Towards data-driven turbulence modeling for wind turbine wakes

Steiner, J.

DOI
10.4233/uuid:72116acd-c5aa-4b3b-8fc7-52f1b2fa9958

Publication date
2023

Document Version
Final published version

Citation (APA)
Steiner, J. (2023). *Towards data-driven turbulence modeling for wind turbine wakes*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:72116acd-c5aa-4b3b-8fc7-52f1b2fa9958

TOWARDS DATA-DRIVEN TURBULENCE MODELING FOR WIND TURBINE WAKES

TOWARDS DATA-DRIVEN TURBULENCE MODELING FOR WIND TURBINE WAKES

Dissertation

for the purpose of obtaining the degree of doctor
at the Delft University of Technology
by the authority of the Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Thursday 14 December 2023 12:30 o'clock

by

Julia STEINER

Master of Science in Aerospace Engineering,
Technical University of Delft, Delft, Netherlands,
born in Frauenfeld, Switzerland.

This disseration has been approved by the promoters.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. S. Watson	Delft University of Technology, promoter
Prof. dr. ir. A.C. Viré	Delft University of Technology, promoter
Dr. R.P. Dwight	Delft university of Technology, co-promoter

Independent members:

Prof. dr. H. Xiao	Stuttgart University, Germany
Dr. M. Abkar	Aarhus University, Denmark
Prof. dr. S. Basu	Unversity at Albany, United States of America
Prof. dr. D. von Terzi	Delft University of Technology
Prof. dr. ir. J.W. van Wingerden	Delft University of Technology, reserve member

Keywords: RANS, turbulence modeling, wind energy & machine learning.

ISBN 978-94-6366-793-7

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

CONTENTS

SUMMARY

The Dutch energy strategy expects renewable energy sources like wind and solar to provide around 70% of the yearly electricity by 2030. In order to achieve these targets, it is crucial to ensure cost-effective design and operation of wind farms. Optimizing design variables, such as wind farm layout, requires fast and accurate numerical models. Simple engineering models fail to capture the flow physics associated, for example, with wind turbine wakes. An alternative is to use Reynolds-Averaged Navier-Stokes (RANS) solvers. However, these models have structural shortcomings for industrial applications and development of better models has stalled in the past decades. More recently, data-driven techniques have been used to try and derive better, application-specific models. The field is still very young and it is unclear how much impact can data-driven techniques make in the field of RANS-based turbulence modeling.

In this work, a combined methodology between a k-corrective frozen-RANS approach and sparse symbolic regression is used to derive data-driven nonlinear RANS turbulence models for flow around scaled wind turbines. The resulting models give significantly better predictions than the baseline k-ε model for both velocity and turbulent kinetic energy (tke) in the near and far wake of the turbines with a very close match between reference LES and corrected RANS data. For the velocity and the tke fields, the errors as compared to the baseline models are reduced by 95% and 80%, respectively. The dataset includes three test cases with either two or three turbines. The training is first done on one test case, and then validated on the two other ones.

The present approach was initially developed by Schmelzer et al. [79], who referred to it as SPARTA, and applied to 2D test cases with moderate Reynolds numbers. For this approach to become usable for wind engineering applications, significant changes were necessary. Besides this, in literature, many recent examples of data-driven turbulence modeling can be found. However, the large majority of these publications focus on 2D benchmark cases with low to moderate Reynolds numbers. In this work, results from 3D cases at a Re $\approx 100,000$ were presented on a more industrially relevant case. This highlights the novelty of this work.

While good results were obtained, similar to traditional Nonlinear Eddy Viscosity Models, the models initially showed numerical instability. These instabilities were more or less severe depending on model complexity, mesh density, and input feature set. Ensuring numerical stability of the models was possible by introducing two simple general limiters that are active only in a very limited amount of cells, where a positive coupling loop was responsible for the divergence of the entire simulation.

To conclude, the present work demonstrates the applicability and shows the limitations of the SPARTA data-driven modeling approach for wind turbine wake predictions, which is far more complex than the cases investigated before. For the future, the application of the methodology to a larger dataset that also includes full-scale wind turbines is recommended. This would allow to assess how general the derived models are and how

much training data is necessary to obtain usable models. Additionally, the efficiency of the algorithm used for training the models can be a limiting factor and smart subsampling of the input dataset or modification of the algorithm may become necessary for larger datasets.

SAMENVATTING

De Nederlandse energie strategie bevat de verwachting dat hernieuwbare bronnen zoals wind en zon 70% van de jaarlijkse elektriciteitsvoorziening zal beslaan tegen het jaar 2030. Om dit doel te bereiken is het cruciaal de kosteneffectiviteit te waarborgen van het ontwerp en de uitvoering van windmolenparken. Het optimalizeren van ontwerpvariabelen, zoals de indeling, vergt snelle en accurate numerieke modellen. Simpele technische modellen zijn niet in staat om de vloeistofdynamica te beschrijven van bijvoorbeeld de zog van de turbines. Een alternatief hiervoor is het gebruik van Reynolds-gemiddelde Navier-Stokes (RANS) computermodellen. Echter, deze modellen bevatten structurele gebreken voor het gebruik in industrie en de ontwikkeling van betere modellen is achtergebleven in de afgelopen decennia. Recenter is gepoogd datagedreven technieken te gebruiken om applicatie-specifieke modellen af te leiden. Het onderzoeksveld is echter nog jong en het is onduidelijk hoeveel datagedreven technieken teweeg kunnen brengen in het veld van RANS-gebaseerde turbulentiemodellering.

In het huidige werk is een gecombineerde methodologie van een k-correctieve bevroren RANS benadering en ijle symbolische regressie gebruikt om datagedreven niet-lineaire RANS turbulentiemodellen voor stroming rondom geschaalde windturbines af te leiden. Het resulterende model maakt significant betere voorspellingen dan het $k-\epsilon$ basismodel voor zowel snelheid als tubulente kinetische energie (tke) in de nabije alsmede de verre zog van de turbines met een nauwe overeenkomst tussen de referentie LES en gecorrigeerde RANS data. De fouten van de snelheids- en tke velden zijn in vergelijking met het basismodel afgenomen met respectievelijk 95% en 80%. De dataset bevat drie testsituaties met twee of drie turbines. De training is uitgevoerd met één testsituatie, en dan gevalideerd met de twee andere.

De huidige aanpak was ontwikkeld door Schmelzer et al. [79] onder de naam SPARTA, die het toepaste op 2D testsituaties met Reynoldsgetallen van gematigde waarden. Om deze aanpak toepasbaar te maken voor de windtechniek waren significante aanpassingen noodzakelijk. Daarbij zijn er in de literatuur vele recente voorbeelden van datagedreven turbulentie te vinden. De meesten van deze publicaties richten zich echter op 2D maatstafsituaties waarbij lage tot gematigde Reynoldsgetallen worden gebruikt. In het huidige werk zijn resultaten van 3D situaties met $Re \approx 100,000$ gepresenteerd in een situatie relevant voor de industrie. Hiermee wordt de nieuwigheid van dit werk uitgelicht.

