to the wind energy community. Furthermore, the results of this project were presented at the IEA Wind Task 31 yearly meeting in Amherst, MA (USA) in October 2019. Following some exchanges with the task members and with Javier Sans Rodriguez, IEA Wind Task 31 work package leader, it was decided to convert the Excel table to the following different Google Forms surveys in order to make it more accessible to the users:

- **Registration form:** participants register, set their confidentiality requirements and receive naming conventions for the rest of the forms;
- **Model description form:** participants submit descriptions of their simulation set-ups;
- **Parameters 'before' form:** participants submit estimations for relevant parameters between pre-defined limits related to skill and cost scores before carrying out the simulations, and the scores are calculated via a Python code. Additionally, participants can enter their own parameter weightings, which are compared to the pre-defined values and possibly adjusted, if required;
- **Parameters 'after' form:** participants submit estimations for relevant parameters between pre-defined limits related to cost scores after carrying out the simulations, and the scores are calculated via a Python code;

- **Results upload link:** participants submit their simulation results in a pre-defined format, and the skill scores are calculated via a Python code.

The exact details and links to these forms can be found in an article by Sarah Barber in The Wind Vane Blog [4].

As well as making it more accessible, the following changes were made to the content of the comparison metrics template:

- In the previous version of the template, some parameters for estimating the skill scores were related to the complexity of the terrain. However, these parameters should not be part of the skill score calculation because they do not vary between models, and the goal is to compare different models. Instead, a new section for classifying the terrain complexity was created in the model description form. For this, new parameters were defined based on previous work on terrain classification related to lidar measurements [5]. This splits complex flow into the following categories: (a) Complex terrain (e.g. definition in IEC 61400-12-1 [6]); (b) Surface roughness (e.g. forested land, changes in ground cover); (c) Presence of obstacles (e.g. buildings, towers and wind turbines); (d) Local meteorology (e.g. low-level jets, divergent flows and fronts). The parameters used for this work should not be too difficult or time-consuming to calculate, and the site classification given in IEC 61400-12-1 and 61400-12-2, for example, are too complex for this application. Therefore, the following parameters were used here in order to simplify the process. The goal of this section in Stage 1 was to test, improve and validate the classification method for use in Stage 2, in which the site classification is key:
    - General terrain complexity - how steep are the slopes on average?
    - General terrain complexity - how many slopes are there?
    - Validation mast position - in how many 30° sectors is there a positive slope steeper than 30° less than 250 m away from the validation position in any direction?
    - Surface roughness complexity - approximately how many different surface roughness regions are you using?
    - Surface roughness - how rough is the surface in general?
    - Atmospheric stability - what is the average value of the vertical temperature gradient? (if relevant)
    - Atmospheric stability - are low-level jets present?
    - Degree of turbulence - what is the approximate Reynolds number, calculated based on the input flow velocity and the distance from the inlet to the calibration met mast?
- For the definition of the skill scores, some parameters related to the intended operational envelope of the model was added. Specifically, this reduces the skill score if WAsP or another linear model is applied in a terrain complexity for which the model is not intended, based on studies that quantify the expected increase in uncertainty [7]. Additionally, a parameter related to the wind speed calibration method was added. For example, a lower skill score is obtained if the results are scaled linearly for the average wind speed at one calibration mast position and height than if the process is carried out for different wind sectors and measurement heights. Finally, further parameters that are used to calculate a separate skill score for the AEP calculations were added. These include the wind speed extrapolation method, the method of taking account of different wind speeds, and the long-term AEP extrapolation method. All the parameters that are used for calculating the skill score can be accessed at [8].
- For the definition of the cost scores, the number of simulations in total carried out for obtaining the AEP were added. All the parameters that are used for calculating the cost score can be accessed at The Wind Vane Blog [4].

3.1.3 Publishing

In order to publically release the challenge Stage 1, the following steps were first carried out:

1. Final definition of challenge goal.
2. Identification of a suitable site.
3. Definition and preparation of the input and validation data sets.
4. Development of a process allowing participants to enter their results both for predicted and actual cost and skill scores
5. Definition of the data to be submitted by the participants.
6. Choice of the data format and storage platform.

Following the final design of the challenge, it was published on the website "The Wind Vane Blog" [4], which is run and managed by Javier Sans Rodrigo, leader of IEA Wind Task 31. This link was then shared on LinkedIn and sent to as many contacts as possible. Additionally, a launch webinar was carried out on April 7th, 2020, which attracted more than 20 attendees.

As well as this, a poster was presented at the WindEurope Wind Resource Assessment Workshop in June 2020 (online) and a poster and paper were presented at the Torque2020 conference in September 2020 (online) in order to attract attention.

The above-mentioned steps are discussed further below:

1. Final definition of challenge goal.

The goal of Stage 1 of this challenge was to collect comparison metrics data regarding the skill and cost scores of a range of different simulation tools for a complex terrain site, both before and after carrying out new simulations. The results are expected to look something like Figure 2(b) on page 12, where each point represents one tool. The clusters are expected due to different categories of tool. A discrepancy between the metrics predicted beforehand and those determined using the results of the simulations is expected. Transfer functions to better predict the skill and cost scores will be developed based on these results.

2. Identification of a suitable site

The Perdigão site in Portugal [9] was chosen for Stage 1 of the challenge, due to the volume and quality of available measurement data, the complexity of the terrain and the relative lack of simulations already carried out. A large measurement campaign was undertaken between December 2016 and June 2017 as part of a large EU-US collaborative field experiment [10]. This is the ideal situation for a new challenge, because interest in taking part is therefore expected to be high. Measurement data from many met masts as well as perhaps from an operating wind turbine is available [9]. The site consists of flow over two parallel ridges with SE-NW orientation, which are 4 km long and 500-550 m tall and separated by about 1.5 km. The two main wind directions are approximately perpendicular to the ridges. A 3D representation of the site as well as an overview of all the measurement sensors is shown in Figure 4.

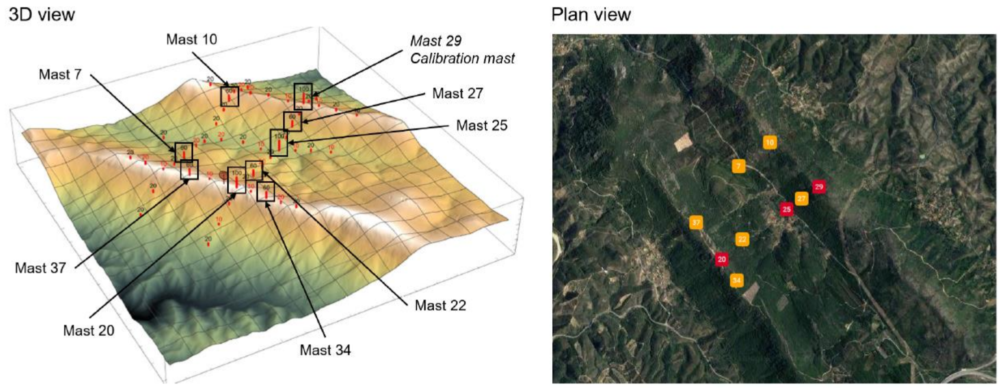


Figure 4. Overview of the Perdigão site: 3D view (left) and plan view (right).

3. Definition and preparation of the input and validation data sets

The input and validation measurement data for the simulation challenge was chosen by firstly downloading all the available ten-minute averaged wind data and assessing its quality and availability. It was decided to focus on the data from the nine 60 m and 100 m high masts; numbers 7, 10, 37, 22, 27 and 34 at 60 m and numbers 20, 25, 29 at 100 m. The three 100 m masts are positioned on a straight line along the main wind direction as can be seen in Figure 5(a), which also shows the position of the wind turbine (WTG). The mean measured wind profiles for these three masts over the entire measurement period are shown in Figure 5(b) together with logarithmic ts using the measured wind speeds at 20 m and 100 m for fitting purposes. It is clear to see that the wind speed is much lower at mast 25, which is expected due to its location between the two ridges. The measured wind roses for masts 25 and 29 at 100 m and 40 m are shown in Figure 6. The main flow directions for mast 29 are SW and NE, agreeing with previous analysis [9]. This previous analysis showed that the main wind direction is the SW direction, and a mesoscale circulation leads to flow from this direction actually entering the simulated region from the NE direction at certain times of day. Additionally, the presence of the valley forces wind to travel up it in a SSE direction, reflected in the wind roses for mast 25, which is positioned in the valley.

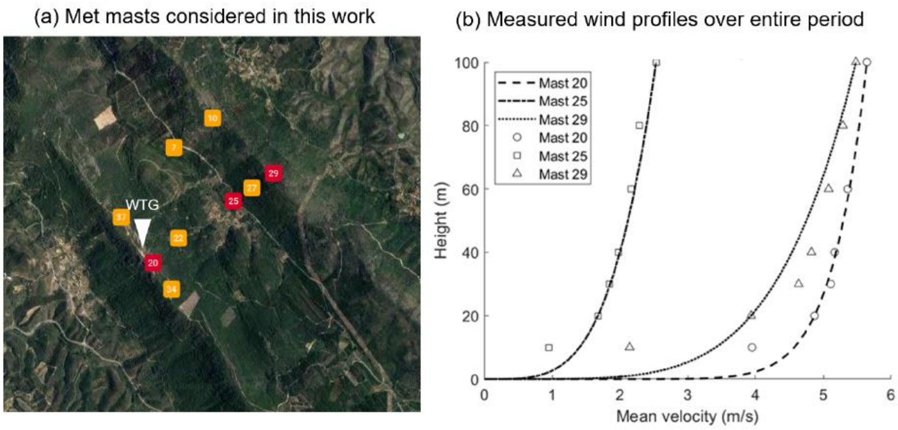


Figure 5. (a) Met mast considered in this work (plan view from Google Maps); (b) Measured wind profiles over entire measurement period, Mast 29, Mast 25 and Mast 20.

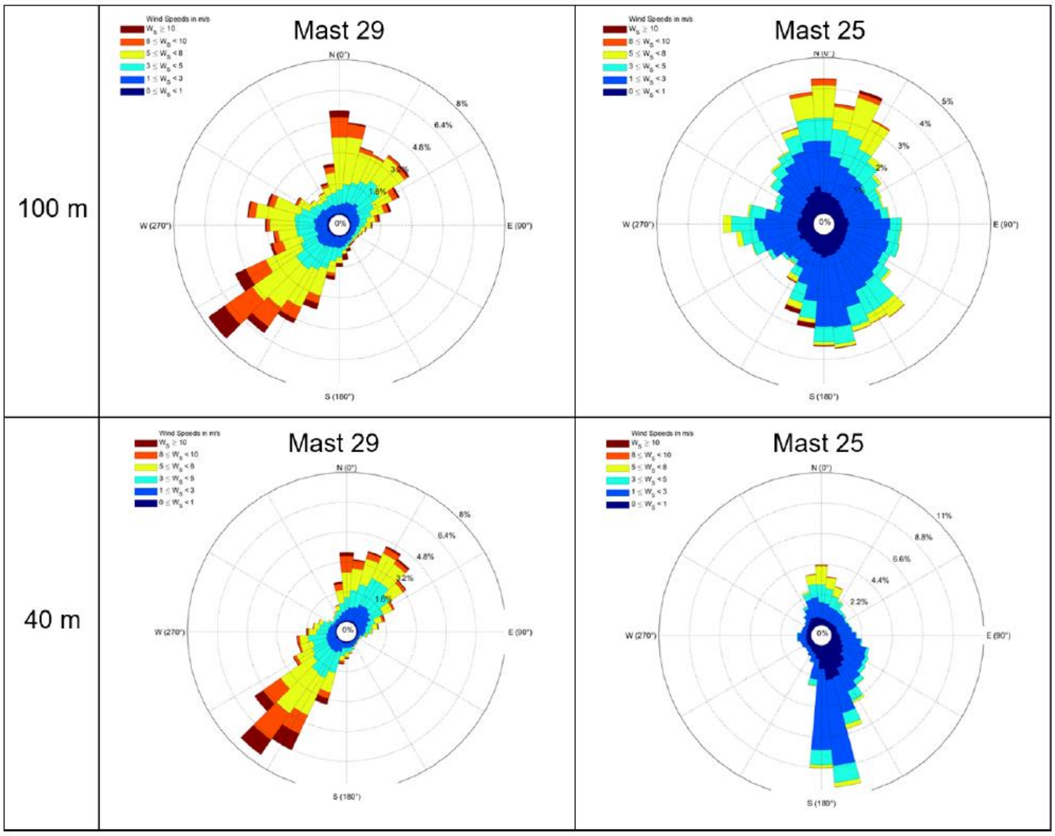


Figure 6. Measured wind roses for Mast 29 (left) and Mast 25 (right), both at 100 m (top) and 40 m (bottom) heights.

In wind resource assessments, CFD simulations are typically calibrated by linearly scaling the simulation results in order to achieve the wind speed that equals the wind speed at a 'calibration mast'. The accuracy can then be assessed by comparing the scaled simulation results to measurements at a different location ('validation mast'), which is ideally far away from the calibration mast. In order to reduce calibration inaccuracies, it is important to choose a met mast location for the input data that represents the wind behaviour at the boundaries of the simulated domain as well as possible. Therefore one of the met masts on the ridge should be used. Mast 29 was chosen for the input data ('calibration mast') due to its distance away from the wind turbine (marked on Figure 5), in order to allow validation using the wind turbine data as well as the validation masts.

As different wind turbines could be positioned in various locations in a wind resource assessment, it was decided to take all eight remaining masts (7, 10, 20, 22, 25, 27, 34 and 37) as the 'validation masts', in order to assess the capabilities of different tools for calculating flow in separated regions (mast 25, 27, 7 and 22) as well as on top of hills (masts 10, 37, 20 and 34). A data period of 02.02.2017- 15.06.2017 was chosen in order to ensure overlapping time periods between all the masts. As well as the measurement data, the following other input data will be provided to the modellers:

- Topography and roughness maps;
- Description and set-up of measurement equipment;
- Wind turbine height, coordinates and power curve;
- A Python script for writing data to NetCDF format correctly.

- The exact details and links to the data can be found at The Wind Vane Blog link on page 9. As this is a blind test, the validation data will only be provided after the challenge window has been closed.

4. Development of a process allowing participants to enter their results both for predicted and actual cost and skill scores.

This was adapted as discussed in Section 3.1 above.

5. Definition of the data to be submitted by the participants.

For each simulation run, participants will be asked to provide 3D wind vector components of vertical wind speed profiles at each validation met mast and the wind turbine location for each 30° wind direction sector, the calculated AEP in each sector, horizontal planes of 3D wind speed vectors at 100 m and 40 m above ground, as well as vertical planes through each validation met mast in the SW direction. The upload link for submitting the results was only sent after submitting the model description, parameter `before' and parameter `after' forms.

6. Choice of the data format and storage platform.

It was decided to provide and submit data in NetCDF format, the details of which are given in The Wind Vane Blog article [4].

3.1.4 Results evaluation

In order to evaluate the submitted results, a Python package was developed with the elements shown in Table 2. The current version of the package can be found on Gitlab[2] (access can be granted on request). The current status of the code is shown in the right-hand column of the table. The main part as well as the detailed analysis part of the code have been finished.

[2] https://gitlab.com/windenergie-hsr/pragmaticchallenge/comparisontool

Table 2. Python code elements and status.

Step	Task
1. Preparation	Read in assigned/submitted simIDs*
	Create offsets for coordinate system (see Model Description)
	Read in met mast data of all validation met masts
	Filter all metmast data and create .nc files
2. Read submission data	nameID_modelID_simXX_z
	nameID_modelID_simXX_xy
	nameID_modelID_simXX_xyz
3. Basic handling	Correct for coordinate system used
	Check plausibility of data (check met mast positions, correlate the data, compare to a baseilne, correct for outliers)
	Calculate absolute velocities from wind speed components (all files, all sectors)
	Calculate angle (horizontal) and angle (vertical) of flow from wind speed components (all files, all sectors)
4. Analysis of nameID_modelID_simXX_z	For each met mast and the WTG, plot one graph of absolute wind speed vs. z for each sector and for the average compared to met mast data
	Repeat for wind direction (horizontal angle)
	Repeat for wind direction (vertical angle)
	Calculate a baseline AEP value in order to make comparisons later (no measurements available)
	Calculate the theoretical AEP at the WTG position using the frequency distribution of met mast 29 in each sector, the average wind speed simulated at that position in each sector and the power curve provided (1. Using wind speed at hub height (interpolate if necessary). 2. Using Rotor-Equivalent Wind Speed)
	Calculate the capacity factor (AEP / (365*24*rated power))
	For each met mast and the WTG, interpolate the absolute wind speed and directions and calculate values at exact met mast heights (for the RMSE calculation in the next step)
	Based on these interpolated values, calculate the RMSE of the absolute wind speed and the two directions (horizontal and vertical) compared to met mast, separately for each sector: 1. Using all heights. 2. Only using heights covering rotor area. 3. Only using hub-height (= absolute difference)
	For each of the three methods, average the RMSE values over all sectors: 1. Unweighted average. 2. Average weighted using frequency distribution of met mast 29
	For each met mast and the WTG position, plot a bar chart of % difference in wind speed between simulation and measurement vs. height (see example on right)
	Average the RMSE over all heights and make a polar diagram of these values vs. sector (see example on right)
	Calculate absolute and % difference between calculated and baseline AEP for each sector and for total
	Make a polar diagram of these values vs. sector (same as RMSE polar diagram)
5. Analysis of nameID_modelID_simXX_xy	For each xy plane, plot contours of absolute wind speed and directions. Note: No data available for this task.