Hoewel er goede resultaten zijn behaald, vergelijkbaar met traditionele niet-lineaire wervelviscositeitsmodellen, toonden de modellen numerieke instabiliteit. De ernst van deze instabiliteiten was variabel, afhankelijk de complexiteit van het model, de dichtheid van de mazen, en de meegenomen karakteristieken. Numerieke stabiliteit van de modellen was gewaarborgd met behulp van twee simpele algemene begrenzers die alleen actief waren in een zeer beperkt aantal cellen waar een positieve koppelingscyclus verantwoordelijk was voor de divergentie van de gehele simulatie.

Ter conclusie, het huidige werk demonstreert de toepasbaarheid en toont de limitaties van het gebruik van het SPARTA datagedreven model voor voorspellingen van de zog van windturbines, wat veruit complexer is dan eerder bestudeerde situaties. Voor de toekomst wordt toepassing van de methodologie op een grotere dataset, die ook windturbines bevat op volledige schaal, aangeraden. Hierdoor is het mogelijk om te kunnen beoordelen hoe algemeen de afgeleide modellen zijn en hoeveel trainingsdata nodig is voordat de modellen als bruikbaar kunnen worden beschouwd. Daarnaast kan de efficiëntie van de algorithmes die zijn gebruikt voor het trainen van de modellen een limiterende factor zijn, waardoor slimme onderbemonstering van de invoerdata of aanpassing van het algorithme noodzakelijk mag blijken voor grotere datasets.

1

INTRODUCTION

The Dutch energy strategy expects offshore wind farms to provide 49 TWh/yr of electricity by 2030, whereas onshore wind energy and solar should provide 35 TWh/yr, which together corresponds to about 70% of the current yearly electricity consumption [75]. Even in Switzerland where average wind speeds are much lower, a recent estimate puts the potential around 10 TWh per year which is roughly 15 % of the yearly electricity consumption, two-thirds of which is in winter when solar and hydropower are insufficient [25].

Crucial for cost-effective wind farms are the optimal design of individual turbines, the optimal design of the farm layout, and accurate models of the energy output of the farm once it is operational for optimizing control strategy and prediction of the farm performance. Wind power is an interdisciplinary field and hence accurate structural, aerodynamic, electrical, cost, and weather models are necessary to achieve such optimization. Furthermore, the interaction between these models is complex. This thesis will focus on aerodynamic models that are used at different levels in the design phase of wind farms, to analyze: (i) airfoil performance, (ii) the performance of a full blade or rotor, (iii) the interaction between turbines, and their surroundings (e.g. terrain), in a farm, and (iv) the interaction of entire wind farms with the atmospheric boundary layer.

1.1. LIMITATIONS OF CURRENT AERODYNAMIC MODELS

In the design phase, wind farm layouts, and inflow conditions have to be considered. Models that are used once the farm has been built need to be accurate and fast to allow for smart control of the farm. Generally speaking, this requires a trade-off between, on the one hand, fast but inaccurate models and, on the other hand, computationally expensive but accurate models. Engineering models for predicting the aerodynamics of wind turbines and farms exist but only work well for very specific configurations such as when there is little interaction between the different turbines wakes. An alternative is to use computational fluid dynamics (CFD) models, which should improve the applicability of the wake models to configurations with wake interaction, but may be very computationally expensive. Currently, the state of the art in CFD-based aerodynamic modeling

of wind turbines on all the different mentioned design levels from the small turbulent structures on the blade to the interaction of the wind farm with the atmospheric boundary layer are large-eddy simulation (LES) or hybrid RANS/LES approaches which resolve a large part of the spectra of the turbulent fluctuations [11]. Reynolds-Averaged Navier-Stokes (RANS) models are low-fidelity CFD models that are about two orders of magnitude less computationally expensive than LES-based approaches. If CFD is used for wind engineering purposes in industry, it is mostly RANS-based models that are used. RANS models provide mean values for velocity and pressure, and first-order statistics for turbulent properties. However, they suffer from structural shortcomings for flows of industrial interest and are better at predicting tendencies than accurate values for complex flows. For example, the stall angle for thick wind turbine airfoils is frequently overpredicted especially for thick airfoils used in the root region of wind turbines [108]. Also, for wind farms, standard RANS models such as the k-ε and the k-ω (SST) tend to overpredict the wake recovery and hence over-predict the energy yield of a wind farm [77].

1.2. DATA-DRIVEN RANS TURBULENCE MODELS

Shortcomings of RANS turbulence models are well known beyond the wind energy community and the development of better RANS models has stalled in the past decades. Application-specific tuning of the model parameters can help, but this cannot fully overcome the structural shortcomings of the models. More recently, with the advent of machine learning in all other scientific areas, data-driven techniques have been used to derive more accurate, albeit still application-specific, models [18]. In CFD modeling, data-driven approaches are still in their infancy, but there is a need to explore if they have the potential for improving existing RANS-based models. In the short term, data-driven techniques may fail to derive improved general-purpose models, but can quite possibly help derive case-specific improvements given some high-fidelity data.

1.3. MOTIVATION

For wind farms, the most frequently used turbulence model is the k-ε model, however, it suffers from two main shortcomings: (i) the Boussinesq hypothesis which is not a good model in strongly non-equilibrium flow such as the near wake of a wind turbine, and (ii) the rotor model which interacts with the turbulence model only indirectly. The Boussinesq hypothesis can have different meanings in the field of flow modeling, in this instance, it refers to assuming that the turbulence anisotropy is proportional to the mean velocity based strain rate in the flow. As a consequence, it over-predicts the turbulent kinetic energy (tke) in the near wake, and this over-predicted turbulent mixing leads to a wake that recovers too quickly. In literature, improved models that limit the tke production in the near wake through some sort of shear limiter exist, but they do not reliably improve the velocity, tke, and turbulence anisotropy profiles in the wake. Hence, there is room for improvement. Further, if a wind farm is placed in complex terrain, the limitations of the turbulence model will cause inaccurate predictions for the same reasons.

Given how well-defined the shortcomings of the existing models are, there is a clear motivation for testing how a data-driven approach can help improve predictions for wind engineering flows without increasing the computational cost further. To put it

more concisely:

The objective of this thesis is to work towards an efficient methodology for data-driven turbulence modeling for wind turbine wakes by using time-averaged LES data, comparing the LES data with a baseline RANS model, learning corrections to a baseline RANS model using simple explicit regression methods and finally testing said data-enhanced RANS models.

The starting point for this work is the previously described known shortcomings of RANS turbulence models and the method employed by Schmelzer et al. [79] to derive data-driven turbulence models. The authors developed a framework which injects data from a high-fidelity model into the turbulence equations of a low-fidelity model using a so called frozen approach. The results of the frozen approach are optimal correction terms which ideally lead to a perfect match between the low- and the high-fidelity model when injected into the equations of the low-fidelity model. As a second step, the authors used a regression procedure referred to as SpaRTA (Sparse Regression of Turbulent Stress Anisotropy) to derive explicit expressions for enhanced turbulence models. This work improves this methodology and makes it more industrially applicable through adjustments to the methodology, the efficiency of the underlying code base and the usage of larger datasets at higher Reynolds numbers. The dataset in this work contains about fifty times more data points than the original ones and the Reynolds number is around six times higher as well.