6. Analysis of nameID_modelID_simXX_xyz	For each met mast xyz plane, plot contours of absolute wind speed and directions. Note: No data available for this task.
7. Comparison between simIDs of same organisation	For each met mast and the WTG position, for each sector plot a bar chart of RMSE wind speed vs. simulation number (RMSE calculation 1. Using all heights. 2. Only using heights covering rotor area. 3. Only using hub-height (= absolute difference))
	Repeat for average RMSE over all sectors: 1. Unweighted average. 2. Average weighted using frequency distribution of met mast 29 (for wind speed and both angles)
	Repeat for AEP (% difference to baseline)
8. Comparison between all organisations	Same as 7 but for all results
9. Detailed comparisons	Timestamp binning
	Speed up factors and flow turning between metmasts.
10. Documentation	Write API
	Comment the code
	Create documentation

3.2 Results and discussion

In the following sections, the results for Stage 1 are presented. First, in Section 3.2.1 we look at the submitted data of participants outside of OST. Additionally to the submitted data, we simulated the site with two different tools, WAsP and Ansys Fluent. An overview of the whole dataset is given in Section 3.2.2. The site complexity score is presented in Section 3.2.3. Having gathered data from various tools and participants, the before and after score metrics described above can be calculated. The rationale behind this process is described in Sections 3.2.4 and 3.2.5. Lastly, the skill versus cost score plots for the wind resource assessment are presented in Section 3.2.6.

3.2.1 Data Submission

Following the launch of Stage 1, we received a total of 20 registrations from the companies using the tools shown in Table 3:

Table 3. Submissions to Challenge Stage 1.

Organisation	**Country**	**Tool**	**Status**
Von Karman Institute of Fluid Dynamics	Belgium	OpenFOAM	Submitted
Freelancer	France	ZephyCFD	Not submitted
US Forest Service	USA	WindNinja-COM and WindNinja-CFD	Not submitted
JH Wind GmbH	Germany	WASP, ZephyCFD	Not submitted
IFPEN	France	waLBerla	Not submitted
GEO-NET Umweltconsulting GmbH	Germany	FITNAH-3D	Not submitted
UL Renewables	Spain	WRF	Not submitted
ZHAW School of Engineering	Switzerland	PALM	Not submitted
The University of Texas at Dallas	USA	UTD-WF	Submitted
Wind Engineer	Singapore	??	Not submitted
Srh hochschule Berlin	Germany	Fluent	Not submitted
ALTEN	Spain	WAsP	Submitted
EDFR	France	Meteodyn	Not submitted
EMD International A/S	Denmark	WAsP-IBZ, WAsP-CFD, OFWind-CFD	Not submitted
Meridian Energy Limited	New Zealand	Meteodyn WT	Not submitted
GE	Germany	OpenWind	Not submitted
EDF Renouvelables	France	Meteodyn WT	Not submitted
ENERCON GmbH	Germany	E-Wind	Submitted
Eindhoven University of Technology	Netherlands	??	Not submitted
ZephyScience	France	E-Wind	Not submitted

In the final column of this table, the submission status can be seen. Unfortunately, we did not receive as many submissions as expected. The other benchmarks developed within IEA Wind Task 31 also suffered similar problems in the same time period. The correspondence with the participants indicated the following reasons for the lack of submissions:

- Many people find the challenge very interesting; however, the industrial focus means that the companies signed up usually have a higher task on the priority list. Research is less important than their daily business. The COVID-19 situation did not help with this.
- Some people had difficulties reading in the input data: we decided upon the NetCDF format following recommendations from IEA Wind Task 31; however, this type of file is not easy to

understand and deal with for non-programmers. To overcome this, we provided a guideline and sample code for users.

- We did not provide the input data in sufficient length or quality: some participants complained that the data wasn't pre-filtered or long enough to calculate the Annual Energy Production.
- The orientation of the wind speed vectors in the input data provided by us was confusing. We provided a sketch that should have allowed a correction; however, it was difficult to understand
- The required format of the submission data was confusing (NetCDF). We eventually decided to provide Excel templates to overcome this; however, this is not an ideal solution.
- Some of the more detailed required submission data scared off some participants. We had originally required the participants to submit several different wind speed planes in order to attract academic partners; however, this was too challenging and time-consuming for the industry participants. We overcame this by making some of the submissions optional, in order to try and please both types of participant.

Two participants (EMD and EDFR) even dropped out following a number of phone calls and discussions with them.

This process has allowed us to learn a lot about public simulation challenges, which we was used for designing Stage 2.

In total four external participants (Von Karman Institute of Fluid Dynamics, The University of Texas at Dallas, ALTEN, Enercon GmbH) submitted sufficient data for determining the skill and cost scores. However, each submission represents a different category of tool (e.g. linear model, RANS-CFD, LES-CFD), helping to highlight the mentioned clusters in the skill vs cost score figure. As well as this, we submitted our own results using two different tools in order to ensure enough data points for the evaluation.

3.2.2 Data overview

An overview and details of the data used to determine the skill and cost metrics are shown in Table 4. A more extensive overview can be found in the Appendix in Table 11. In total five different simulation tools from six different organisations were used with various configurations and conditions resulting in a total of ten different simulations. In the following sections the various simulations are denoted by the name given in the respective table column. Names starting with *orgXX* are used for organisations that prefer to stay anonymous in future publications.

Table 4. Overview and details of the submitted data

Name	Organisation	Model	Description
UTD_UTD-WF_LES	University of Texas at Dallas	UTD-WF	In-house high-resolution large-eddy simulation code
UTD_UTD-WF_LES-canopy	University of Texas at Dallas	UTD-WF	With canopy model
org01_OpenFOAM_k-epsilon-structured	Von Karman Institute	OpenFOAM	k-epsilon turbulence model Structured mesh
org01_OpenFOAM_k-epsilon-unstructured	Von Karman Institute	OpenFOAM	k-epsilon turbulence model Unstructured mesh
org04_wasp_sim01	Alten	WAsP	
ENERCON_E-Wind_k-epsilon	Enercon	E-Wind	k-epsilon turbulence model
ENERCON_E-Wind_k-L	Enercon	E-Wind	k-L turbulence model
ENERCON_E-Wind_k-omega	Enercon	E-Wind	k-omega turbulence model
ost_fluent_k-omega-SST	OST	Fluent RANS	k-omega-SST turbulence model In-house automated workflow
ost_wasp_sim01	OST	WAsP	

3.2.3 Site complexity score

First of all, the site complexity score for the Perdigao site had to be determined. This makes the results comparable to the other simulation sites mentioned in [1] as well as the sites submitted for Stage 2. To do so each participant filled out a questionnaire with questions regarding the terrain, the surface roughness and the complexity of local flows. The answers are summarised in Figure 7. The first thing to notice is the large variation of scores for some of the questions. This might either indicate that the question is too generic or too hard to quantify objectively leading to these discrepancies.

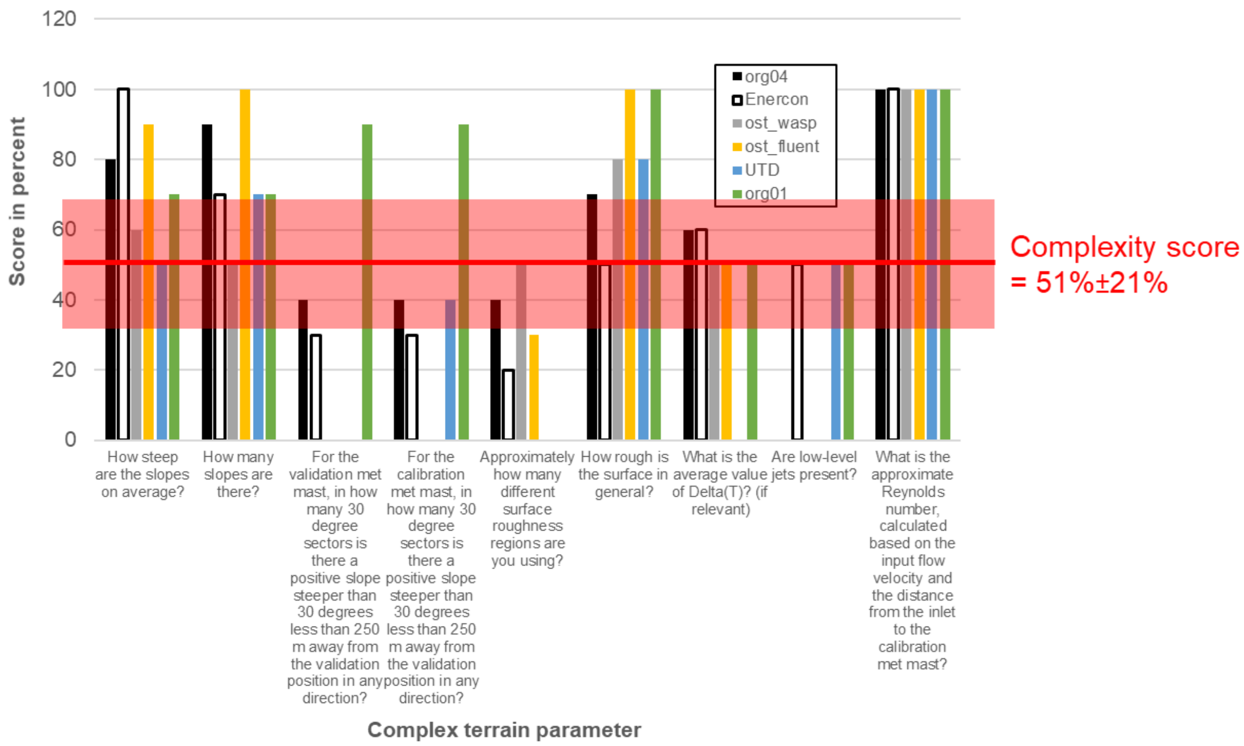


Figure 7. Scores given to each question by participants.

The complexity score is then calculated by a sum of each of the given scores and comprises two parts, namely the mean complexity value and the complexity maximum deviation. The higher the mean score, the more complex the site. A high maximum deviation, however, indicates two important points. Either the questions themselves are not precise enough, which leads to the observed discrepancy, or the site is so complex that it is hard to pin down an accurate score for some of the questions. Hence, the questions with large maximum deviations have to be reconsidered for Stage 2 in order to be able to make this distinction.

Given the method above, the resulting complexity score for the Perdigao site considering all questions is 51% ± 20%. The value is marked with a red line and the range with a light red box on Figure 7.

3.2.4 Before score metrics

Similar to the complexity score, the before skill and cost scores were determined based on the questionnaires filled out by each participant and their respective simulations. The questions used to determine the before skill scores are shown in Figure 8. The scores for each of these questions was additionally weighted based on its importance. In further iterations the determination of the weights could, however, be done by formulation of an optimisation problem, where the before skill scores are tried to be matched with the after skill scores. More on this in the Stage 2 section.

Skill A. Model parameters
A1. Mathematical model - what is the complexity of the underlying aerodynamic equations?
A2. Mathematical model - what is the complexity of the underlying thermodynamic equations (including temperature and coriolis force modelling)?
A3. Time step - how well does your chosen time step conform to the recommendations for your model and set-up?
A4. Simulation period - how well does the simulation period reflect the entire measurement period?
A5. Grid quality - how well are the software recommendations regarding orthogonal quality, skewness and aspect ratio fulfilled?
A6. Grid quality - has grid independency been proven via a grid study?
A7. Grid quality - are the estimated y+ values within the software recommendations of the applied boundary layer model?
A8. Domain size - how well do the domain dimensions allow the flow to be properly captured?
A9. Convergence - how well does your model convergence meet the required convergence criteria?
A10. Operating envelope - if you are using WAsP or a linear model, what is the average dRIX value for all wind sectors at the validation mast 1?
A11. Operating envelope - if you are using WAsP or a linear model, what is the average dRIX value for all wind sectors at the validation mast 2 (if applicable)?
A12. Operating envelope - if you are using WAsP or a linear model, what is the average dRIX value for all wind sectors at the validation mast 3 (if applicable)?
Skill B. Input data quality
B1. Measurement data - how do you assess the general quality of wind data you have been provided with?
B2. Measurement data - how well did you filter the measurement data?
B3. Quality of terrain data - what is the ratio of horizontal resolution as a % of your domain length?
B4. Quality of land cover data - how well does the land cover data you have been provided with depict the real situation?
B5. Quality of atmospheric data - how do you assess the general quality of atmospheric data you have been provided with?
Skill C. Wind model calibration and validation
C1. Validation of model - if you have previously validated the wind model for other sites, what was the average deviation between simulations and measurements for a site of similar complexity?
C2. Wind speed calibration quality - which wind speed calibration method did you use? (please describe in detail in the model description)
C3. Wind speed calibration quality - how well does your input data method capture the wind profile actually measured at the calibration mast?
Skil D. AEP calculations
D1. Wind sectors - which wind sector extrapolation method are you applying?
D2. Wind sectors - how are different wind speeds in different sectors taken into account?
D3. How well did you account for flow turning?
D4. Long-term AEP - which long-term extrapolation method is applied?
D5. Rotor area extrapolation - which rotor area extrapolation method is applied?
Skill E. Other parameters
E1. Skill of user - how many years of experience does the user have with the applied model?
E2. Skill of user - on a scale of 0-100%, how do you rate your own ability to apply the chosen model correctly?
E3. Robustness of model - what is the status of the applied model?

Figure 8. Questions for determining the before skill score metrics

The results of the questionnaire are presented in Figure 9. In order to get the final before skill scores, the weighted sum for each individual simulation is calculated. The concrete values are given in the comparison plots in Section 0.

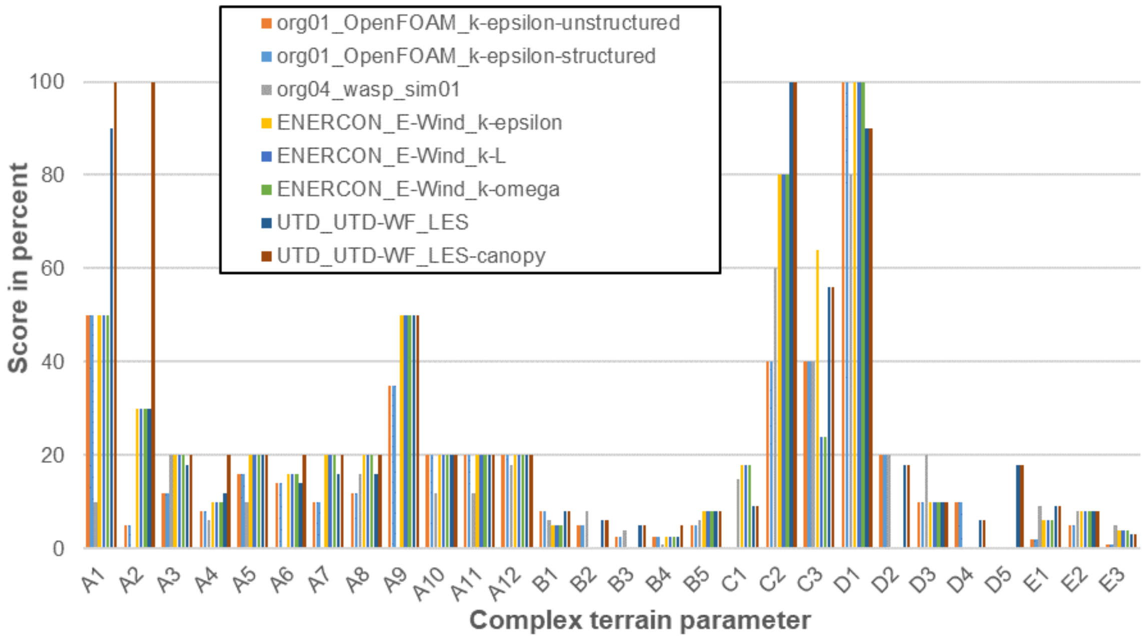


Figure 9. Results for the before skill score questionnaire

3.2.5 After score metrics

For each simulation, 3D wind speed components along a vertical line for the nine met masts as well as AEP values per 30° sector at two met mast positions were submitted. These values are used to calculate the after skill scores. Before moving on to these scores, the specific results are discussed in more detail in the following. This will help to better understand how the scores based on wind speed and annual energy production were determined.

For the sake of brevity, only met masts 29 (calibration), 25 and 20 are considered. To give an overview of the wind conditions, the wind roses for each met mast are shown in Figure 10. Considering met mast 29 and met mast 20 is very interesting in that they are in line with the main wind direction, but on different ridges (see Figure 5). Met mast 25 is located in the valley and is separated by the ridge from met mast 29.

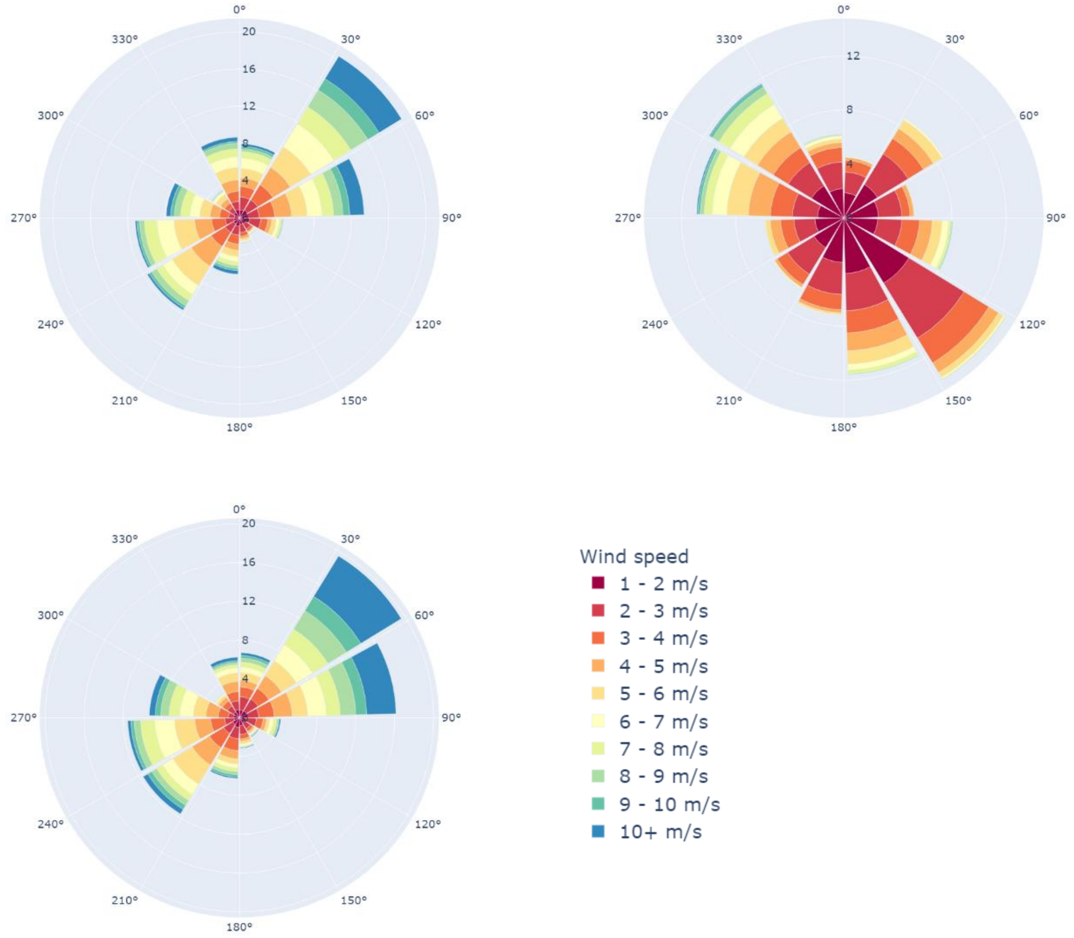


Figure 10. Wind roses for met mast 29 (top left), met mast 25 (top right) and met mast 20 (bottom left)

Wind speed scores

Figure 11, Figure 12 and Figure 13 show the root mean squared error (RMSE) values between the simulated and the measured wind profiles for different evaluation heights for met mast 29, met mast 25 and met mast 20, respectively. The RMSE is defined by

$$RMSE = \sqrt{\frac{1}{N}\sum_{i}^{N}\left(ws_{sim,i} - ws_{meas,i}\right)^2}$$

with the simulated and measured wind speeds, $ws_{sim,i}$ and $ws_{meas,i}$, respectively, at vertical position i. In case of a single evaluation point, e.g. at hub height, this expression simply reduces to the absolute difference.

The dashed lines represent the mean RMSE values for all simulations at the specific height profile. For each of the twelve sectors the RMSE values were calculated for each simulation. The bars represent the mean value the twelve sectors whereas the vertical bars represent the standard deviation.

The first thing to notice is that the lowest errors occur at met mast 29, which is not surprising as it serves as the calibration location. Furthermore, an increase in errors when considering heights closer to the ground can be observed due to the simulated and the measured wind profiles having larger deviations in that region. Here, for example, a special roughness model seems beneficial when looking at met mast 29 and met mast 20 by comparing *UTD_UTD-WF_LES* versus *UTD_UTD-WF_LES-canopy*, where a canopy model was used in the latter case.

At met mast 25 the overall RMSE values are larger than for met mast 29. This is due to the extreme flow turning as visible from the wind roses and hence the increased difficulty to simulate the wind profiles accurately. Moreover, the different evaluation heights have less impact on the error values. The simulations done with E-Wind and OpenFOAM show the lowest errors. The LES simulations show comparatively and unexpectedly large RMSE values and further investigations on the side of the participant is expected. Interestingly, the two WAsP simulations show fairly different simulation performances, which might potentially be due to different experience levels of the modellers.

For the farthest location, met mast 20, E-Wind and OpenFOAM show larger errors compared to met mast 25. The most surprising candidate, however, is *org04* with the WAsP simulations, showing great performance for being the most simple model and also being commonly considerate unsuitable for terrains such as the one seen at the Perdigao site.

These somewhat counterintuitive results at the three met mast positions indicate already the difficulty to draw definite conclusions based on such a small data set.

These results were used for the skill score vs. cost score plots presented in Section 3.2.6.

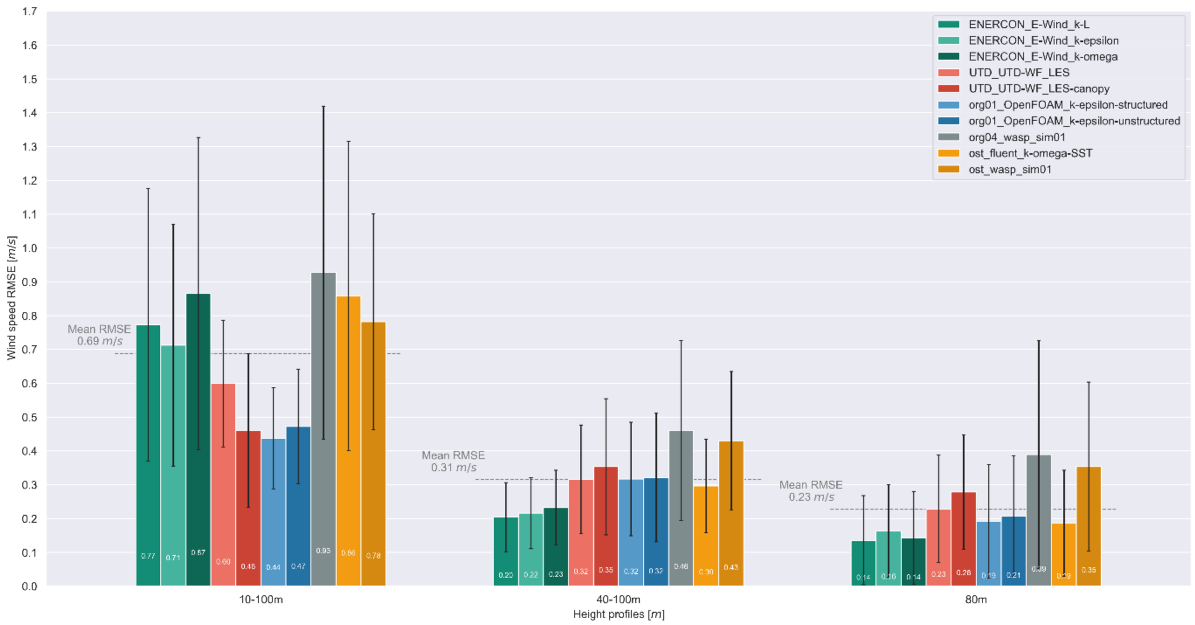


Figure 11. RMSE values between simulated and measured wind profiles for different heights for met mast 29. The dashed lines represent the mean RMSE values for all simulations at the specific height profile.

Speed-up and turning

Continuing with the wind speed results, Figure 14, Figure 15 and Figure 16 show the speed-up factors and flow turning values for each sector of met masts 29, 25 and 20, respectively. These values were calculated with respect to the measurement results of met mast 29, the calibration mast, at height 80 meters. The speed-up factor is defined as follows

$$speed\text{-}up_{metmast,sector} = \frac{Wind\ speed_{metmast,sector}}{Wind\ speed_{29,sector}}$$

and the turning is defined by

$$Turning_{metmast,sector} = Wind\ direction_{29,sector} - Wind\ direction_{metmast,sector}.$$

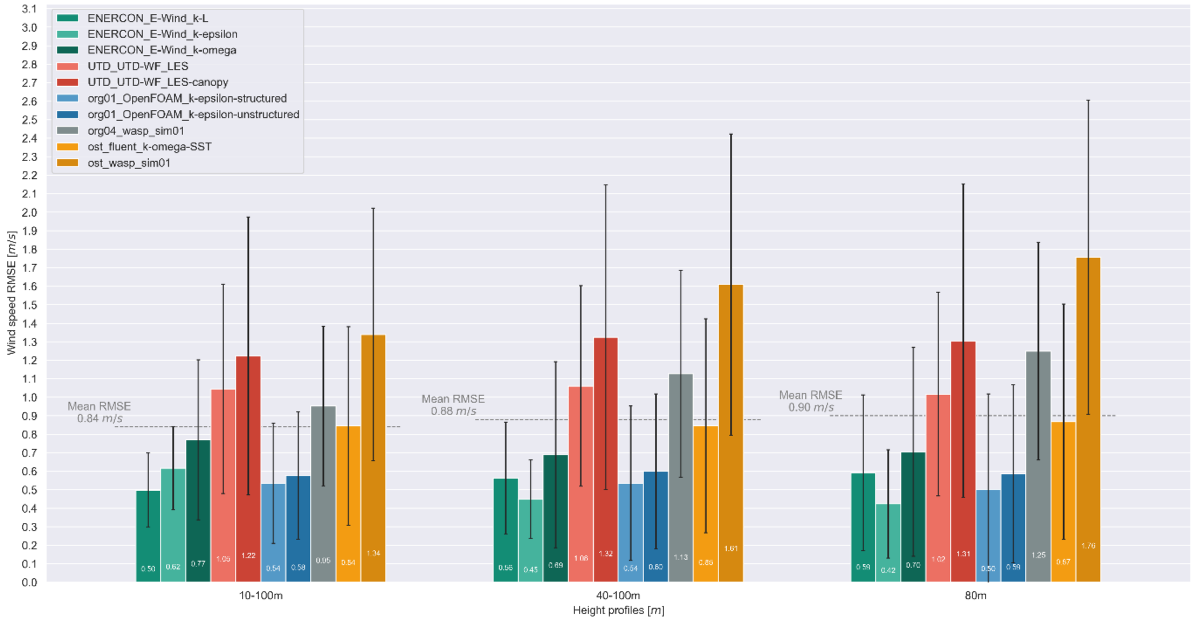


Figure 12. RMSE values between simulated and measured wind profiles for different heights for met mast 25. The dashed lines represent the mean RMSE values for all simulations at the specific height profile.

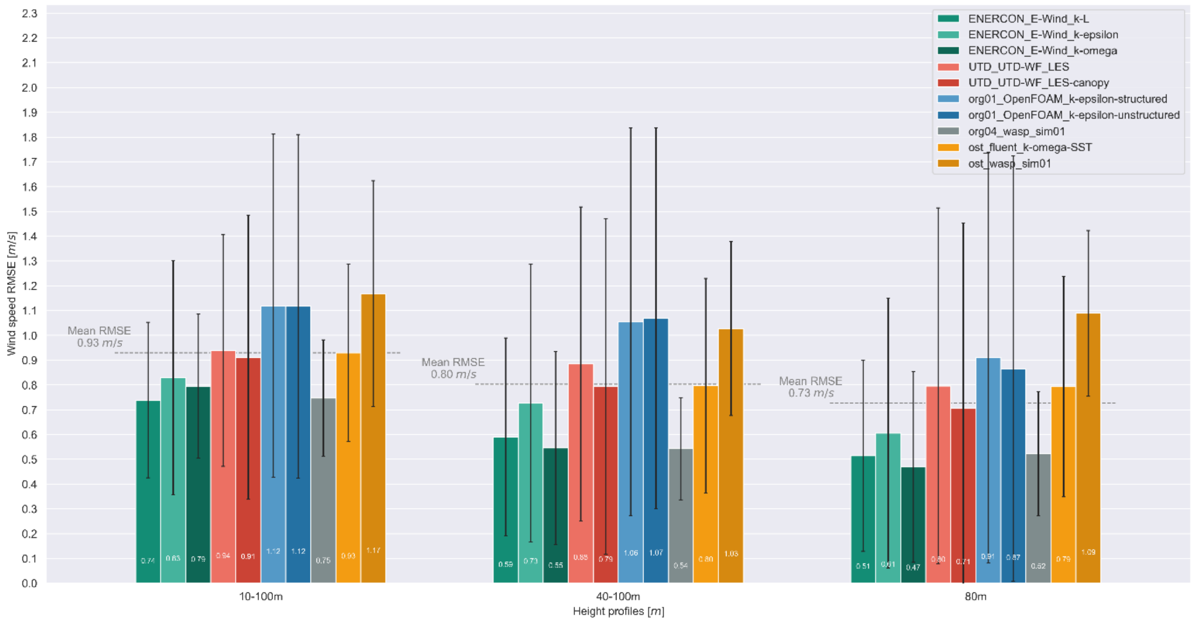
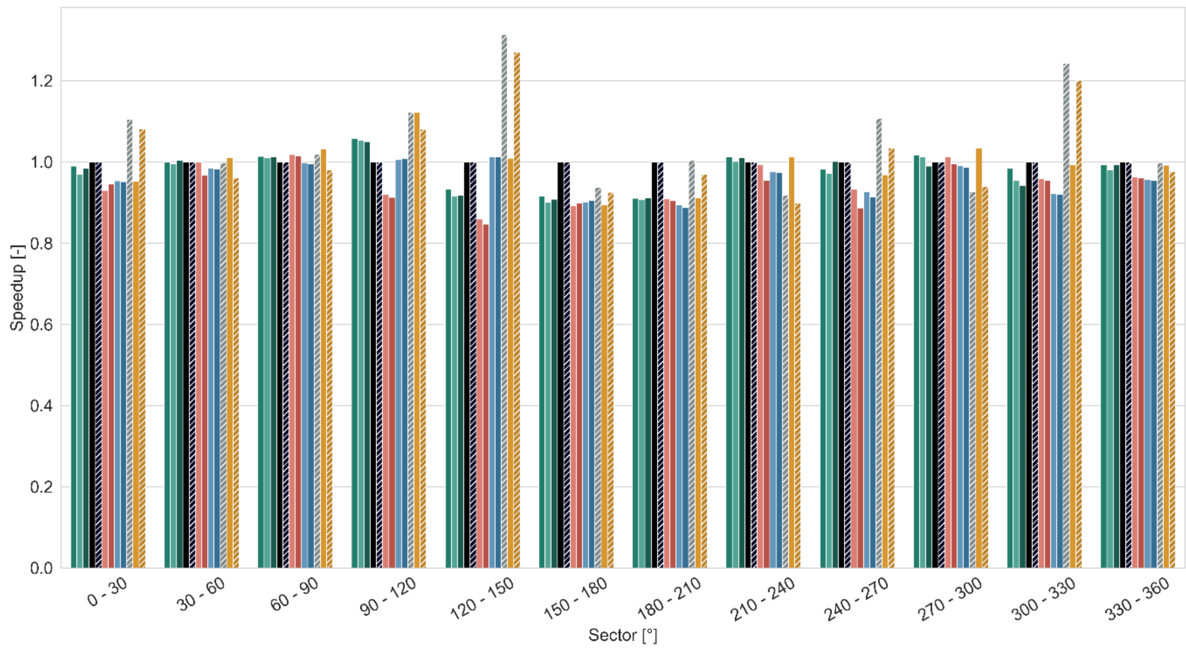


Figure 13. RMSE values between simulated and measured wind profiles for different heights for met mast 20. The dashed lines represent the mean RMSE values for all simulations at the specific height profile.

No speed-up is thus defined by a value of 1.0 and no turning is defined by a value of 0 degrees, as both seen in Figure 14 for the measurement data. A deviation from these values for met mast 29 therefore shows the quality of the calibration of the simulation data. As can be seen, in terms of speed-up most of the models are well calibrated, however, both WAsP simulations show some outliers. Considering the wind rose for met mast 29, the various simulation tools, especially in case of WAsP, least accurately predicted wind directions with the lowest frequency of occurrence.



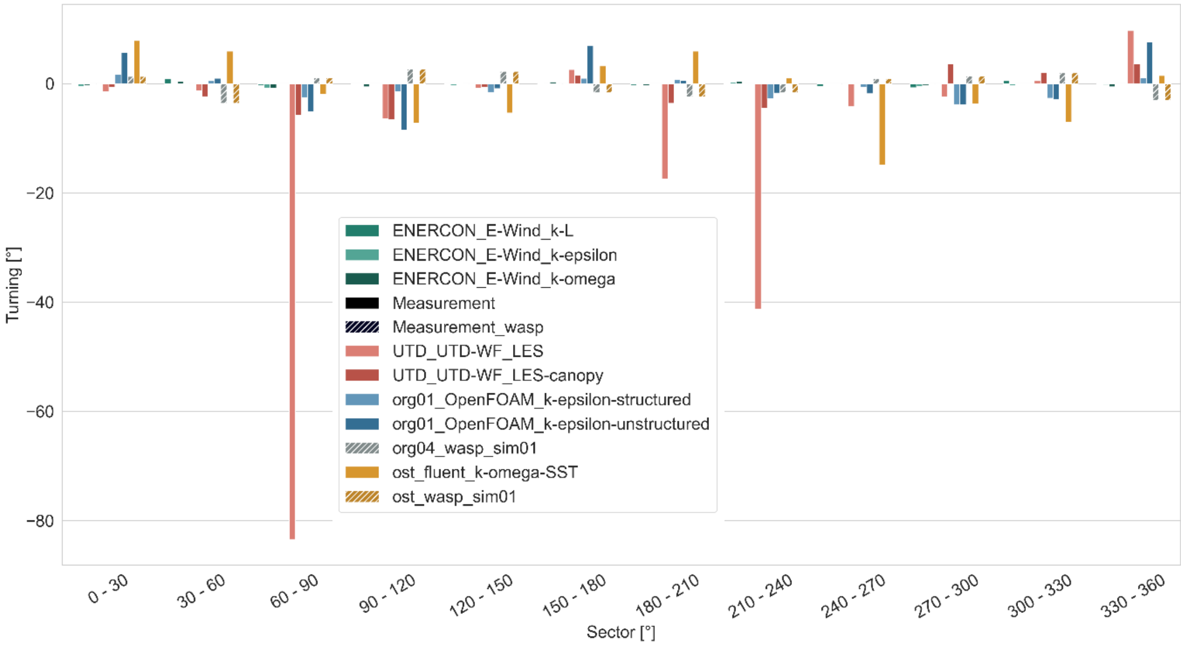


Figure 14. Speed-up factors (top) and flow turning values (bottom) for the twelve sectors of met mast 29.