Finally, it is important to stress that the aim of this work is not to derive a methodology that is capable of deriving a general-purpose turbulence model. Instead, this dissertation works towards a methodology that is capable of deriving models which are helpful for a specific situation or application. Or more specifically, the goal is to be able to build a better RANS model for a specified location and turbine type given a time-averaged LES simulation of a few turbines. This is already a step in the right direction for addressing the shortcomings mentioned above. Such an improved model could then be used for layout and control optimization. Once the methods of data-driven turbulence modeling are mature enough to efficiently derive data-enhanced models that are accurate and numerically robust, the question of generalizability can be addressed.

1.4. OUTLINE

The dissertation is structured as follows. Chapter 2 reviews the existing literature encompassing the derivation of the most widely-used turbulence models and their limitations, the state-of-the-art in wind farm physics modeling, specific analysis of shortcomings of the baseline k-ε model, and finally, an introduction to the world of data-driven turbulence modeling. In Chapter 3, the methodology from data generation to the training of the correction terms is presented. In Chapter 4, the results from each step of the methodology is presented and critically analyzed. Finally, the conclusions and outlook on future developments are presented in Chapter 5.

2

LITERATURE REVIEW

The literature review presented in this chapter gives an overview of how computational fluid dynamics is used to model the flow in wind farms and what the limitations are. This thesis aims to develop a methodology to improve low-fidelity turbulence models given high-fidelity reference data. In this work, the low-fidelity model will be a Reynolds-Averaged Navier-Stokes (RANS) model, and the high-fidelity model will be time-averaged Large Eddy Simulation (LES). Other combinations are possible. For example, some authors use a Direct Numerical Simulation (DNS) high-fidelity model to provide data, while others use measurement data. In Section 2.1, this review first briefly introduces the reader to the relevant low- and high-fidelity turbulence models and what assumptions are made in the derivation thereof. Understanding how those models are derived will help understand why they fail to predict wind turbine wakes accurately. The focus is on CFD-based models, thus engineering models, such as the Jensen model [37], are not presented here. Then, in Section 2.2, a review of different case studies on wind turbine wake modeling using CFD is presented. This is meant to introduce the state of the art for both RANS and LES models as used for wind engineering purposes. Following this, Section 2.3 presents a detailed comparison between the most frequently used RANS model and a state of the art LES model. This will highlight the very specific shortcomings of the low-fidelity model and tie together the first two sections of this review. Finally, Section 2.4 presents an overview of data-driven approaches in turbulence modeling for all applications.

2.1. INTRODUCTION TO TURBULENCE MODELING

The main characteristic of turbulent flow is unsteadiness; the flow can be seen to be comprised of eddies of different sizes. The largest eddies of the flow are subject to geometry (for example a wall or an obstacle) and forcing (for example, a driving pressure gradient or an actuator disk) and are responsible for the majority of the turbulent kinetic energy (tke) production. As most of the turbulent kinetic energy is carried in the large scales, they are referred to as the energy-containing scales. Energy is transferred from

larger to smaller eddies, as eddies continuously break down into smaller ones. This is called the turbulent energy cascade. The rate at which the kinetic energy in the eddies is converted into internal thermal energy is the *dissipation rate* ε. The smallest eddies pertain to the dissipative range, have dynamics dominated by viscous effects, and do not significantly affect the large scales except by acting as an energy drain by converting tke to heat. If the Reynolds number is sufficiently large, there exists an intermediate range of scales in the so-called inertial subrange. The energy contained in these scales has a universal form determined by ε; and this range acts mainly in transferring energy towards the dissipation range. In contrast to the dissipation range, the inertial range is dominated by inertial effects. The size of the eddies in the dissipation range scales with the viscosity ν and ε, whereas the size of the eddies in the energy-containing range is related to the geometry of the flow. Consequently, for flows with high Reynolds numbers, the range of scales of the flow becomes very large, such that in computational fluid dynamics not all scales can be resolved. Instead, the turbulent scales have to be (at least partially) modeled by a turbulence model.

There are many ways of achieving this aim within computational fluid dynamics. The two main categories of methods are Large Eddy Simulations (LES) and Reynolds-Averaged Navier-Stokes (RANS) methods. In LES, parts of the turbulence spectrum are resolved, usually including some portion of the inertial range, and the energy removed from the flow through the unresolved smaller scales is modeled through a sub-grid scale model. By contrast, in RANS, a model is used for the effects of all turbulent scales. Generally, RANS simulations are computationally less expensive than LES ones but also less accurate. Hence, there is a trade-off between computational efficiency and modeling quality. It is worth mentioning that many more turbulence modeling frameworks have been developed, including: (i) "hybrid" models which use both RANS and LES in different regions of the flow such as Delayed Eddy Simulation (DES), Delayed Dettached Eddy Simulation (DDES) & Improved Delayed Dettached Eddy Simulation (IDDES), [85] (ii) unsteady variants of RANS models such as unsteady RANS (URANS) and Partially-Averaged Navier-Stokes (PANS) [27], and lastly (iii) variants of LES models such as Wall-Modeled LES (WMLES) [9]. In this work, only RANS and LES are used, hence a brief introduction to both methods is given hereafter. Note that the equations given apply to the modeling of wind turbines in the atmospheric boundary layer. As such, they are formulated to include Coriolis forcing, non-neutral stratification (buoyancy), and actuator forcing [1]. The framework uses the incompressible formulations of the equations. However, it includes small variations in density through the Boussinesq approximation for density - a linearized version of compressibilty effects for atmospheric flows assuming $(\rho - \rho_0)/\rho_0 << 1$ [51].