In terms of flow turning, E-Wind shows accurate results, whereas the LES simulations have very large outliers, the reason of which remains to be investigated. One possible explanation for this might be that only eight instead of at least twelve wind directions could be simulated because of time and cost constraints and thus the splitting into twelve sectors caused those differences. The Fluent and the OpenFOAM (unstructured) simulations also show flow turning up to ten degrees for some sectors.

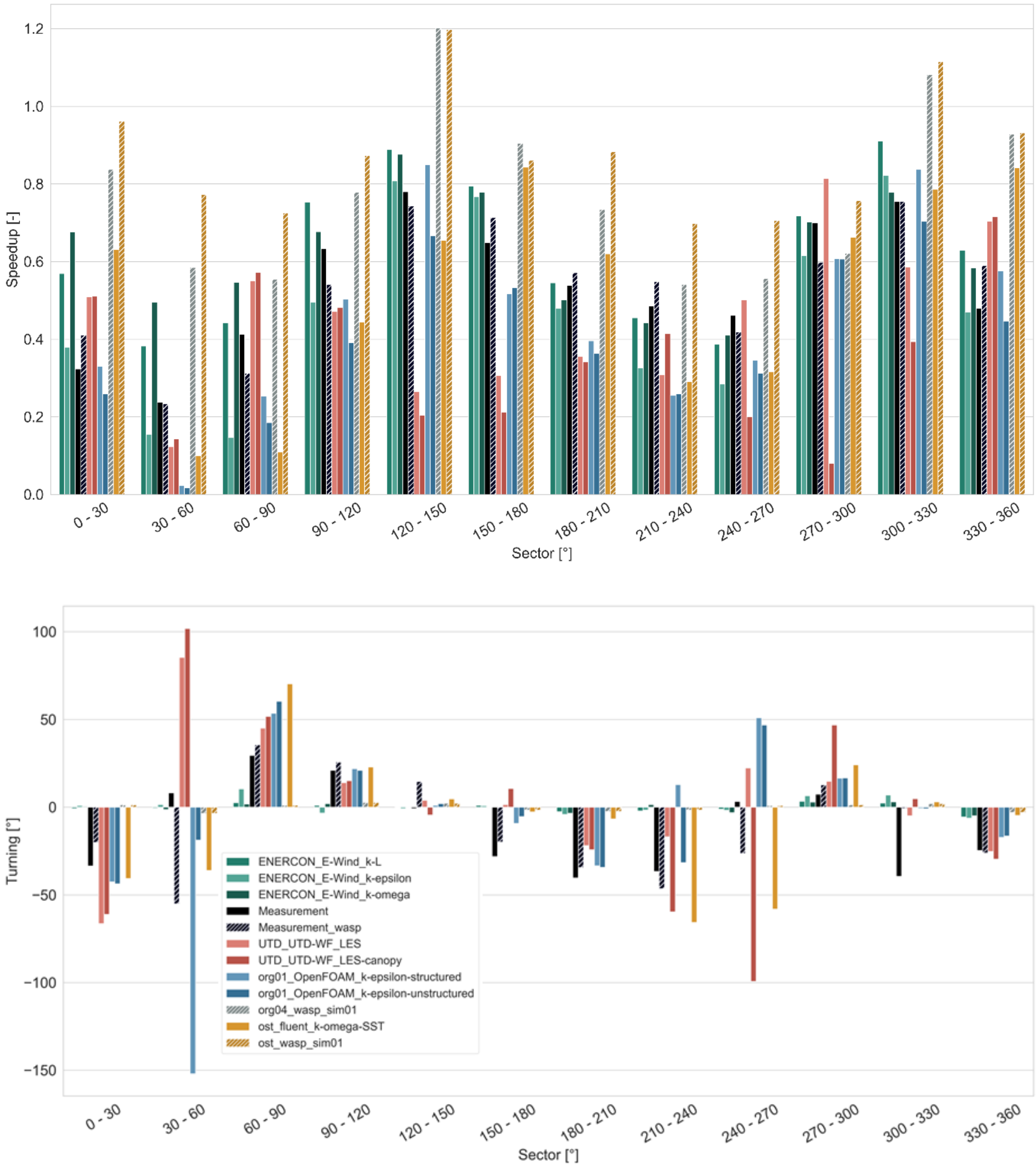


Figure 15. Speed-up factors (top) and flow turning values (bottom) for the twelve sectors of met mast 25.

For met mast 25 the ideal speed-up would be the same as the one observed for the measurement, as seen in Figure 15. Here the WAsP simulations over-predict the wind speed, in some cases significantly, for all sectors. In general, large deviations between the measured data and the simulation results can be observed, even though the overall trend for each sectors seems to match.

As to the flow turning, again the ideal values would match the ones of the measurements. Significant differences for all simulations are apparent. Met mast 25 seems to be in an extremely difficult location and none of the models is able to reliably give accurate results.

For met mast 20 (Figure 16), the farthest location, the prediction accuracy of all simulation tools seems to be much better as compared to met mast 25. The speed-up factors match more closely the ones of the measurements and the general trend fits as well. Here, E-Wind is the most accurate and consistent. Moreover, the measurement data shows little flow turning, in the range between -10° and +15° and all simulations lie within that range.

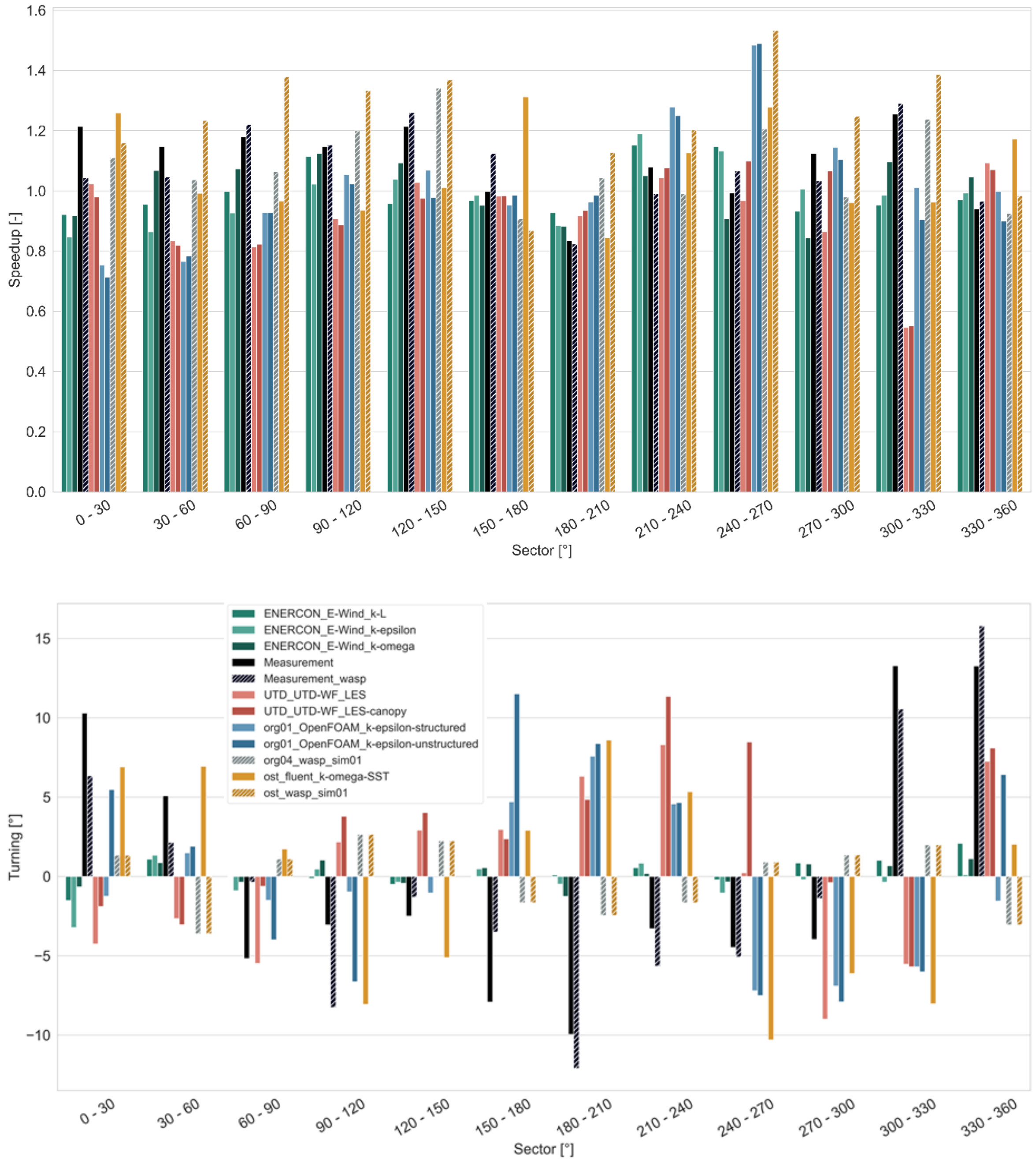


Figure 16. Speed-up factors (top) and flow turning values (bottom) for the twelve sectors of met mast 20.

AEP scores

Next, we look at the values for the Annual Energy Production (AEP). Each participant submitted these values for each of the twelve sectors for met masts 29 and 25 in the template mentioned above. The summation of all sectors yields the overall AEP. Figure 17 shows the normalised AEP values per sector for met mast 29. The normalisation was done based on the overall production value and clearly highlights the relative deviation from the AEP based on the measurement data. The measurement based AEP values, denoted by "Measurement" and "Measurement_wasp", were calculated by multiplying the wind speed time series data with the power curve of the wind turbine. The wind turbine is located close to met mast 20 and its power curve was used for the AEP calculation at all met mast locations for comparison purposes. For "Measurement_wasp" a different sector division was used in order to match the one of the WAsP models, which defined the northern sector from -15° to +15° instead of 0° to 30°. The AEP prediction accuracy mainly depends on two measures, the wind speed and the wind speed frequency per sector. As can be observed, all models somewhat under-predict the energy production for most of the sectors. All E-Wind models are very consistent and the LES simulations are most accurate for the main wind direction, 30° to 60°.

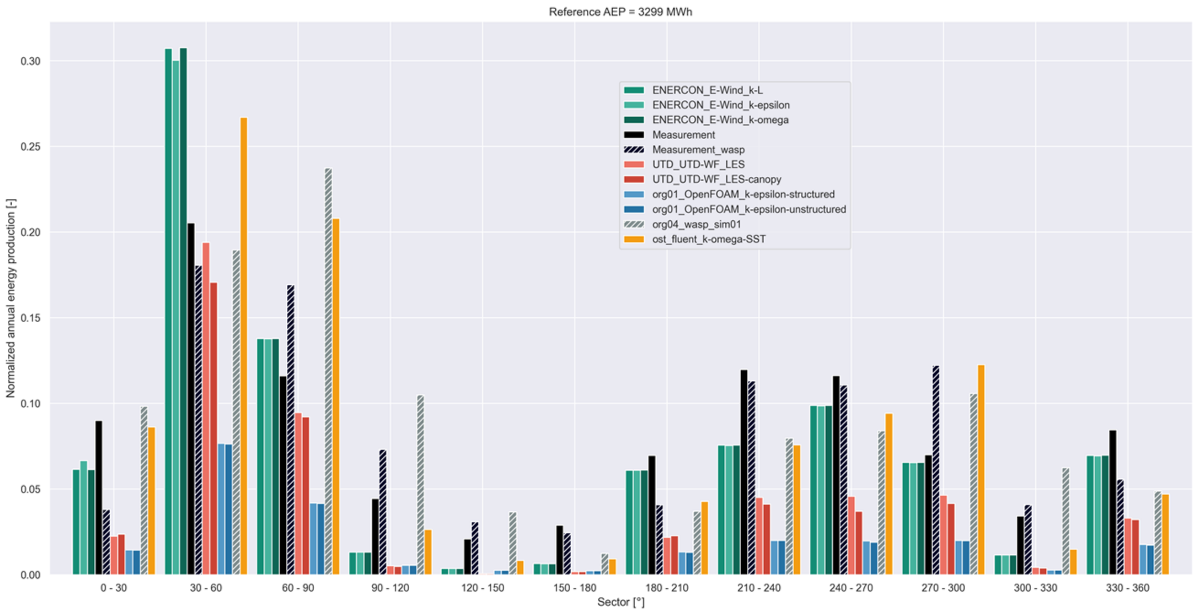


Figure 17. Normalised AEP values per sector for met mast 29

For met mast 25, see Figure 18, the South-Eastern direction is largely under-predicted by all models but WAsP, which shows large over-estimations for all sectors. This is consistent with the over-estimations of the speed-up factors, see Figure 15. The North-Western direction with higher wind speeds is only captured by E-Wind with the k-L turbulence model.

Figure 19 shows the normalised overall AEP values for met masts 29 and 25. Notable are the performance of the Fluent model for met mast 29, accurately predicting the AEP, and the prediction performances of the three E-Wind simulations. The WAsP simulation also shows good agreement with the results based on the measurements. The poor performances of the OpenFOAM and LES simulations remains to be discussed with the participants. For met mast 25 the E-Wind simulations with the k-L and the k-omega turbulence models show good agreement with the measurements. As already observed in the AEP values per sector, the large over-estimation of the WAsP simulation very pronounced for the overall values.

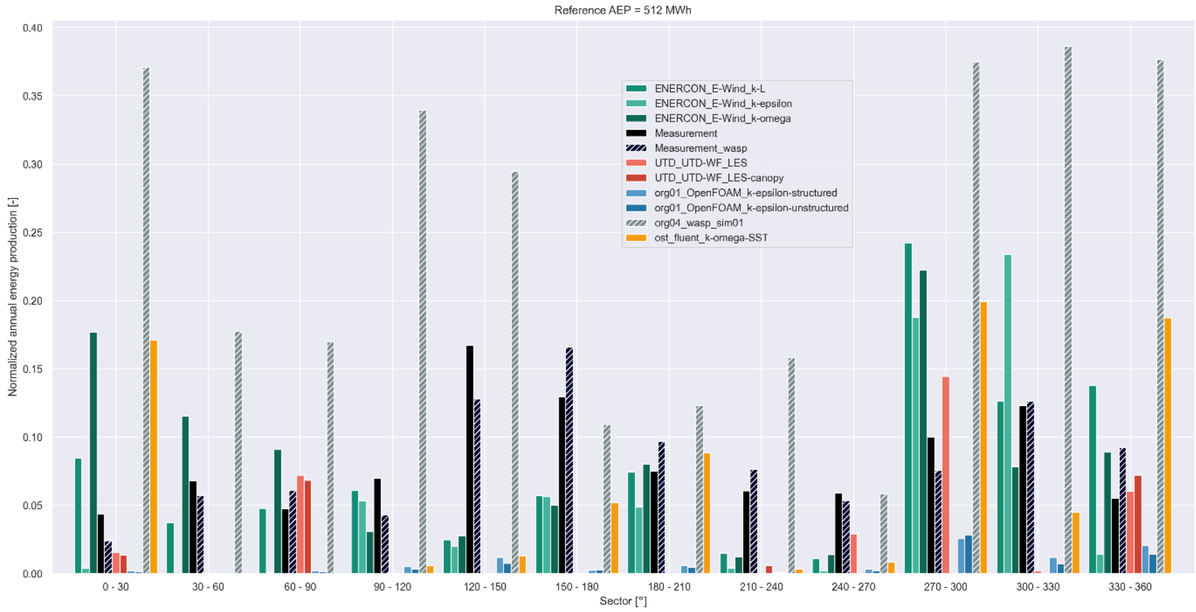


Figure 18. Normalised AEP values per sector for met mast 25

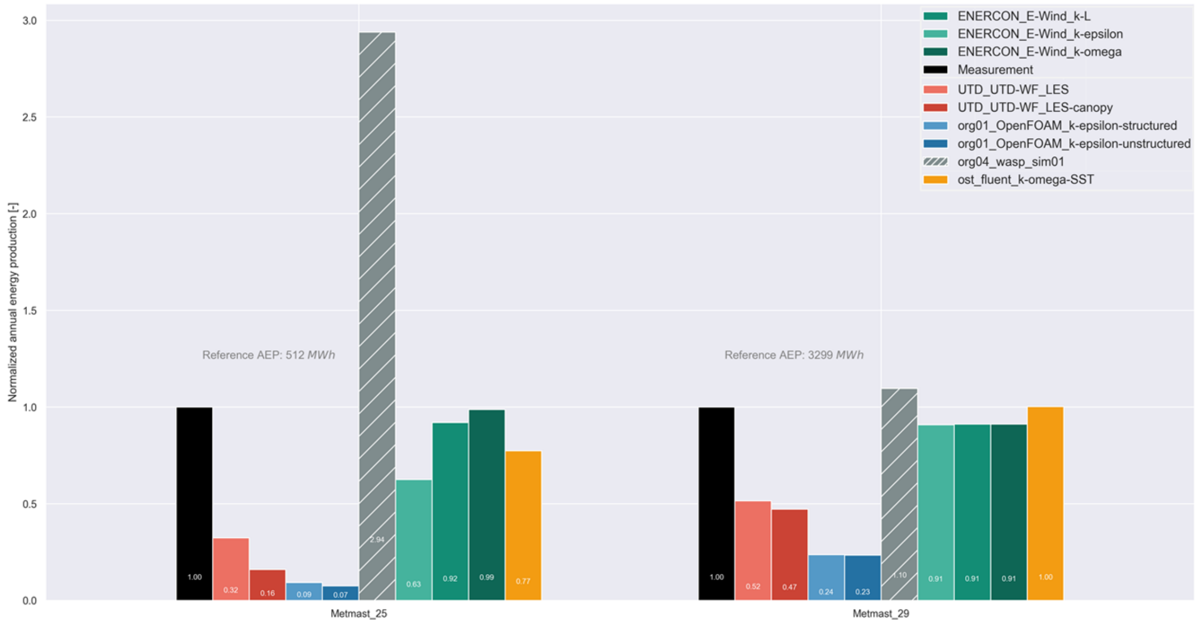


Figure 19. Normalised overall annual energy production values for met mast 29 and met mast 25

The overall RMSE values between the measured and simulated wind profiles as well as the overall AEP values are used to determine the after skill scores, which are discussed in more detail in the next section.

3.2.6 Skill versus costs

In the previous section the overall RMSE values between the measured and simulated wind profiles and the overall AEP values were considered in more detail. In this section these values are reduced to a single number, namely the after skill score. A distinction is made between the *after skill score* based on the wind profiles and the *after skill score* based on the AEP values, both of which are considered in this section.

For Stage 1 the *after skill score* based on the wind profiles was defined by

$$score(ws) = \begin{cases} \dfrac{(3 - RMSE_{ws})}{3}, & ws \leq 3 \\ 0, & ws > 3 \end{cases} \tag{1}$$

and the *after skill score* based on the AEP was defined by

$$score(aep_{sim}, aep_{meas}) = 1 - \left| \frac{aep_{meas} - aep_{sim}}{aep_{meas}} \right| \tag{2}$$

The deviation error value of 3 was used, because it corresponds to the maximum observed wind speed error between the simulations and the measurements. However, the exact value chosen here is not important because we are only making comparisons.

Figure 20 shows the relative skill versus relative cost scores based on wind speed for different evaluation heights at met mast 29. Here relative means that all *after skill scores* were normalised by the highest achieved score at the calibration met mast (met mast 29). The light and darker red regions indicate possible ranges of insufficient or unacceptable scores, respectively, that could be defined by the user. Here, for example, relative skill scores below 40% would not be worth considering no matter how low the cost. In turn relative cost scores above 60% would be deemed too costly independent on the simulation accuracy. The dots denote the *before skill score* metrics, as discussed in Section 3.2.4, whereas the cross-marks denote the *after skill scores*. The before metrics predict that the E-Wind tool should be the most effective for this location, with most effective being defined as the region with relatively high skills and low costs. The after metrics, however, show that E-Wind k-epsilon, OpenFOAM k-epsilon and WAsP are most effective here. The highest scores are achieved at evaluation height 80m, whereas for the profile evaluation from 10 to 100 meters the before approximate the after skill scores better. Based on a qualitative analysis, the wind speed prediction quality at met mast 29 reaches 9 out of 10 points.

At met mast 25, see Figure 21, the evaluation height plays less of a role and the before and after scores are closer to each other in general, potentially indicating that at this location the site and complexity is more in line with the evaluation based on the answers given in the questionnaire. As this location is more complex to simulate accurately, the wind speed prediction quality was given a score of 8 out of 10.

For both met masts the immense cost needed for LES simulations is revealed. Given the lower accuracy compared to simpler models, the additional costs are not justified and so cannot be recommended for wind modellers in the industry. Hence, the application of tools such as LES and DNS will remain solely in the academic field in the foreseeable future. As mentioned before, the reason for the lower accuracy might be due to the restriction of wind directions that can be simulated, limited by time and costs.

Figure 22 and Figure 23 show the relative skill versus cost scores based on the annual energy production for met masts 29 and 25. As observed in the previous section, the E-Wind and Fluent perform best for both met masts. Particularly bad performance can be seen for the LES and OpenFOAM simulations and, as mentioned above, remains to be discussed in further detail with the respective participants. The AEP prediction quality for met mast 29 was assessed to be 8 out of 10, whereas the quality for met mast 25 achieves a score of 6 out of 10. This shows the difficulty of accurately modelling flow at more complex locations.

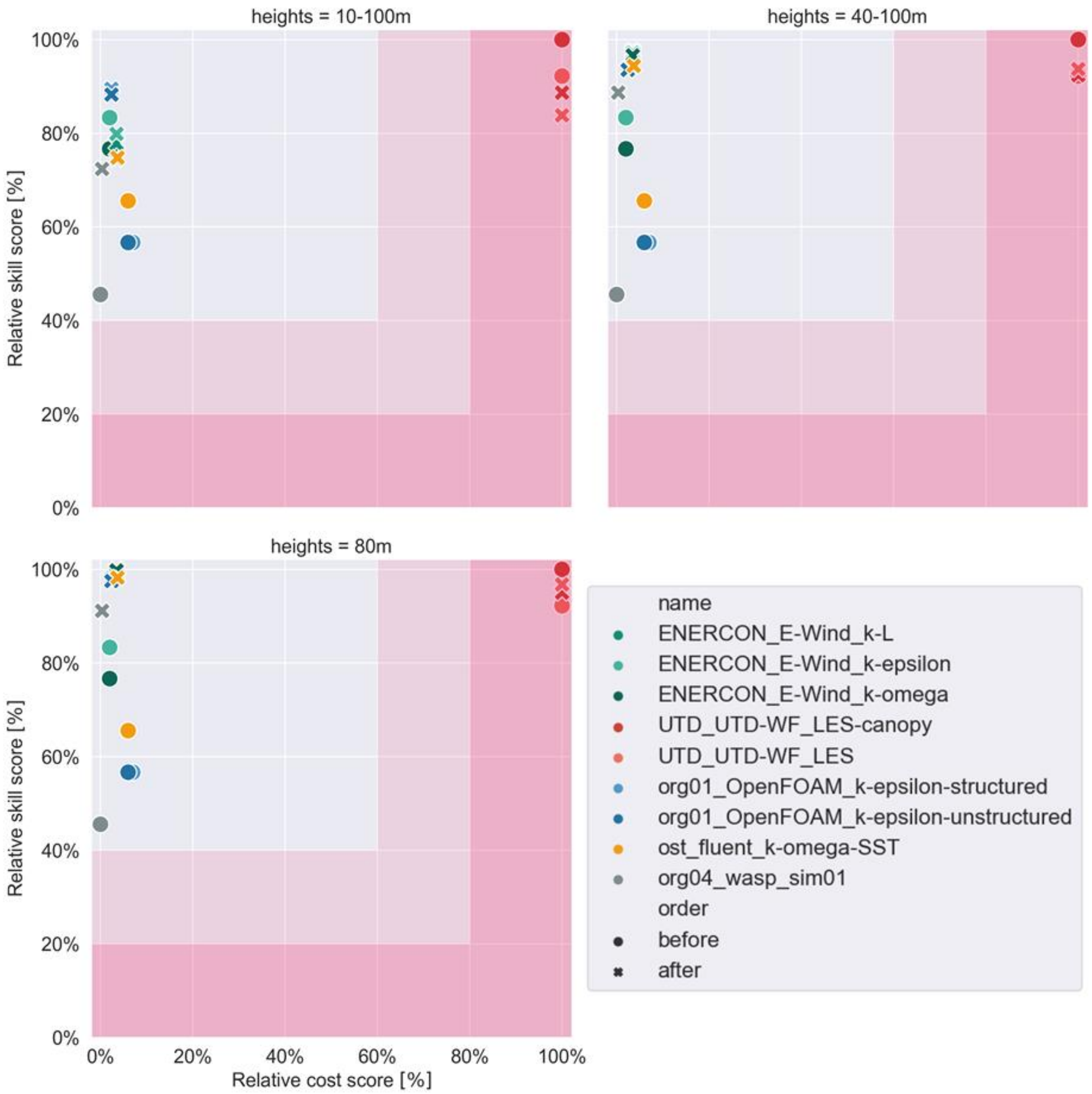


Figure 20. Skill versus cost scores based on wind speed for met mast 29

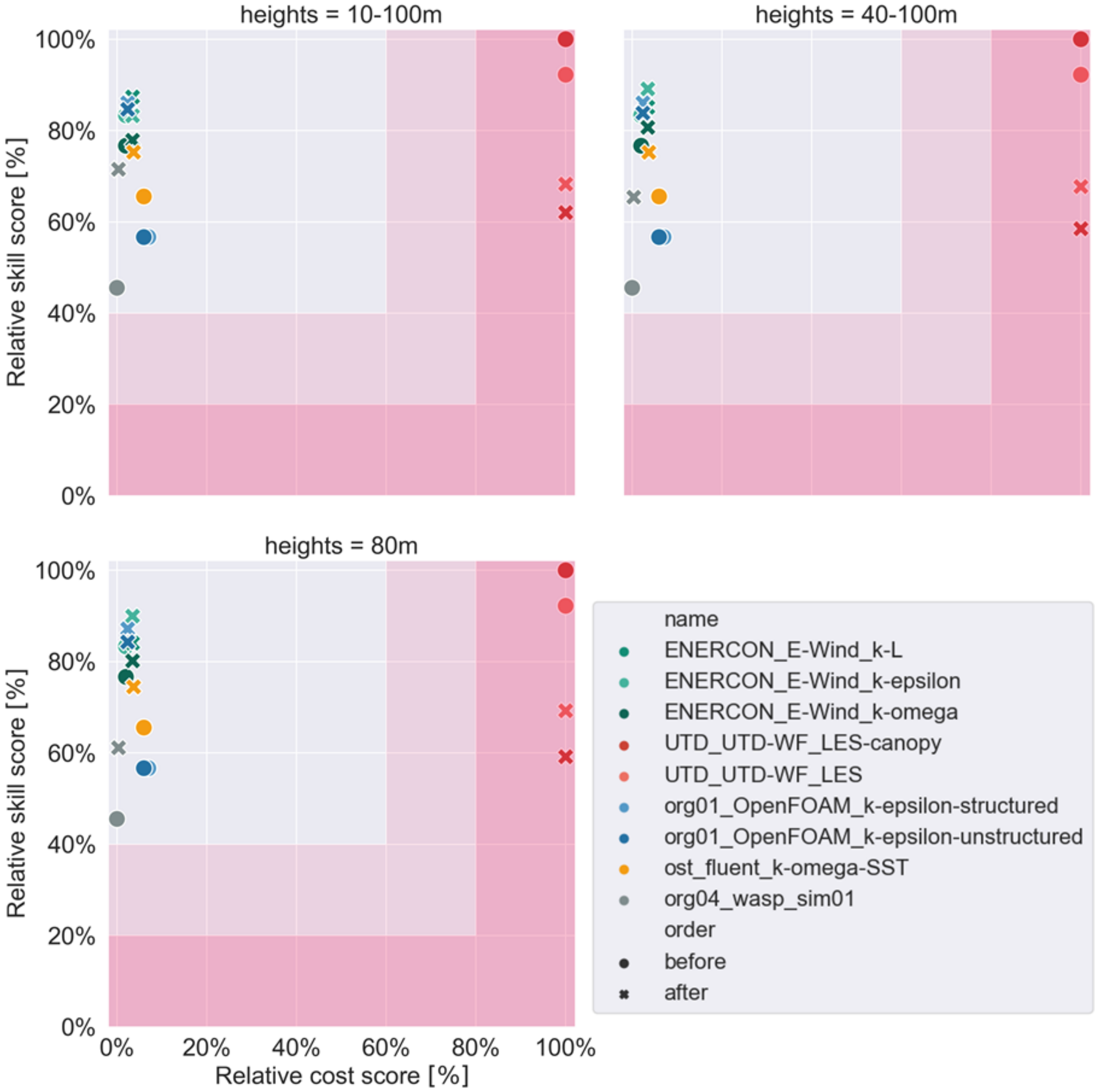


Figure 21. Skill versus cost scores based on wind speed for met mast 25

Figure 22. Skill versus cost scores based on annual energy production for met mast 29

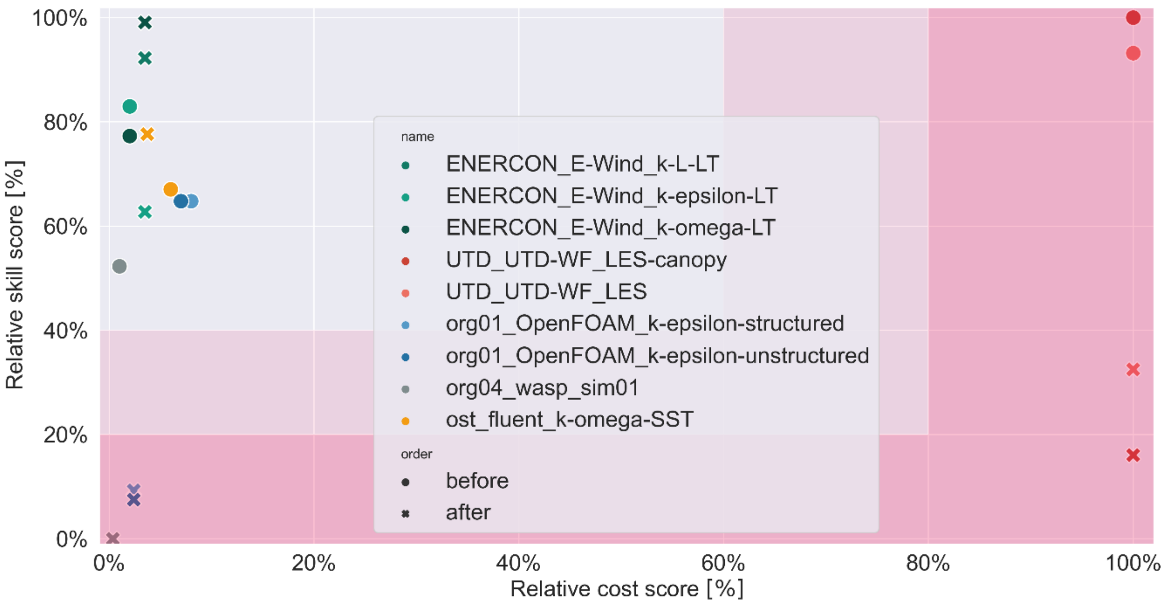


Figure 23. Skill versus cost scores based on annual energy production for met mast 25

3.3 Summary of Stage 1

The content of this chapter can be summarised by the following points.

General summary:

1. Received many different simulation results:
    1. Large variation in terms of used software and models.
    2. Makes for a great comparison.
2. Predefined and standardised result templates worked very well.
    1. Can be used as basis for the new decision tool in Stage 2.
3. More understanding and evaluation of Comparison Metricss needed:
    1. More data to better draw generalised conclusions.
4. Learnings for Stage 2 challenge:
    1. Prepare and transform the data to avoid confusion in pre-processing.
    2. Three months of data not necessarily ideal for WRA focus.
    3. Use a sharing platform to improve collaboration.
    4. Organise webinars / workshops to accompany the challenge.

Improvements to CM process needed:

5. Defining the Model Description in a Google Form and then estimating skill and cost score parameters in a separate Google Form not optimal:
    1. User enters set-up details and then estimates skill parameters → time-consuming and confusing, large room for interpretation.
    2. Final decision tool for Stage 2: collect parameters related to set-up, automatically convert to skill score.
6. The values entered into the Google Forms were transferred to an Excel template by hand:
    1. Template then had to be adjusted for each project.
    2. Will be automated in Stage 2

Skill score before estimation:

7. Many of the parameters can only be estimated once the simulations have been fully set up.
    1. The whole point of this method is to be able make an estimation of the most optimal model without carrying out any simulations.
    2. Skill score parameters need to be adjusted → user can estimate their values.
    3. Assign each parameter a confidence score → apply tool even if very little information about the set-up is known.
8. Improve weightings:
    1. Some should be dependent upon the terrain complexity classification.
    2. Some of the parameter descriptions were confusing or unclear.

Cost scores before and after:

9. Difficult to define the Actual Total Costs related to carrying out a WRA project:
    1. Staff hourly cost and number of projects per year had to be normalised for a fair comparison.
    2. Further consideration for the final version of the tool in Stage 2.

Complex terrain classification:

10. Difficult to classify the complexity of the terrain by asking questions that the participants could easily answer.
11. Many questions difficult to understand and interpret.
12. Definitions need improvement.

4 Challenge Stage 2

In this chapter, the design of Stage 2 is first discussed, followed by the results and then a summary of the findings.

4.1 Challenge design

The goal of this Stage 2 was to collect simulation results and site complexity descriptions in order to develop a new decision tool for the optimal choice of model for a given project. The challenge was published on the WeDoWind Platform[3], giving detailed information and instructions for the participants. Participants were asked to submit results of already existing wind simulations and/or AEP estimations for any site**.** As we wanted to assess the actual skill of the simulations, the site needed to have validation measurements for determining the after skill scores, as described above.

The results were used to develop the decision tool, which is introduced in more detail in the next section, enabling the estimation of skill and cost scores of the submitted results using pre-defined weighted parameters, as well as classifying the terrain complexity of the submitted sites.