2.1.1. REYNOLDS-AVERAGED NAVIER-STOKES SIMULATIONS

RANS methods solve the time-averaged Navier-Stokes equations and are obtained by decomposing all fields ϕ into mean and temporally fluctuating components $\phi = \overline{\phi} + \phi'$. Then, the Navier-Stokes equations are solved for the mean fields and all the turbulent fluctuations are modeled. The equations for the mass, momentum, and energy in the

RANS framework are:

$$\frac{\partial \overline{u}_i}{\partial x_i} = 0, \tag{2.1}$$

$$\frac{\partial \overline{u}_i}{\partial t} + \overline{u}_j \frac{\partial \overline{u}_i}{\partial x_j} = -\frac{1}{\rho_0}\frac{\partial \overline{p}}{\partial x_i} + \nu_0 \frac{\partial}{\partial x_j}\left(\frac{\partial \overline{u}_i}{\partial x_j} + \frac{\partial \overline{u}_j}{\partial x_i}\right) - \frac{\partial \tau_{ij}}{\partial x_j} + \overline{f}_i + f_C \varepsilon_{ij3}\overline{u}_j + g\frac{\overline{\theta} - \overline{\theta}_{ref}}{\rho_0 \theta_0}, \tag{2.2}$$

$$\frac{\partial \overline{\theta}}{\partial t} + \overline{u}_j \frac{\partial \overline{\theta}}{\partial x_j} = -\frac{\partial q_j}{\partial x_j}, \tag{2.3}$$

where $\overline{u}$, $\overline{p}$ and $\overline{\theta}$ are the mean (i.e. ensemble-averaged) velocity, pressure, and potential temperature fields, $\overline{f}$ is a volume forcing term for the actuator disk, Θ_0 and ρ_0 are the reference values for the Boussinesq approximation, ϵ_{ij3} is an alternating unit tensor, and f_C is the Coriolis constant. The influence of the turbulence on the mean flow is contained in the turbulent flux terms τ_{ij} and q_j. The nomenclature for the Reynolds stress and the turbulent heat flux is $\tau_{ij} = \overline{u_i' u_j'}$ and $q_j = \overline{u_j'\theta'}$, and they must be modeled.

Similar to the Navier-Stokes equations for the mean fields, exact equations for the Reynolds stresses can be derived. While these equations cannot be directly solved - "the closure problem of turbulence" - because they contain an even higher-order combination of turbulent fluctuations, they are a helpful starting point for developing turbulence models. The exact transport equation is:

$$\underbrace{\frac{\overline{D}}{\overline{D}t}\overline{u_i' u_j'}}_{\text{Convection}} + \underbrace{\frac{\partial}{\partial x_k}\mathscr{T}_{kij}}_{\text{Diffusion}} = \underbrace{\mathscr{P}_{ij}}_{\text{Production}} + \underbrace{\mathscr{R}_{ij}}_{\text{Redistribution}} - \underbrace{\varepsilon_{ij}}_{\text{Dissipation}} \tag{2.4}$$

$$\mathscr{P}_{ij} \equiv -\overline{u_i' u_k'}\frac{\partial \overline{u}_j}{\partial x_k} - \overline{u_j' u_k'}\frac{\partial \overline{u}_i}{\partial x_k} \tag{2.5}$$

$$\varepsilon_{ij} \equiv 2\nu \overline{\frac{\partial u_i'}{\partial x_k}\frac{\partial u_j'}{\partial x_k}} \tag{2.6}$$

$$\mathscr{R}_{ij} \equiv \overline{\frac{p'}{\rho}\left(\frac{\partial u_i'}{\partial x_j} + \frac{\partial u_j'}{\partial x_i}\right)} \tag{2.7}$$

$$\mathscr{T}_{kij} = \frac{1}{\rho}\overline{u_i' p'}\delta_{jk} + \frac{1}{\rho}\overline{u_j' p'}\delta_{ik} + \overline{u_i' u_j' u_k'} + -\nu\overline{\frac{u_i' u_j'}{\partial x_k}} \tag{2.8}$$

By taking the trace of these equations, an exact transport equation for the turbulent kinetic energy k can be derived:

$$\frac{\overline{D}k}{\overline{D}t} + \frac{\partial}{\partial x_i}\left(\frac{1}{2}\overline{u_i' u_j' u_j'} + \overline{\frac{u_i' p}{\rho}}\right) = \nu\frac{\partial^2 k}{\partial x_i^2} + \mathscr{P} - \varepsilon \tag{2.9}$$

where the total turbulence production is half the trace of the turbulence production tensor $\mathscr{P} = \frac{1}{2}\mathscr{P}_{ii}$ and the same for the total dissipation $\varepsilon = \frac{1}{2}\varepsilon_{ii}$.

The third-order moments and the derivatives of the velocity fluctuations in equation (2.4) need to be modeled for developing a turbulence model. A brief overview of how different classes of turbulence models are derived based on the Reynolds stress transport equation is given in the following. More information can be found in Chapters 10 and 11 of Pope [68].

From the Reynolds stress transport equation, the terms that require modeling are the transport term $\mathscr{T}_{ij}$, the pressure-redistribution-tensor $\mathscr{R}$, and the dissipation tensor ε_{ij}. The production tensor $\mathscr{P}_{ij}$ does not contain any higher order terms, and hence it does not need to be modeled.

The simplest term to model is the dissipation tensor. For high Reynolds number flows away from the wall, dissipation isotropy is a reasonable assumption. Hence the term is modeled as

$$\varepsilon_{ij} \simeq \frac{2}{3}\varepsilon\delta_{ij}, \tag{2.10}$$

where $\varepsilon := \frac{1}{2}\varepsilon_{ii}$. While this simplifies the dissipation tensor to a scalar, a model for this scalar still needs to be constructed. This is done through a transport equation that is empirically constructed, in a very similar way as for simpler two-equation models explained below. Thus, it is not repeated here.

The Reynolds-Stress transport term $\mathscr{T}_{ij}$ is responsible for the turbulent transport of the Reynolds stress, and the material derivative is responsible for the transport due to the mean flow. Simple modeling approaches seem to yield satisfactory results, such as the one from Harlow [31]:

$$\mathscr{T}'_{kij} \simeq -C_s \frac{k}{\varepsilon}\overline{u'_k u'_l}\frac{\partial \overline{u'_i u'_j}}{\partial x_l} \tag{2.11}$$

Lastly, the pressure-rate-of-strain tensor is responsible for the redistribution between the different Reynolds stress components. This is the most challenging term to model, and many approaches exist. No specific ones will be presented here. However, how exactly this term is modeled will determine which class of turbulence model is used. This has important implications for the model's accuracy and computational robustness. Physics dictates that for inhomogeneous flows, redistribution is not a purely local phenomenon, so the flow at one location depends on the flow properties in its proximity. Models which incorporate this and model the pressure-rate-of-strain tensor nonlocally are referred to as **elliptic relaxation models** because an additional elliptic equation is solved to obtain the tensor. If the pressure-rate-of-strain tensor is modeled through the local Reynolds stress, dissipation, and mean velocity gradient only, the model class is referred to as a **Reynolds Stress model (RSM)**. This then requires solving six transport equations for the Reynolds stresses and one for the dissipation.

A simpler, but also less accurate, class of models named **algebraic stress models (ASM)** can be obtained by simplifying the left hand side of (2.4) such that the transport equation for the Reynolds stresses turns into an algebraic equation. Rodi [76] first proposed the following simplification which is referred to as the weak equilibrium condition:

$$\frac{D\left(\overline{u_i' u_j'}/k\right)}{Dt} = 0. \tag{2.12}$$

Decomposing the Reynolds stress accordingly into intensity and anisotropy

$$\overline{u_i' u_j'}/k = 2b_{ij} + \frac{2}{3}\delta_{ij}, \tag{2.13}$$

highlights why this is only a weak and not a full equilibrium condition, as only the anisotropic part of the tensor are assumed to be in equilibrium with its surroundings, and variations in the Reynolds stresses due to the turbulent kinetic energy are retained.

Applying the simplification in (2.12) to the left hand side of the Reynolds stress transport (2.4) leads to the following simpler implicit algebraic expression:

$$\frac{\overline{u_i' u_j'}}{k}(\mathscr{P} - \varepsilon) = \mathscr{P}_{ij} + \mathscr{R}_{ij} - \frac{2}{3}\varepsilon\delta_{ij}. \tag{2.14}$$

For algebraic stress models, in addition to the above equation, also a transport equation for the tke and dissipation needs to be solved.

The next simpler class of models is **Nonlinear Eddy Viscosity Models (NLEVM)**. In this class, an explicit expression for the k-normalized Reynolds stresses is available that depends on local flow variables. If one assumes that the mean velocity gradient is the only relevant non-scalar parameter for modeling the normalized Reynolds stress tensor, then the strain $\mathbf{S}$ and rotation rate tensor $\mathbf{\Omega}$ can be used to construct a finite set of tensors from which every possible symmetric second order tensor can be created. This is referred to as Pope's integrity basis [67]. Hence, a generalized expression for the anisotropy tensor $b_{ij} = \frac{\overline{u'v'}}{2k} - \frac{1}{3}\delta_{ij}$ for NLEVM reads

$$b_{ij} = \sum_{\lambda=1}^{10} G^{(\lambda)}\left(\eta_i\right)\mathbf{T}^{(\lambda)}\left(\mathbf{S},\mathbf{\Omega}\right), \tag{2.15}$$

where $G^{(\lambda)}$ are arbitrary scalar functions of the invariants η_i. The integrity basis contains ten tensors and five invariants; their exact expression is given in Tables 2.2 and 2.1, respectively. The first tensor is simply the strain rate tensor. There are five invariants which are also functions of the strain and rotation rate tensor.

Table 2.1: Pope's integrity basis: Invariants. The curly bracket denotes the trace operator: $\{X\} = tr(X) = \sum_i X_{ii}$.

Invariants λ_1 to λ_5
$\{\mathbf{S}^2\}$ $\{\mathbf{\Omega}^2\}$ $\{\mathbf{S}^3\}$ $\{\mathbf{\Omega}^2\mathbf{S}\}$ $\{\mathbf{\Omega}^2\mathbf{S}^2\}$

Table 2.2: Pope's integrity basis: Tensors

Tensor number	Tensor expression
$\mathbf{T}^1$	$\mathbf{S}$
$\mathbf{T}^2$	$\mathbf{s\Omega} - \mathbf{\Omega S}$
$\mathbf{T}^3$	$\mathbf{S}^2 - \frac{1}{3}\mathbf{I}\{\mathbf{S}^2\}$
$\mathbf{T}^4$	$\mathbf{\Omega}^2 - \frac{1}{3}\mathbf{I}\{\mathbf{\Omega}^2\}$
$\mathbf{T}^5$	$\mathbf{\Omega S}^2 - \mathbf{S}^2\mathbf{\Omega}$
$\mathbf{T}^6$	$\mathbf{\Omega}^2\mathbf{S} + \mathbf{S}^2\mathbf{\Omega} - \frac{2}{3}\mathbf{I}\{\mathbf{s\Omega}^2\}$
$\mathbf{T}^7$	$\mathbf{\Omega S\Omega}^2 - \mathbf{\Omega}^2\mathbf{S\Omega}$
$\mathbf{T}^8$	$\mathbf{S\Omega S}^2 - \mathbf{S}^2\mathbf{\Omega S}$
$\mathbf{T}^9$	$\mathbf{\Omega}^2\mathbf{S}^2 + \mathbf{S}^2\mathbf{\Omega}^2 - \frac{2}{3}\mathbf{I}\{\mathbf{S}^2\mathbf{\Omega}^2\}$
$\mathbf{T}^{10}$	$\mathbf{\Omega S}^2\mathbf{\Omega}^2 - \mathbf{\Omega}^2\mathbf{S}^2\mathbf{\Omega}$

The scalar functions $G^{(\lambda)}$ still need to be determined. If the expression is substituted in an ASM model, explicit expressions for the scalar functions can be found. Hence, for each implicit ASM, an explicit NLEVM formulation exists. Standalone NLEVMs that are not derived from an ASM exist as well, which can also be expressed through Pope's integrity basis, assuming they rely solely on ∇u.

Now if all terms with $\lambda > 1$ are dropped, yet an even simpler class of models is obtained, **linear eddy viscoity models (LEVM)**, where the anisotropy tensor is assumed to scale linearly with the mean flow stresses as only the strain rate tensor remains. This is the most frequently used type of model. The underlying turbulent-viscosity hypothesis (aka. the Boussinesq hypothesis) constitutes the theoretical basis for these models and is written

$$a_{ij} = 2k \cdot b_{ij} = \overline{u_i' u_j'} - \frac{2}{3} k \delta_{ij} = -2 \cdot \nu_t \overline{S_{ij}}, \tag{2.16}$$

where different expressions exist for the eddy viscosity ν_t. If one looks at the units of the expressions, it has to be the product of a relevant length and velocity scale.

From the simplest to the most complex LEVM there are: (i) **algebraic models**, (ii) **one-equation models**, and (iii) **two-equation models**. As the name implies, for algebraic models, an algebraic expression is used for ν_t. For one-equation models, frequently, a transport equation is solved for k, which is related to the velocity scale, and a fixed expression is used for the length scale. The Spalart-Almaras model in contrast directly solves a transport equation for the eddy viscosity ν_t. For two-equation models, a transport equation for the tke is used, and then a second transport equation is use to model, most frequently, either the dissipation rate ε or the specific dissipation rate $\omega = \varepsilon / k$. Other variables for the second equation have been proposed as well.

The k-ε model is the most widely used model for wind farm physics, and it is used as a baseline model in our work. The transport equation for the turbulent kinetic energy is reasonably similar to the exact equation for tke:

$$\frac{\overline{D}k}{\overline{D}t} = \nabla \cdot \left(\frac{\nu_T}{\sigma_k} \nabla k \right) + \mathscr{P} - \varepsilon \tag{2.17}$$

For the dissipation rate, an exact equation exists. However, in this case it is not very useful for modeling because it is driven by unclosed terms. Instead, a purely empirical transport equation is typically used [49], modeled after the tke equation:

$$\frac{\overline{D}\varepsilon}{\overline{D}t} = \nabla \cdot \left(\frac{\nu_T}{\sigma_\varepsilon} \nabla \varepsilon \right) + C_{\varepsilon 1} \frac{\mathscr{P}\varepsilon}{k} - C_{\varepsilon 2} \frac{\varepsilon^2}{k}, \tag{2.18}$$

where C_μ, $C_{\varepsilon 1}$, $C_{\varepsilon 2}$, σ_k, and σ_ε are constants, that may require flow-specific tuning.