The challenge was split into the following five parts. Participants had the opportunity to submit results for multiple sites and multiple simulation set-ups:

1. For each site, details of the terrain and the site had to be uploaded.
2. For each set-up for each site, details of the used model had to be uploaded.
3. For each set-up for each site, simulation and measurement results had to be uploaded.
4. Contributing to a discussion on the WeDoWind Platform about the classification of complex terrain could be joined in order to improve understanding.
5. Running tests with the decision tool as soon as a sufficient amount of results come in.

The final decision tool results are dependent on the amount of participation. However, we hope for:

1. For all sites:
    - Complex terrain classification results.
    - Summary of submitted results.
    - Summary of predicted skill and cost scores compared to actual skill and costs.
2. For sites for which three or more results are submitted:
    - Direct comparison of wind speed and AEP accuracies of different models / workflows.
    - Direct comparison of costs of different models / workflows.
    - Skill score vs. cost score scatter plots (using wind speed accuracy).
    - Skill score vs. cost score scatter plots (using AEP accuracy).
3. Description and code for the new decision tool on GitLab.

[3] https://www.wedowind.ch/wedowind-ecosystem

In order to be able to process, analyse and transform the submitted data into the *skill* and *cost score metrics* for developing the new decision tool, the participants were asked to provide one or more of the following simulation and measurement results, depending on their availability:

Wind speed:

1. Simulated and measured wind speed profiles at a validation location (including the validation location and measured heights) AND/OR
2. Root Mean Square Error (RMSE) between simulated and measured wind speed profiles at a validation location for all wind direction sectors (including the validation location, measured heights and wind direction sectors) AND/OR
3. Average RMSE between measurement and simulation at a validation point over all sectors (including the location and measured heights) AND/OR
4. Absolute difference between simulated and measured wind speed profiles at a validation location for all sectors at hub height (including the validation location, hub height and sectors) AND/OR
5. Average absolute difference between measurement and simulation at a validation point over all sectors (including validation location and hub height) AND/OR

Annual Energy Production (AEP):

1. Simulated and measured AEP at a validation location (including validation location, wind turbine hub height and type, power curve used and information of how losses were corrected for) AND/OR
2. Absolute or percentage difference between simulated and measured AEP at a validation location (including validation location, wind turbine hub height and type, power curve used and information of how losses were corrected for) OR
3. Simulated and theoretical AEP at a validation location (including validation location, wind turbine hub height and type, power curve used and information of how the theoretical AEP was calculated). *Note: the theoretical AEP can be calculated from the wind speed measurement at that location and the power curve)*
4. Absolute or percentage difference between simulated and theoretical AEP at a validation location (including validation location, wind turbine hub height and type, power curve used and information of how the theoretical AEP was calculated).

In order to evaluate the results of this challenge, a new decision tool was developed, as described below.

4.1.1 Decision Tool design

The software application is named *Decision Tool.* The tool estimates the costs and accuracies, defined as *cost* and *skill scores*, of wind resource assessment tools based on user/costumer inputs and site complexities. It helps customers choose the right simulation tool for a given site based on their accuracy and cost expectations. The *Decision Tool* will provide a recommendation for the most suitable tool for a given site and therefore helps the modeller to rely less on gut feeling and instead make an informed decision.

In order to better understand why, what and how to develop the decision tool, some user stories were gathered, allowing for a high level view and to give some context.

User stories

1. As a wind resource modeller I want use the information and experience I have about a site so that I can choose the best modelling tool.
2. As a wind resource modeller I want the best tool so that I get the most accurate results within a defined cost range for a given site.
3. As a researcher I want to have a recommendation for the best tool to use for a given site so that I can focus more on the research aspects rather than spending time and costs on the process of finding the right tool myself.
4. As a company we want to get a license for a wind resource assessment tool that works for most of our sites. The maximum cost of X should not be exceeded.
5. As a researcher I want to compare my modelling tool with the most suitable tool for a given site in order to improve my methodology.

Main purpose and functionality

The tool is mainly intended for wind resource assessment tasks for under-explored sites of various types of complexity. The users have experience in the field of wind energy with focus on resource assessment. They have or should have access to the following data in order to take the most out of this tool:

1. Terrain complexity: Ideal would be a roughness map to determine an objective value. Additionally, a trained eye for assessing the terrain in terms of slopes and other obstacles will greatly aid this process.
2. Weather complexity: Ideally a good understanding of the underlying synoptic, meso- and microscales of the chosen site should be known in order to sufficiently assess the weather complexity. At least met mast data should be available to get an overview of the flow structures on the site.
3. How accurate do the wind speed simulations need to be? (Skill score)
4. How much should be the ideal and maximum cost of the whole process?

The fewer available information and experience the lower the chance that the tool will be able to make a reliable decision in terms of optimal wind resource assessment tool for the given site.

In order to build and validate the *Decision Tool* it is necessary to establish a database with data and results of already finished wind resource assessment projects. Four important parameters are needed and have to be determined in order to build the database. In the following, these parameters will be introduced and the process to obtain them explained in some detail. 1) The *after skill score* describes the actual accuracy of the employed wind resource assessment tool. The value is calculated based on the comparison of measurement data and simulation results at specific locations. It is differentiated between an *after skill score* based on wind speed predictions and annual energy production predictions. 2) The *after cost score* is the overall cost of the wind resource assessment project for the chosen site. The parameter is deemed more complex compared to the *after skill score* in that it involves less transparent influencing factors such as computing resources, hourly rates, hours spent working on the project and licensing costs. 3) The *before cost score* denotes the estimated costs before conducting the wind resource assessment. It includes factors similar to the *after skill score* with the difference of having only estimates instead of concrete values. 4) The *before skill score* is by far the most complex parameter to determine. It describes the estimated accuracy of a tool for a site with a given complexity before running any simulations. The value is solely estimated based on a questionnaire filled out by the modeller beforehand, including questions such as

- How do you judge the general complexity of this site? (1-10)
- What kind of simulation did you run?

- Which turbulence model was used?
- How did you model surface roughness?

Additionally, roughness and terrain maps or descriptions as well as meteorological data can be provided, if available, in order to obtain a more objective estimation compared to mere human judgement.

Based on these results a transfer function between the *before* and *after scores* will be developed and constitutes the *core* of the decision tool business logic. The transfer function can then be developed based on techniques such as multiple regression, k-nearest neighbours, decision tree based models and more.

Besides having a *business logic layer* with the transfer function as its central part and the *data access layer* with the mentioned database, a *presentation layer* is created serving as the user interface. The user interface is a web interface in form of a GUI. Users/modellers can simply access it through a modern web browser and interact with the *Decision Tool*.

The interaction between the three layers will be as follows.

1. The wind resource modeller fills out a survey through the web interface. (*Presentation layer*)
2. The data will be stored in the database. (*Data layer*)
3. The transfer function uses the submitted user data to determine the most appropriate modelling tool. (*Business logic layer*)

User interface

As mentioned above, the user interface will be a web interface in form of a GUI, where access is allowed via a modern web browser. The interface is in the presentation layer and is the only part of the tool that is exposed to the user. An input mask can be created where the user inserts all necessary data, and the tool then calculates the scores automatically and presents the results in the web application.

The user inputs consist of:

- **Site details:** The user enters details of the site complexity, such as the number and steepness of the slopes, similar to the questions asked in the Model Description related to site classification in the existing model. The questions will be improved to make them easier to quantify and less open to interpretation.
- **Model description:** For each model, the user enters details of the planned model set-up, based on the information already given in the existing Model Description. However, the fields will be directly related to the skill score parameters, and the tool will calculate the parameters automatically, without the user having to interpret (or misinterpret) the parameter description and scale. The improvements will be added as discussed above.
- **User experience with model:** For each model, the user answers questions similar to the questions in the skill score parameters (Skill E).
- **Cost estimations:** For each model, the user estimates the costs using the categories introduced in this work. Few improvements are required to this part.
- **Optimisation constraints:** The user defines what they are trying to optimise and what the constraints are. This helps the decision tool to automatically choose the optimum tool for a given application. In this work so far, this "choice" was only done by looking at the Comparison Metrics charts and assessing which tool offers the best compromise of costs and accuracy. The user may choose, for example, to always choose the model that is furthest towards the top left of the

chart (maximum skill, minimum costs), or to always choose the model that has the highest skill below a certain cost limit.

- **Confidence scores:** Every one of the fields related to the site details, model description, user experience with model and cost estimations will also require a confidence score. This will allow users to apply the decision tool without knowing many details of the planned set-up. The more details that are known, the higher the confidence score will be and the more accurate the final decision will be.

The main and ideal use of the tool are called the happy flow. In case special actions are needed or errors occur the user enters into and alternative flow. In the following the happy flow and three alternative flows, depending on the situation, are described.

Happy flow

1. User accesses web application through web browser by entering address.
2. The user sees the landing page with two options – login and register.
3. User logs in with user credentials.
4. The user can start a new tool decision process by filling out a survey.
5. After submitting the survey, the most suitable tool for the given site is presented, as explained above.
6. The results are encrypted and saved in a database and can later be accessed by the same user again.
7. The user may repeat the survey as many times as needed.

Alternative flow 1

3. User registers by entering email and password.
4. After registration, the user is directed to the login page and can log in.

Alternative flow 2

4. The user can visit the account to inspect user information and saved results from previous sessions.

Alternative flow 3

3. In case the user does not remember the login in credentials anymore, a password reset request can be send.
4. The user receives an email with a password reset link and has now 30 minutes change the password.
5. Clicking the link in the mail will open a page where the user can enter the new password.

GUI

The web application is represented by a GUI, which is presented to the user. In this section screenshots of the application are shown in order to give a better picture of the presentation layer of the tool.

The first view that is presented to the user is a login screen, as shown Figure 24 on the left. A warning is raised in case incorrect login credentials were used (right).

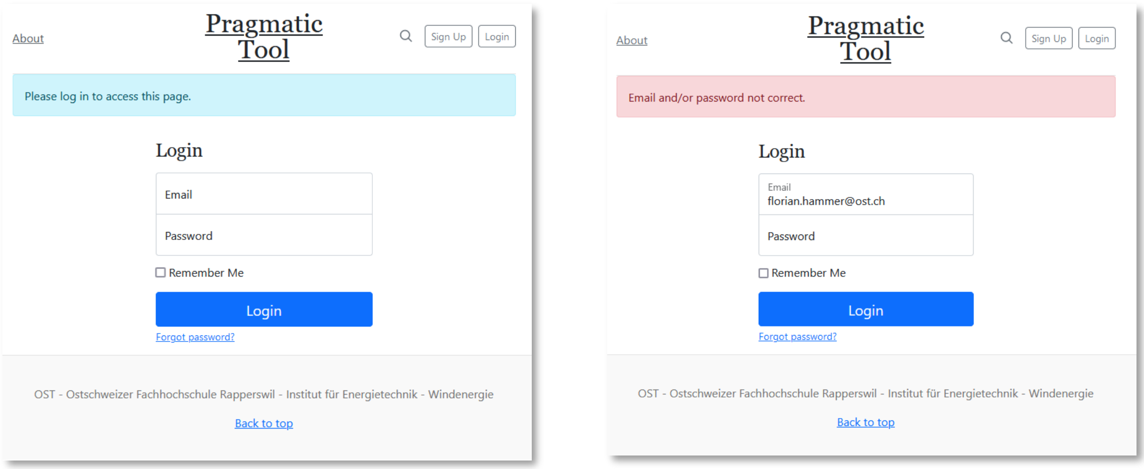


Figure 24. (Left) Login screen and (right) failed login attempt

Figure 25 shows the registration page in case the user does not have an account yet.

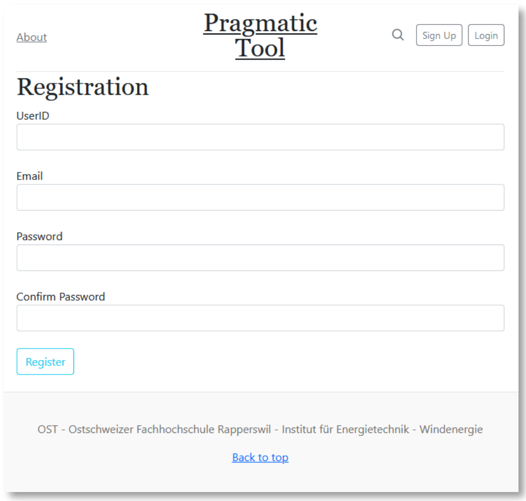
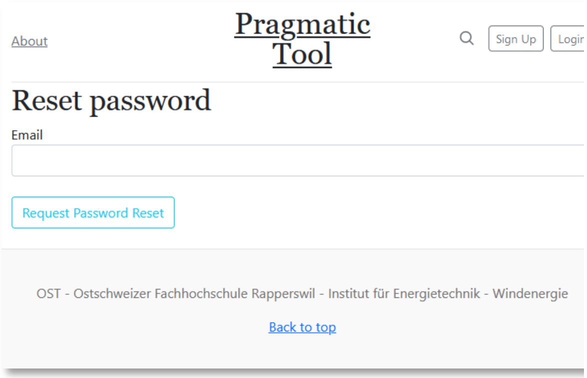
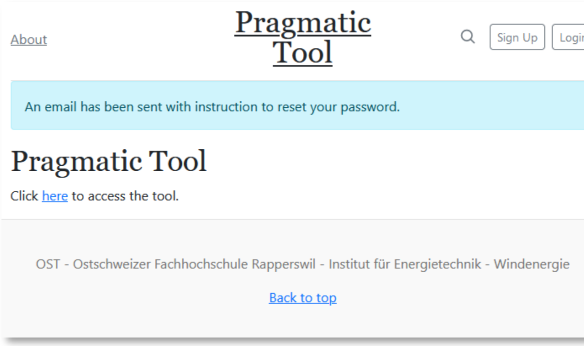


Figure 25. Registration page

It is also possible to request a password reset. The process is shown in Figure 26.





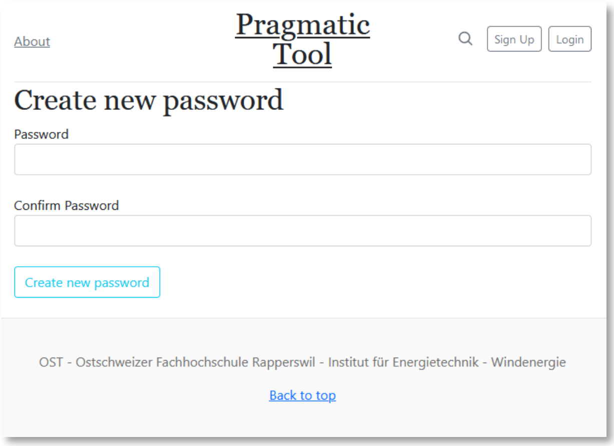


Figure 26. (Top left) Reset password page (top right) Confirmation of password reset (bottom) New password creation page

In order for the tool to determine the most suitable tool for a given site the user and site information mentioned above are necessary. This information can be entered in a survey, see Figure 27, which is encrypted and saved in a database.

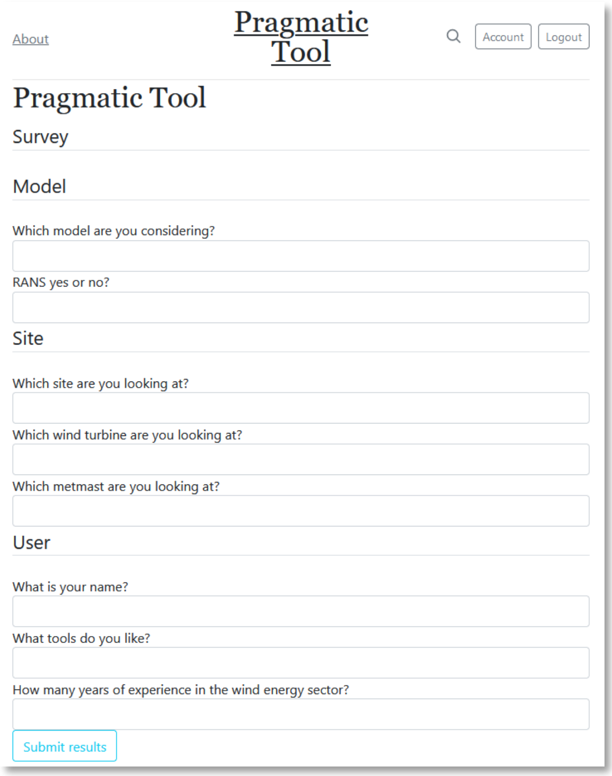


Figure 27. Survey to be submitted

Technical specifications

The Decision Tool web application is based on Flask (version 2.0.2), a Python micro web framework. The front-end is developed with Bootstrap v5.0, which is a free and open source CSS framework. The business logic is written with Python 3.8 and the relational database is based on the SQLite management system.

The concept of the tool is depicted in Figure 28 and consists of:

- **Complexity classifier:** The tool will automatically calculate a complexity score based on the site details input by the user. This will be a number from 0-100%. The user can optionally upload the digital map of the site.
- **Skill score calculator:** For each model, the model description and user experiences inputs made by the user will be converted to a skill score by converting the user inputs to scores, weighting them according to the in-built weightings, and averaging them to get one score between 0% and 100%. Some of the weightings will be altered for the complexity class. The weightings can be adjusted by the decision tool operator but not by the user.
- **Skill score confidence class calculator:** For each model, the confidence scores that have been input by the user will be combined to give an overall confidence score.
- **Cost score calculator:** For each model, the total costs will be calculated using the inputs made by the user.