Finally, the eddy viscosity ν_t is given by

$$\nu_t = C_\mu k^2 / \varepsilon. \tag{2.19}$$

Now that several different classes of turbulence models have been introduced, it remains to specify what the specific shortcomings of each of them are. For anything other than simple 2D flows with minimum streamline curvature and adverse pressure gradient, two-equation LEVM will give imperfect results. For most applications, at the very least, tendencies will be predicted correctly. Application-specific tuning may render some improvements, but no general purpose LEVM works well for all complex flows. Nevertheless, LEVMs are used for many industrial applications, and despite their shortcomings, they provide added value over engineering models and measurements. NLEVMs may give better results than LEVMs for flows with some streamline curvature and/or adverse pressure gradients. However, they are more difficult to converge and may not perform well on fine meshes, which negates the purpose of using them. In theory, RSM or elliptic relaxation models should yield more accurate results for flows with strong streamline curvature, adverse pressure gradients, and strong hysteresis effects, but this only applies if the modeling of the pressure-rate-of-strain tensor is appropriate and wall effects are well handled. Furthermore, this does not only come at the cost of computational resources but also at the expense of numerical robustness. These models not only require that more equations per iteration are solved but also that in total more iterations are necessary for convergence because the resulting system of equations is stiff. Finally, even higher-order models rely on an empirical equation for the total dissipation, and this can affect the accuracy for lower and higher order models alike. A more complete list of shortcomings and benefits of specific model classes is available in the literature [68].

So far, only RANS modeling for the Reynolds stresses has been presented. For atmospheric flows with non-neutral stratification, the **turbulent heat flux** also needs to be modeled. This is done using methods similar to those used for the Reynolds stresses. The integrity basis shown above can be modified by adding the temperature gradient, and then models can be derived which are similar to NLVEM models [101]. However, for wind engineering applications, the eddy diffusivity assumption – analogous to the eddy viscosity assumption for the Reynolds stresses – is the most frequently used model for the turbulent heat flux [4]

$$q_j = \overline{u_j'\theta'} = -\frac{\nu_T}{Pr_t}\frac{\partial\overline{\theta}}{\partial x_j}, \tag{2.20}$$

where $Pr_t \approx 1.0$ is the turbulent Prandtl number.

2.1.2. LARGE-EDDY SIMULATIONS

LES methods solve the filtered Navier-Stokes equations. Turbulent scales that are smaller than the filter width are modeled using a subgrid scale (SGS) model, and the rest of the scales are fully resolved. The equations for mass, momentum, and energy conservation are:

$$\frac{\partial\tilde{u}_i}{\partial x_i} = 0, \tag{2.21}$$

$$\frac{\partial\tilde{u}_i}{\partial t} + \tilde{u}_j\frac{\partial\tilde{u}_i}{\partial x_j} = -\frac{1}{\rho_0}\frac{\partial\tilde{p}}{\partial x_i} - \frac{\partial\tau_{ij}^{SGS}}{\partial x_j} + \tilde{f}_i + f_C\varepsilon_{ij3}\tilde{u}_j + g\frac{\tilde{\theta}-\tilde{\theta}_{\text{ref}}}{\rho_0\theta_0}, \tag{2.22}$$

$$\frac{\partial\tilde{\theta}}{\partial t} + \tilde{u}_j\frac{\partial\tilde{\theta}}{\partial x_j} = -\frac{\partial q_j}{\partial x_j}, \tag{2.23}$$

where $\tilde{u}$, $\tilde{p}$, and $\tilde{\theta}$ are the resolved (i.e. filtered) instantaneous velocity field, pressure and potential temperature, respectively, $\tilde{f}_i$ is a volume forcing term for the actuator disk representing a wind turbine, θ_0 and ρ_0 are the reference values for the Boussinesq approximation, and f_C is the Coriolis constant. The SGS part of the stress τ^{SGS} and the SGS heat flux q^{SGS} require further modeling and are meant to capture the effect of the unresolved scales on the resolved ones.

There are many different subgrid scale models. For wind engineering flows, relatively simple models are used and are derived in a similar way as RANS models. The two simplest models use the eddy viscosity and diffusivity assumption for the unresolved fluxes:

$$\tau_{ij}^{SGS} = -\nu_{SFS}\cdot\left(\frac{\partial\tilde{u}_i}{\partial x_j} + \frac{\partial\tilde{u}_j}{\partial x_i}\right) \tag{2.24}$$

$$q_j^{SGS} = -\frac{\nu_{SFS}}{Pr_t}\frac{\partial\tilde{\theta}}{\partial x_j} \tag{2.25}$$

where different approaches exist for determining the SGS viscosity ν_{SGS} and the turbulent Prandtl number Pr_t. The simplest, and also frequently used, approach is the Smagorinsky model

$$\nu_{SGS} = C_s^2\cdot\Delta^2|\tilde{S}| \text{ with } \tilde{S}_{ij} = \frac{1}{2}\left(\frac{\partial\tilde{u}_i}{\partial x_j} + \frac{\partial\tilde{u}_j}{\partial x_i}\right) \tag{2.26}$$

where Δ is the local(ly averaged) filter width and C_s is the Smagorinsky constant. There is a simple formula to calculate the Smagorinsky constant based on the local filter width and Kolmogorov constant. However, from numerical studies, it was found that the best value depends on the flow type [26]. This realization led to the development of the Dynamic Smagorinsky model, where the constant is determined by comparing the local results from two different filter widths [26]. While this is more accurate, calculating

the relevant values for the additional smaller filter width is computationally intensive and may introduce numerical instabilities. Codes where spectral discretization is used and filtering is cheap are better suited to this. The additional accuracy that the Dynamic Smagorinsky model brings is important when strongly stably stratified flows are modeled as the turbulence scales can become very small and it may no longer be possible to reach the required resolution depth. In that case, the importance of a good SGS model increases.

Many more options exist to model the SGS tensor, including those which use a transport equation for the SGS turbulent kinetic energy or those where the turbulent Prandtl number is varied according to a local Richardson number.

In this work, the Wall-Adapting Local Eddy-viscosity (WALE) SGS model is used [17]. Compared to the (dynamic) Smagrinsky model, the WALE model is more accurate and stable near walls. While it still uses the eddy viscosity assumption, the formula for the eddy viscosity is modified using both the resolved strain and rotation rate to read

$$\nu_t = \overline{\rho} L_s^2 \frac{\left(\varsigma_{ij}\varsigma_{ij}\right)^2}{\left(S_{ij}S_{ij}\right)^{\frac{5}{2}} + \left(\varsigma_{ij}\varsigma_{ij}\right)^{\frac{5}{4}}} \tag{2.27}$$

with $\tilde{S}_{ij} = \frac{1}{2}\left(\frac{\partial \tilde{u}_i}{\partial x_j} + \frac{\partial \tilde{u}_j}{\partial x_i}\right) - \frac{1}{3}\delta_{ij}\frac{\partial \tilde{u}_k}{\partial x_k}$ and $\varsigma_{ij} = \frac{1}{2}\left(\frac{\partial \tilde{u}_i}{\partial x_k}\frac{\partial \tilde{u}_k}{\partial x_j} + \frac{\partial \tilde{u}_j}{\partial x_k}\frac{\partial \tilde{u}_k}{\partial x_i}\right) - \frac{1}{3}\delta_{ij}\left(\frac{\partial \tilde{u}_l}{\partial x_k}\frac{\partial \tilde{u}_k}{\partial x_l}\right)$ and where the length scale $L_s = \min\left(\kappa z, C_w V^{\frac{1}{3}}\right)$ in which z is the distance from the wall and C_w is a constant.