- **Cost score confidence class calculator:** For each model, the confidence scores that have been input by the user will be combined to give an overall confidence score.

The results consist of:

- **Complexity class:** The complexity score will be converted to a number between one and four. This number will be used to scale certain weightings that are dependent upon the complexity class. The details of this step have to be developed in more detail.
- **Skill score:** For each model, the skill score will be presented to the user.
- **Skill score confidence class:** For each model, the skill score confidence score will be converted to a class (a number between one and four) and presented to the user.
- **Cost score:** For each model, the cost score will be presented to the user.
- **Cost score confidence class:** For each model, the cost score confidence score will be converted to a class (a number between one and four) and presented to the user.

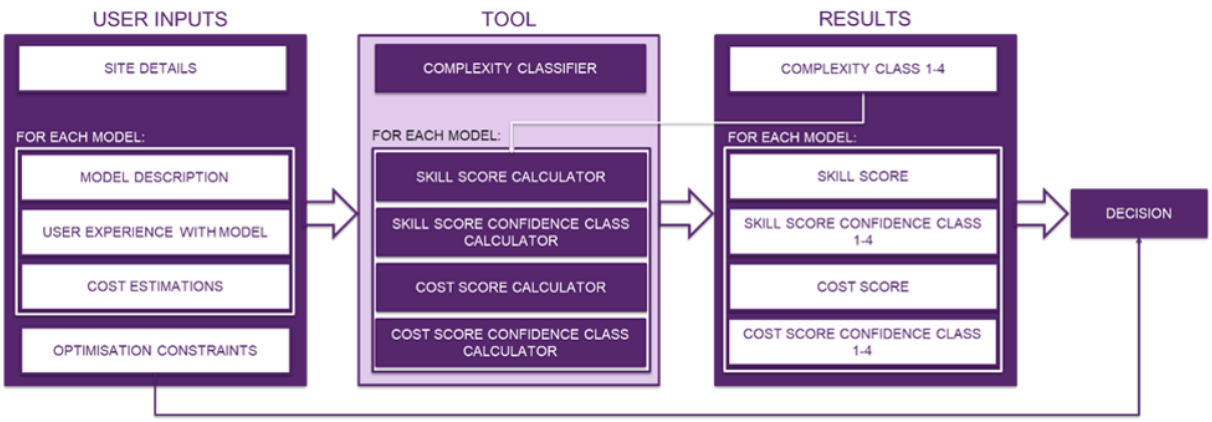


Figure 28. Concept of the decision tool

The decision is made as follows:

- For all the models considered, the resulting relative skill and cost scores are displayed on a chart of skill score vs. cost score (relative to the maximum score amongst the considered models). This chart is similar to the Comparison Metrics charts displayed in this project, but without the "after" scores (because the simulations have not been carried out). Each point represents one model, and will have error bars associated with it that relate to the confidence scores.
- Depending upon the optimisation constraints entered by the user as described above, a decision will be made regarding the most optimal tool for the application, together with a confidence score.

A simplified UML class diagram for the decision tool is depicted in Figure 29. These are the components that comprise the data and business logic layer.

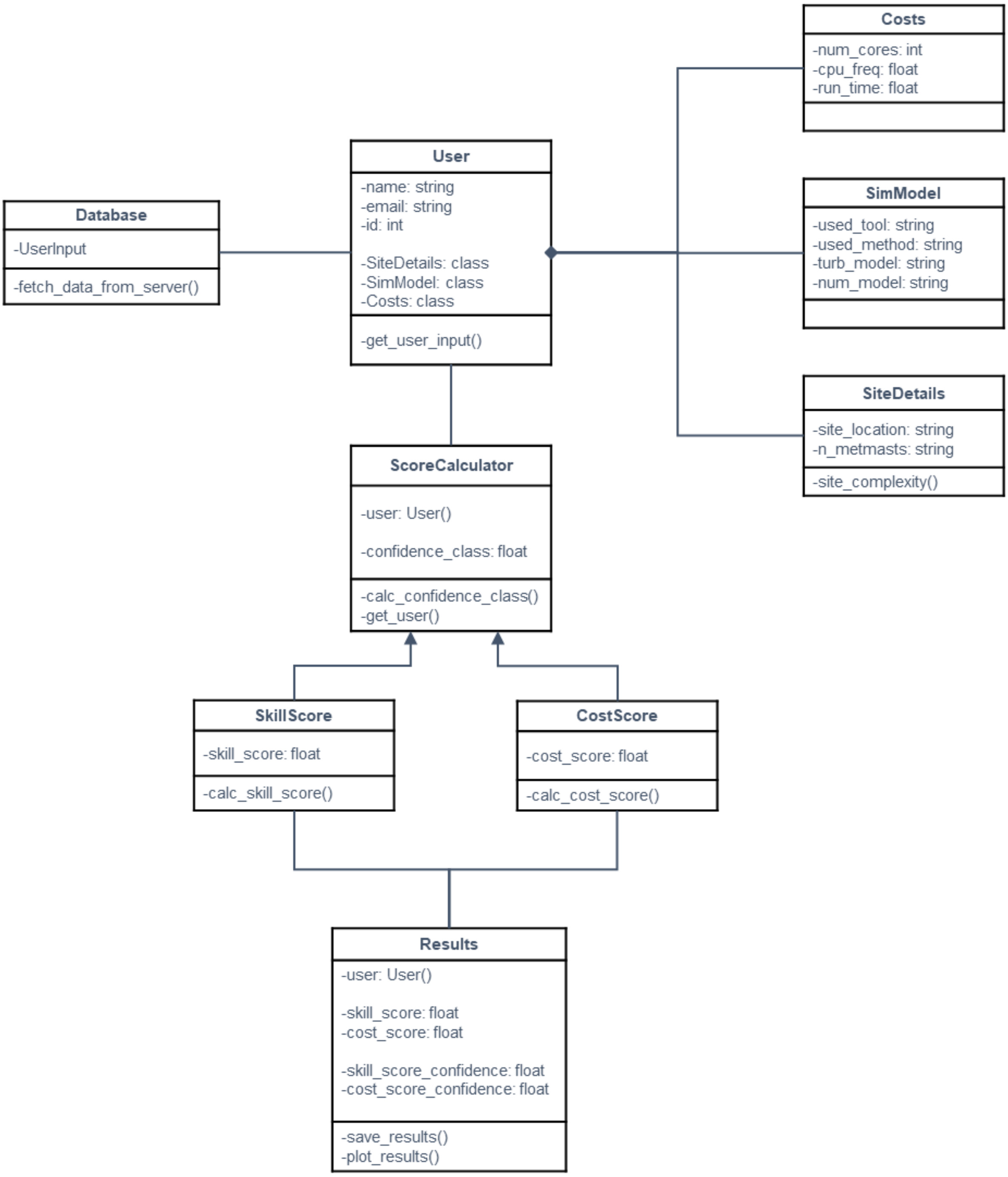


Figure 29. Simple UML class diagram for the decison tool

4.2 Results and discussion

In the following sections, the results for Stage 2 are presented. First, in Section 4.2.1 we look at the submitted data of participants. Having gathered data from various tools and participants, the site complexity, before and after score metrics are discussed in Sections 4.2.3, 4.2.4 and 4.2.5. Lastly, the skill versus cost score plots for the different submissions are presented in Section 4.2.6, followed by a short summary.

4.2.1 Data Submission

Following the launch of Stage 1, we received a total of 7 registrations from the companies using the tools shown in Table 5:

Table 5. Submissions to Challenge Stage 2

Organisation	Country	Tool	Status
BREG BV	Belgium	-	Not submitted
University of Exeter	UK	-	Not submitted
Barcelona Super Computing Center – Participant 1	Spain	-	Not submitted
Barcelona Super Computing Center – Participant 2	Spain	-	Not submitted
ZAMG	Austria	-	Not submitted
ENERCON	Germany	E-Wind	Submitted
UL Renewables	Spain	WRF	Not submitted

In the final column of this table, the submission status can be seen. Unfortunately, as for Stage 1 we did not receive as many submissions as expected. In Stage 2, we attempted to address the challenges mentioned in Section 3.2.1 by not demanding any particular project or input data and allowing any format of submissions. However, this reduced the specificness of the project and we believe it was therefore difficult for participants to engage. In the future, a compromise between the approaches used for Stage 1 and Stage 2 seems to be the most sensible idea, and this will be considered for future projects.

4.2.2 Data overview

An overview and details of the submitted data is shown in Table 6. In total, four different sites with varying site complexities from one organisation were used. As for Stage 1, Enercon used their in-house tool E-Wind. For the different sites mainly the turbulence models and grid representation were changed in order to account for the site complexity.

Table 6. Overview of submitted data

Name	Organisation	Simulation	Turbulence model	Site complexity	Grid
Site01	Enercon	RANS	k-epsilon	30	Combination
Site02	Enercon	RANS	k-epsilon	70	Combination
Site03	Enercon	RANS	k-epsilon	20	Mesh-conform
Site04	Enercon	RANS	k-L	100	Combination

4.2.3 Site complexity score

For Stage 1 it was concluded that the site complexity classification based on the given questions was rather difficult, leading to inaccurate scores. Hence, for Stage 2 we decided to determine the site complexity based on terrain, roughness and weather station data accompanying the respective site. For this, the participants were asked to upload this kind of data, if available. Additionally, the participants were asked to rate the site complexity on a scale from 1 to 10. The goal was to assess and compare the self-evaluation of the participant together with his/her experience and the objective complexity measure based on the submitted terrain, roughness and weather data. This would in turn help us to further refine questions for the site complexity classification for sites where this additional data is not available.

Unfortunately, this task proofed more difficult and time consuming than initially anticipated. Moreover, the lack of data would have rendered the results of this analysis meaningless. Hence, we decided to reschedule this task to a future project and continue with solely the self-evaluated site complexity scores. This should not cause major problems in our case, as the values of the self-evaluation for the different sites is based on a single person, using the same simulation tool. The site complexity can hence be assessed in a relative sense.

In order to make it more comparable to Stage 1, the complexity scores on the scale from 1 to 10 were scaled by a factor of 10.

4.2.4 Before score metrics

Similar to Stage 1, the before skill scores were determined based on the answers given by the participants to a questionnaire. While conducting the Stage 1 challenge it was also found that some of the questions were either too subjective or too hard to interpret for the participant. Hence, only the question that seemed most objective were picked and slightly reformulated in order to make it simpler to answer. Furthermore, the before skill score is now comprised of two parts, the *user skill score* and the *model skill score*, who are determined based on the questions shown in

Table 7 and

Table 8, respectively. The weighted sum of both of these scores then results in the before skill score. This allows for another level of tuning and matching before and after skill scores.

Table 7. Questionnaire for determining the user skill score

Questions	Answers	Scores
How many years of experience do you have in the field of wind energy?	Less than 1 year	20
	1 - 2 years	60
	2 - 4 years	80
	4 - 6 years	90
	More than 6 years	100
How many years of experience do you have with complex terrain?	Less than 1 year	20
	1 - 2 years	60
	2 - 4 years	80
	4 - 6 years	90
	More than 6 years	100
How many years of experience do you have with the used simulation tools and applied methods?	Less than 1 year	20
	1 - 2 years	60
	2 - 4 years	80
	4 - 6 years	90
	More than 6 years	100
Approximately how many different sites have you simulated?	1	20
	2 - 4	60
	5 - 8	80
	9 - 14	90
	15+	100

Table 8. Questionnaire for determining the model skill score

Questions	Scores
What kind of simulation did you run? (RANS, URANS, LES, LBM-LES, etc.)	Score from 0 – 100 depending on the chosen answer. Additional weighting is applied depending on the site complexity.
How did you model surface roughness?	Score from 0 – 100 depending on the chosen answer
What grid type did you use? (e.g. structured/unstructured, mesh-conform/immersed boundary)	Score from 0 – 100 depending on the chosen answer. Additional weighting is applied depending on the site complexity.
Did you conduct a grid independency study?	Score from 0 – 100 depending on the chosen answer
How did you calibrate the wind profile to the one at the calibration / reference mast?	Score from 0 – 100 depending on the chosen answer
How many different wind directions did you use for the AEP calculation? (e.g. 12 or 24)	Score from 0 – 100 depending on the chosen answer
Which method did you use for calculating the wind speed at the rotor?	Score from 0 – 100 depending on the chosen answer
How did you extrapolate the measured data for the wind farm lifetime?	Score from 0 – 100 depending on the chosen answer
How did you correct for air density at the site?	Score from 0 – 100 depending on the chosen answer

What is more is the addition of site complexity dependency for some of the scores in

Table 8. It can be safely assumed that the kind of simulation (RANS, URANS, LES, LBM-LES, etc.) as well as the type of grid are dependent on the complexity of the site. This dependence within the decision tool was modelled by the following function

$$f(c) = b \cdot e^{-w \cdot c} \tag{3}$$

with the site complexity score c, the weight w, controlling the decline of the function, and the scaling factor b. For each simulation and grid type one of these functions is determined by assigning a weighting for the highest complexity score. In

Table 9 the used weights for each of these for a fictitious site with complexity score of 100 are listed. The resulting functions are then used as an additional weight, which is dependent on the site complexity. Figure 30 shows the site complexity dependent functions for the simulation types (left) and the grid types (right). These functions should only be interpreted as an initial attempt, and depending on the available data these functions can be easily extended and tuned in the future.

Table 9. Weights for different kinds of simulations and grid types for a fictitious site with complexity score of 100

Simulation kind	Weighting for site complexity 100	Grid type	Weighting for site complexity 100
WAsP	0.1	Not represented	0.0001
RANS	0.5	Model	0.01
URANS	0.6	Mesh-conform	0.85
LES	0.85	Combination	0.95
DNS	0.95		

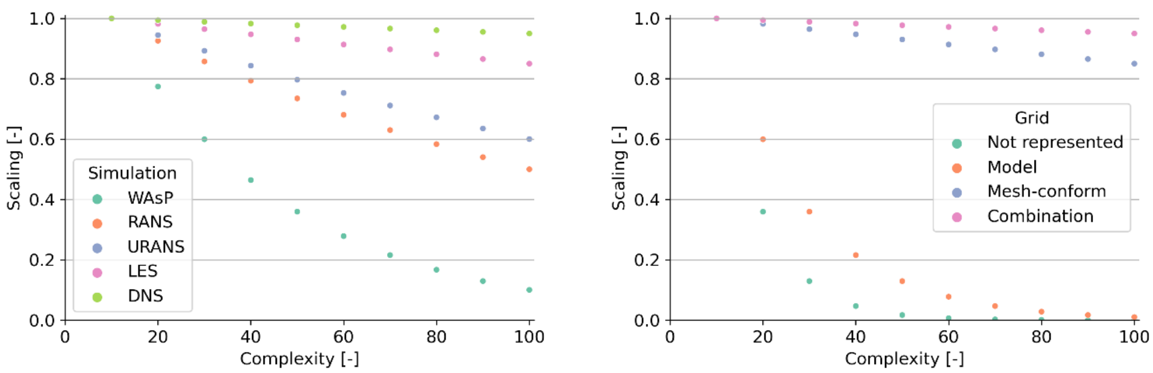


Figure 30. Site complexity dependent functions for the simulation types (left) and the grid types (right)

Another extension of the new method are a confidence skill score and a confidence cost score class. These are used to determine the confidence score. The determination of this score is based on every one of the fields related to the site details, model description, user experience with model and cost estimations. The confidence score allows users to apply the *Decision tool* without knowing many details of the planned set-up. The more details that are known, the higher the confidence score will be and the more accurate the final decision will be. The consideration of the confidence score adds another level of tuning and matching before and after skill scores.

4.2.5 After score metrics

The after skill and cost score metrics were determined according to the methodology given in Section 3.2.5 for Stage 1.

4.2.6 Skill versus costs

In this section the skill versus cost scores for the four sites with different site complexities are analysed. As shown in the table above, the participant used RANS simulations for all sites. Hence, it should be possible to see a clear trend according to the above assumptions that for increasing site complexity the model accuracy decreases. Due to the lack of submission data we decided to produce artificial, but realistic, test cases, denoted by the labels WAsP and LES. For this data a fictitious modeller with a comparable *user skill score* to the Enercon cases was used and the only variables are therefore the *model skill score* and the site complexity. The *model skill scores* were determined by filling out the questionnaire, shown in Table 10. As no real simulations were conducted for the WAsP and LES cases, only the before metrics are available.

Table 10. Fictitious but realistic test cases for the comparison with the Enercon cases

Questions	LES	WAsP
What kind of simulation did you run? (RANS, URANS, LES, LBM-LES, etc.)	LES	WAsP
How did you model surface roughness?	Model	Model
What grid type did you use? (e.g. structured/unstructured, mesh-conform/immersed boundary)	Mesh-conform	-
Did you conduct a grid independency study?	Yes	No
How did you calibrate the wind profile to the one at the calibration / reference mast?	Iteratively through changing input conditions, For entire measurement period	Direct input at met mast (similar to WAsP)
How many different wind directions did you use for the AEP calculation? (e.g. 12 or 24)	8	24
Which method did you use for calculating the wind speed at the rotor?	Hub-height only	Hub-height only
How did you use the wind simulations to obtain the AEP?	Based on time series and the OEM's power curve	Direct simulation output (similar to WAsP)
How did you correct for air density at the site?	None	None

The next step, before calculating and plotting any *before skill scores*, is to determine the weights for each before score associated to a question in Table 10 based on the *after skill scores*, determined through the simulation and measurement results. This step is also part of the business logic of the *Decision tool* and was planned to be implemented in such a way as to allow for easy updates and extensions. In order to determine the weights, the data has to be split and transformed into the parts shown in Figure 31. The score matrix, **S**, contains all scores for each of the k questions in the questionnaire above for all N submissions or rather simulation cases. Hence, the matrix **S** is of size $N \times k$. For each of the k scores there is a weight w_i to be determined. The final after skill scores, y_j, for each case are stored in the vector **y**.

$$\mathbf{S} = \begin{bmatrix} | & | & & | \\ s_1 & s_2 & \cdots & s_k \\ | & | & & | \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} w_1 & w_2 & \cdots & w_k \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

Figure 31. The scores matrix, **S**, based on the questionnaire, the weight vector, **w**, which is to be determined, and the after skill scores vector, **y**.