This modified formulation considering both strain and rotation rate is beneficial, as the eddy viscosity automatically goes to zero at the wall. Conversely, the Smagorinsky model is non-zero at the wall, but it can be modified by using ad hoc damping such as Van Driest damping.

2.2. WIND TURBINE WAKES AND COMPUTATIONAL MODELING

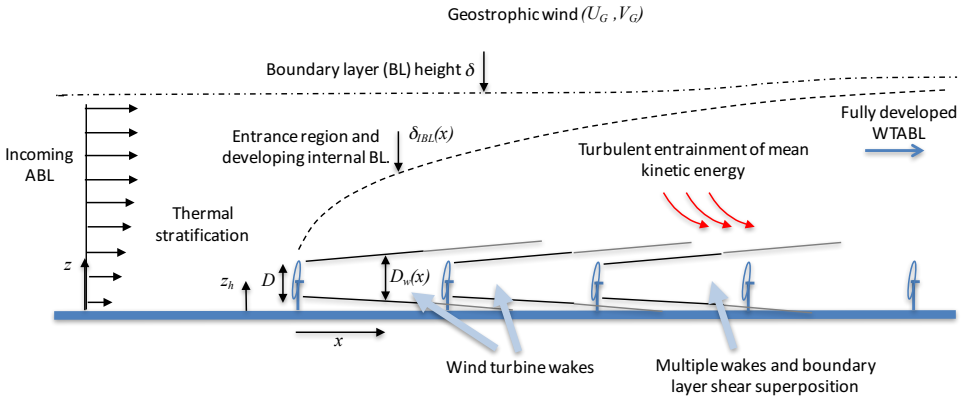


Figure 2.1: Schematic of atmospheric boundary layer with wind turbines, reproduced with permission from [87].

Accurate modeling of wind turbine wakes embedded in the Atmospheric Boundary Layer (ABL) and, in particular, the interaction between the wake of a turbine and the performance of a downstream turbine is relevant for two stages in the life cycle of a wind farm: firstly, during the design stage, when the layout should be chosen according to the wind conditions at the prospective location in terms of wind speed and direction distribution, as well as atmospheric stability and inflow turbulence; and secondly, during the operational period of the farm, when wind farm control strategies seek to find a trade-off between power maximization and load alleviation.

2.2.1. WIND FARM PHYSICS

The wake of a **single free-standing wind turbine** in the ABL can generally be divided into two regions. The near wake is dominated by the tip speed ratio and the shape of the rotor, whereas the far wake is only indirectly affected by the rotor through the velocity deficit and the turbulence intensity. The flow field in the near wake is dominated by the presence of tip and root vortices that lead to strong gradients in the velocity. Due to the momentum and energy extraction at the rotor plane a discontinuity in pressure, an expansion of the wake and a decrease in axial velocity are observed. Generally, the end of the near wake is characterized as the point where the shear layer reaches the wake axis, which is usually about 2 to 5 rotor diameters downstream of the rotor position [16]. In the far wake, the flow is dominated by turbulent mixing, eventually leading to a recovery of the velocity deficit in the wake. The turbulence in the wake has three origins: atmospheric turbulence from terrain and the ABL, mechanic turbulence from the blades and the tower, and wake turbulence from the breakdown of the hub and tip vortices [77]. In existing wind farms, sometimes low-frequency oscillations of the far wake, termed wake meandering, are observed and lead to more variable loading on the downstream turbines. A perturbation analysis showed that the source of the meandering is most likely large-scale atmospheric turbulence in the form of eddies with a length scale roughly equal to the turbine diameter [56].

Once wind turbines are placed closer to each other to form a **wind farm** embedded in the ABL, interaction between the different turbines and the atmospheric boundary layer become the primary influence on the power output and the fatigue loading of the turbines in the farm. Barthelmie et al. [6] analyzed SCADA data from two offshore wind farms in Denmark, Horns Rev and Nysted, for wind directions aligned and $\pm 15^\circ$ from the main axis of the farm. The energy yield of the turbines shows a strong dependence on the wind direction, at least for the first few rows. Independently of wind direction, from the fourth row downwards, the energy output of the turbines drops significantly as compared to the first row. Moreover, in the extreme case where a downstream turbine is directly centered in the wake of a turbine in the first row, a power reduction of up to 40 % for the second row as compared to the first one is observed for a streamwise spacing of around 6 times the turbine diameter with a less steep drop off in energy yield from the third row onwards.

A coupling between the atmospheric flow above the atmospheric boundary layer and the wind farm has even been observed for large wind farms. The main connection between the geostrophic wind and the flow inside the atmospheric boundary layer seems to be the **turbulence-induced vertical flux of mean kinetic energy** given by $\overline{u'w'} \cdot \overline{u}$ [87].

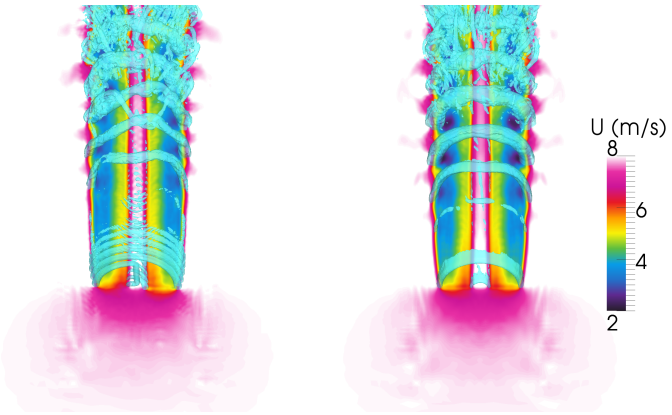
In the asymptotic regime where the farm length exceeds the height of the boundary layer, streamwise changes in the flow field can be neglected, and the kinetic energy transfer between the atmospheric boundary layer and the wind farm occurs in the vertical direction only [13].

2.2.2. MODELING OPTIONS

While there exist a multitude of **engineering models** for wind turbine wakes, generally, better results are obtained with more complete physical models such as **computational fluid dynamics codes** [16, 77, 58, 28, 87]. This holds true in particular when wake interaction is present. Hence, only CFD models are considered here. Due to the large discrepancy between the time and length scales of the eddies in the boundary layer of a wind turbine blade and the ones present in the atmospheric boundary layer, not all relevant scales can be modeled, even with the computational resources available today. Figure 2.1 illustrates the length scales present within the ABL and around the turbine blades. Hence, if the interaction between wakes of different turbines is of interest, the wind turbine blades are indirectly modeled in the CFD code through a simplified rotor model instead of fully resolving the blade boundary layers.