One way to readily obtain the vector of weights, $\mathbf{w}$, by solving the following equation

$$\boldsymbol{S} \cdot \boldsymbol{w} = \boldsymbol{y}$$

However, the constructed matrices and vectors can also be used with more complex algorithms such as random forests, k-nearest neighbours, support vector machines, multiple regression and other machine learning and statistical algorithms. However, due to the low number of submissions, it was not possible to conduct this step. Instead the weights could simply be tuned by hand.

The results of this manual tuning are shown in Figure 32, where the relative skill versus the relative cost scores for the four sites with increasing site complexities are depicted. As already mentioned above, only the *before skill scores* are available for the artificial data. The first thing to notice is that the complexity weighting functions work as intended. For increasing site complexities, the skill scores decrease for all models. As well as that, the relative distances between the models also increases for increasing complexities. This means that for less complex sites the model accuracies approach each other and the skill score is less dependent on the simulation kind and roughness representation, but is rather governed by other parameters such as the number of simulated wind directions for predicting the annual energy production. As LES simulations are inherently expensive with regard to time and costs, less wind direction can be simulated as opposed to RANS and WAsP simulations, where a large number of simulations can be carried out for reasonable time and cost efforts. However, for more complex sites, LES proves is well ahead in terms of accuracy and it is up to the modeller if the additional costs of running such a case are justified. An important point to remember is, however, the fact that the shown analysis of the *before score metrics* are dependent on a very small number of data points and are subject to change as more data comes in. At any rate, the assumptions made and the conclusions drawn from these results appear to be a sensible starting point.

The actual *after skill and cost score metrics* do not show a dependence on the site complexity. A possible explanation for this might be that other parameters play an important role as well, which, however, are not currently known and could therefore not be considered while tuning the weight vector. More data is needed in order to shed light on this issue.

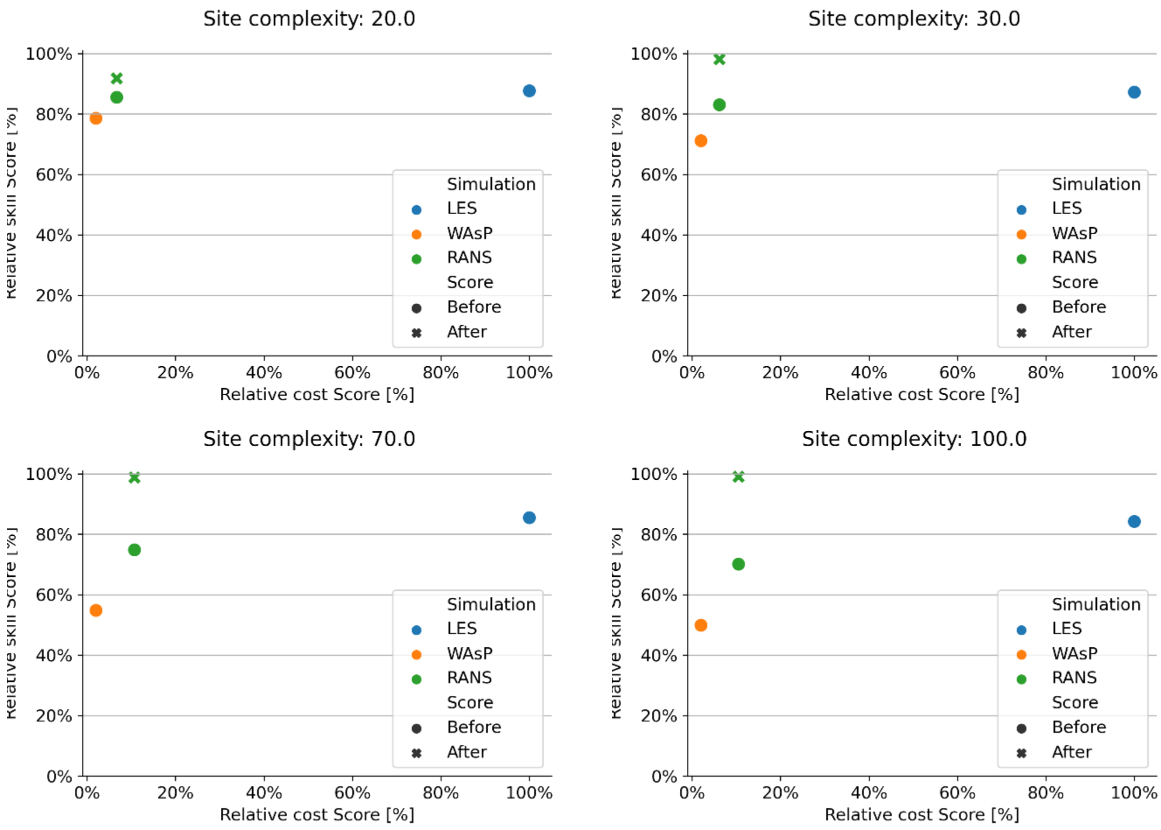


Figure 32. Relative skill versus relative cost scores for four sites with increasing site complexities

4.2.7 Summary of Stage 2

The content of this chapter can be summarised by the following points.

General summary:

1. A new *Decision tool* in form of a web application was developed.
    1. Create surveys that are automatically saved in a database.
    2. Automatic calculation of skill and cost scores.
    3. The business logic is extendable for ML and other statistical models.
2. Received results of four sites with different site complexities:
    1. Not sufficient to draw major conclusions, but good as a starting point.
3. More understanding and evaluation of Comparison Metricss needed:
    1. More data to better draw generalised conclusions.
    2. How to better advertise challenges and increase motivation to participate

Improvements to Comparison Metrics process needed:

4. Defining the Model Description in a Google Form and then estimating skill and cost score parameters in a separate Google Form was not optimal in Stage 1:

   1. User now enters set-up details and the *Decision tool* automatically converts answers to skill score and cost score
5. The values entered into the Google Forms were transferred to an Excel template by hand:
   1. Template then had to be adjusted for each project.
   2. This process was automated in Stage 2

Skill score before estimation:

6. Many of the parameters could only be estimated once the simulations had been fully set up.
   1. A starting point for a method was established that is able to make an estimation of the most optimal model without carrying out any simulations.
   2. The questions used to determine the skill score were adjusted
      1. Reformulated to be easier to answer
      2. More objective questions such as kind of simulation were chosen
   3. Assigning each parameter a confidence score was not possible due to the lack of submissions
7. Improved weightings:
   1. The simulation kind and the roughness representation parameters are now dependent upon the terrain complexity classification.
   2. Weightings can be tuned by machine learning and/or statistical models

Cost scores before and after:

8. After the final workshop for Stage 1 we reached out to the participants to discuss about the calculation of the cost scores. We were able to slightly adjust the process to match the estimations and actual costs of the different simulations for Stage 2.

Complex terrain classification:

9. Difficult to classify the complexity of the terrain by asking questions that the participants could easily answer.
   1. The participants were asked to submit site data comprising terrain maps, roughness maps and weather data, if available, in order to calculate a more objective site complexity score.
10. This task proofed more difficult and time consuming than initially anticipated. Moreover, the lack of data would have rendered the results of this analysis meaningless. Hence, we decided to drop the task and continue with solely the self-evaluated site complexity scores.

5 Conclusions

In this project, two public "simulation challenges" for wind energy sites in complex terrain were implemented, in which participants submit their simulation data and results in a pre-defined template. The goal was to collect hundreds of comparison metrics data regarding the "skill" and "costs" of simulation tools both before and after carrying out the simulations, enabling transfer functions for the accurate prediction of tool "skill" and "costs" to be developed. This aims to help modellers choose the best model for the job for a given wind energy project.

In Stage 1 of the project, a submission template for comparison metrics was developed further and greatly improved compared to an initial version. A simulation challenge for the Perdigao site in Portugal was designed and launched according to plan. Five organisations with a total of 10 different submissions participated in the challenge. Each submission contained vertical wind profiles for a total of nine met mast locations and annual energy production values for two met mast positions. These results allowed for an extensive analysis including speed-up factors and flow turning between various met masts, wind profile comparisons, annual energy production values as well as before and after skill and cost score metrics.As part of the project a Python library for the analysis of the data was developed, which will be made public. The library was enhanced and improved after the final Stage 1 workshop based on feedback and discussions with the participants. The results of Stage 1 showed that sophisticated simulation tools such as LES do not necessarily lead to higher accuracies. Especially for less complex locations one is better off using simpler tools such as RANS or WAsP, reaching high levels of accuracy with a fraction of the costs of LES simulations. Overall, the RANS simulation with the E-Wind software, developed by Enercon, achieved the best and most consistent scores. As the participant was also the most experienced amongst all, this might lead to the conclusion that the user skill plays a crucial role for the overall skill score.

For Stage 2 of this project, the manual process of Stage 1 was automated. The resulting *Decision tool* is able to automatically convert answers of questionnaires into skill and cost scores. In order to develop the business logic of the tool, a new challenge was published. In this challenge participants were asked to upload simulation and measurement results of any available site. Based on these results site complexity dependent functions and score weightings were supposed to be developed and tuned. However, due to a lack of participation the tuning part could not be completed. However, the resulting methods serve as a starting point and can be easily updated and extended as more data comes in. Additionally, the questions of the questionnaire were reformulated to render them easier to answer and understand as well as more objective. The influence of site complexity on the skill versus cost score plots was briefly explored by comparing three different models for four sites with increasing complexity. The used models comprised the submitted RANS simulation results by Enercon as well as artificially generated, but nonetheless realistic, data based on the developed questionnaires. This, however, only allowed for an analysis of the *before score metrics*. It was shown that for increasing site complexity the before skill scores decreased and the relative distance between each model increased dependent on its sophistication, i.e. LES outperformed RANS and WAsP simulations. In turn, for less complex sites RANS and WAsP simulations performed similar to the LES case, but had significantly better cost scores. The developed functions and weights that were able to achieve these insights were tuned by hand due to the very few available data points. With more data machine learning and statistical models can replace this manual process in order to get more generalised and reliable results. The current methods and functions serve as a starting point for further development.

6 Outlook and next steps

This project has been completed successfully. The next steps planned are:

- Complete the documentation of the Python library and make it public on GitLab
- A journal paper is currently written for Stage 1 of this project together with the participants. It is expected to be published in the Wind Energy Science journal by the end of next year.
- Submit a journal paper about the *Decision tool* to the open access journal "Journal of Open Source Software"
- Submit abstracts to the TORQUE2023 conference.
- Possible continuation and data gathering with a commercial partner in order to improve and tune the *Decision tool* using existing contacts with the global companies Enercon, UL and TÜV Süd.

7 National and international cooperation

This project involves the following national and international collaborations:

- International:
    - This work is a collaboration with IEA Wind Task 31. The challenge was first introduced at the IEA Wind Task 31 meeting in Amherst, USA during the AWEA/Windtech conference in October 2019. A progress report was presented at the online IEA Wind Task 31 meeting in June 2020.
    - Participants in this challenge come from all over the world, as can be seen in Table 3.
- National:
    - The work is based on the methods that are being developed together with the Swiss partner Meteotest AG in the separate SFOE-funded project "A new process for the pragmatic choice of wind models in complex terrain".

8 Communication

The following communication has been used in the first year of this project:

- The challenge was first introduced at the IEA Wind Task 31 meeting in Amherst, USA during the AWEA/Windtech conference in October 2019. A progress report was presented at the online IEA Wind Task 31 meeting in June 2020.
- The challenge was published on The Wind Vane Blog [4].
- The challenge has been posted on LinkedIn multiple times.
- The challenge Stage 1 design was presented in a poster at the WindEurope Wind Resource Assessment Workshop in June 2020 (online), which was delayed from April 2020.
- The challenge Stage 1 design was presented in a poster and a paper at the Torque2020 conference in September 2020 (online), which was delayed from May 2020.
- The Stage 2 challenge was published on the WeDoWind Platform.

9 Publications

Stage 1 of the challenge is published here, and is continually updated:
https://thewindvaneblog.com/comparison-metrics-microscale-simulation-challenge-for-wind-resource-assessment-stage-1-3d0f88cff313

The rest of the publications related to this work can all be found under Sarah Barber's profile on the Zenodo platform. The links are given as hyperlinks below:

- S. Barber, "Comparison metrics microscale simulation challenge for wind resource assessment – stage 1", Launch webinar, April 2020.
- S. Barber, M. Buehler, H. Nordborg, "IEA Wind Task 31: Initial results of a new comparison metrics simulation challenge for wind resource assessment in complex terrain", WindEurope Wind Resource Assessment Workshop, June 2020 (poster).
- S. Barber, M. Buehler, H. Nordborg, "IEA Wind Task 31: Design of a new comparison metrics simulation challenge for wind resource assessment in complex terrain Stage 1", J, September 2020 (poster).
- S. Barber, M. Buehler, H. Nordborg, "IEA Wind Task 31: Design of a new comparison metrics simulation challenge for wind resource assessment in complex terrain Stage 1", Journal of Physics: Conference Series 1618 062013, doi: https://doi.org/10.1088/1742-6596/1618/6/062013.

10 References

[1] S Barber, A Schubiger, and H Nordborg. A new process for the pragmatic choice of wind models in complex terrain – quantification of ”skill score” and ”cost”. Wind Energy Science Conference, https://zenodo.org/record/3274236#.X4lcYObiuUk.

[2] J Berg, J Mann, A Bechmann, M S Courtney, and H E Jørgensen. The Bolund Experiment, Part 1: Flow over a steep, three-dimensional hill. Boundary-Layer Meteorology, 141(2):219–243, 2011.

[3] IEA Wind Task 31 https://community.ieawind.org/task31/home, 2019.

[4] https://thewindvaneblog.com/comparison-metrics-microscale-simulation-challenge-for-wind-resource-assessment-stage-1-3d0f88cff313.

[5] A Clifton. Remote sensing of complex flows by doppler wind lidar: Issues and preliminary recommendations; technical report nrel/tp-5000-64634.

[6] IEC International Electrotechnical Commission. Wind energy generation systems - part 12-1: Power performance measurements of electricity producing wind turbines, 03.03.2017.

[7] N G Mortensen, A J Bowen, and I Antoniou. Improving WAsP predictions in (too) complex terrain. Proceedings, European Wind Energy Association, 100(5):799–819, 2006.

[8] S Barber. A new method for the pragmatic choice of wind models for Wind Resource Assessment in complex terrain, 10.23728/b2share.b5650658062047189649c8755dec79fc, 2019.

[9] D F Gomes, J C Lopes, J L. Palma, G Lopes, I Fernandes, R Abreu, and J Fernandes. Perdigao field experiment https://perdigao.fe.up.pt, 2019.

[10] H J S Fernando et al. The Perdigão: Peering into microscale details of mountain winds. Bulletin of the American Meteorological Society, 100(5):799–819, 2019.

[11] DTU Wind Energy. WAsP https://www.wasp.dk/.

[12] T Yamaguchi and G Feingold. Technical note: Large–eddy simulation of cloudy boundary layer with the advanced research WRF model. Journal of Advances in Modeling Earth Systems, 4(3), 2012.

[13] S Barber, A Schubiger, S Koller, A Rumpf, H Knaus and H Nordborg. Actual Total Cost reduction of commercial CFD modelling tools for Wind Resource Assessment in complex terrain. J. Phys.: Conf. Ser. 1618 062012, https://doi.org/10.1088/1742-6596/1618/6/062012, 2020.

11 Appendix

11.1 Simulation setups for Stage 1 challenge

Table 11. Simulation setup details

	Model 1 and 2	Models 3 - 5	Model 6	Model 7	Model 8 and 9	Model 10
Wind model	UTD-WF	E-Wind	Fluent	WAsP	OpenFOAM	WAsP
Institution	UTD	Enercon	OST	OST	VKI	Alten
Model type	LES Smagorinsky SGS	Steady state RANS k-epsilon, k-L and k-omega turbulence models	Steady state RANS k-omega SST turbulence model	Linear wind model	Steady state RANS k-epsilon turbulence model	Linear wind model
Grid dimensions (resolution)	5.4 x 3.7 x 2 km (4 m)	Circular domain with 11 km radius, height 6 km (25 m)	10 x 10 x 1.5 km (20 m)	6 x 4 km (automatic)	5 x 5 x 4 km (Not specified)	11 x 11.5 km (100 m)
Number of cells	Not specified ~ 60 million	1.8 million	20 million	Not specified	Not specified	Not specified ~ 126'000
Input profile	Above topography with a shear exponent of 0.68	Precomputed numerical profiles	Log law profile	Met mast 29 data	Log law profile	Met mast 29 data
Turbulence intensity	None	None	Energy k and turbulence dissipation rate ε in the ABL are based on the Harris and Deaves (1981) model	None	None	None
Other models	PCE surrogate model, Canopy model for second simulation	Buoyancy effects included in the turbulence equations, Coriolis force, forest model, atmospheric stability modeled by surface heat flux	Sand grain roughness model	None	Porosity model	None

Wind directions	8 (7°, 36°, 85°, 147°, 213°, 275°, 323°, 353°)	24	12	12	12	12