In terms of **turbulence modeling** for the Navier-Stokes equations for wind engineering purposes, both RANS and LES models are used [77, 58, 11]. More recently, also hybrid models have been applied to predicting the wake of a single wind turbine [88]. Currently, LES models that use a scale-dependent dynamic SGS model are considered the state of the art in wind farm wake modeling [58, 11].

The two most prevalent **rotor modeling** approaches for LES are a non-uniformly loaded Actuator Disk (AD) and an Actuator Line (AL) [77]. Both of these models represent the blades through distributed forcing terms in the Navier-Stokes equations. However, for the AD model, azimuthal averaging is applied. Consequently, the AL model can resolve the tip and root vortices of the blades, whereas the AD model can only reproduce the shear layer of the wake. However, changes in the circulation along the blades that are different from rotation effects can be captured by the AD model through shedding in an azimuthally continuous sheet. Figure 2.2 visualizes these differences. Comparison between the results for these two models can be found in [77, 105, 70, 57]. The main conclusion from the comparison studies is that if wake rotation is included in the AD model, on average, the results are almost identical in the far wake but are a bit different in the near wake. As the near wake is mostly not of interest here, an AD model is used for most cases in this work. In terms of computational cost, AL is more expensive because it requires a smaller time step. For unsteady simulations the size of the time step is usually governed by accuracy and numerical stability concerns. The Courant–Friedrichs–Lewy (CFL) condition which relates the grid size, time step and velocity is also frequently used as a measure especially if explicit time discretisation schemes are used. These considerations also apply to the AD model. However, for the AL model, the maximum time step is also limited by the resolution of the blade rotation, which usually leads to a larger restriction on the time step than the other mentioned considerations [57]. The blade rotation is sufficiently resolved when the blade tip does not pass through more than a single finite-volume cell within a single time step. Hence, the maximum time step for AL model simulations is governed by the tip speed ratio of the turbine.



(a) ALM (b) ADM

Figure 2.2: AL (a) and AD (b) model LES simulations where the blue isosurface is of the second invariant of the velocity-gradient tensor. The contours are of streamwise velocity. Reproduced with permission from [57].

2.2.3. PERFORMANCE OF DIFFERENT RANS MODELS

While LES is the current state of the start in high-fidelity aerodynamic wind farm modeling, it remains too computationally expensive to be used for wind farm design or control purposes. Using RANS models for the same purpose is estimated to yield a decrease in computation time by two orders of magnitude [77].

The simplest and most commonly used RANS models for wind farm modeling are **Linear Eddy Viscosity Models**. The flow around a wind turbine is highly three-dimensional, and the rotor leads to large changes in the mean strain rate over a short distance. This means the flow is anisotropic and locally not in equilibrium, and these are both conditions that differ vastly from the ones LEVM are derived for, as shown in Section 2.1. Further, the turbulence in the atmospheric boundary layer itself is anisotropic. Things like complex terrain and large surface roughness may further complicate the situation. The most recent and popular results of a literature review for different RANS models in combination with an actuator disc rotor model are listed below. All of these models are derived for neutral atmospheric conditions and flat terrain. Stratification will be considered in Section 2.2.4.

- **Standard k-ε [49] and k-ω models [102]:** These two models underpredict the wake deficit and the turbulence intensity (peaks) because the eddy viscosity is overpredicted and the wake becomes too diffusive [77, 71, 46, 5]. A detailed comparison between LES and RANS results with the k-ε model shows that this particular LEVM RANS model cannot quantitatively predict the correct turbulent kinetic energy and its dissipation [72]. In fact, in the near wake region, the realizability constraints are significantly violated.

- **Modified k-ε and k-ω models:** Based on the results obtained with the standard, simple LEVM models, various modifications were applied most of which aim to

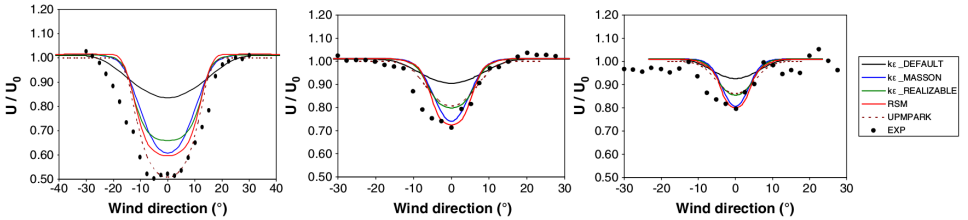
reduce the eddy viscosity in the near wake. El Kasmi and Masson [40] used a modified version of the k-ε model, which introduces a source term in both the turbulent kinetic energy as well as the dissipation transport equation. They argue that the source term in the dissipation rate equation suppresses the overproduction of turbulent kinetic energy in the near wake, where strong shear gradients are present. Prospathopoulos et al. [71] apply an eddy viscosity limiter (Durbin limiter) based on a realizability constraint. Réthoré [72] used two different eddy viscosity limiters based on a realizability constraint and based on the adverse pressure gradient in the near wake region. Van der Laan et al. [46] developed a model named the k-ε-f_P model with a limiter that reduces the eddy viscosity in regions with high-velocity gradients. The limiter is a simplified version of a cubic nonlinear EVM and is applied directly in the relation for the eddy viscosity. In a follow-up publication, van der Laan et al. [45] compare this eddy viscosity limiter to the ones from Shih [80] and Durbin [20], all for the k-ε model. They recommend using either the f_P or the Shih limiter since the Durbin limiter [71] is very sensitive to the ambient turbulence levels. While all of these models offer some improvements over the standard k-ε model, the improvements are test-case dependent, and some of them require tuning parameters. The k-ε-f_P model seems like the most promising one as it consistently increases the prediction of the velocity field. However, the improvement for the tke is less consistent, and it is still an isotropic model.

- **Standard [59] and modified k-ω SST [81] models:** Because the k-ω SST model from Menter is very popular for wall-bounded flow with adverse pressure gradients and already includes an eddy viscosity limiter, Shives and Crawford [81] compared it to the k-ε model with and without tuning of the $C_{\varepsilon 4}$ parameter. Overall, the k-ω SST model yields better predictions of the velocity deficit in the near wake, whereas the k-ε model provides better predictions of the velocity deficit in the far wake. Both models are inadequate at predicting the turbulence intensity in the wake. Both Shives and Crawford [81], as well as Réthoré et al. [73], argued that the underprediction of the turbulence intensity in the near wake originates from the shortcomings of the actuator disk model. Hence, Shives and Crawford introduced a scaling term for the turbulent kinetic energy production in a circular tube that roughly covers the near wake region of the turbine. They calibrated the correction term with existing wake measurement data. The now corrected k-ω SST - referred to as the k-ω SST $+S_k$ model - vastly improved the prediction of the turbulence intensity and to some extent also the velocity deficit in the wake.

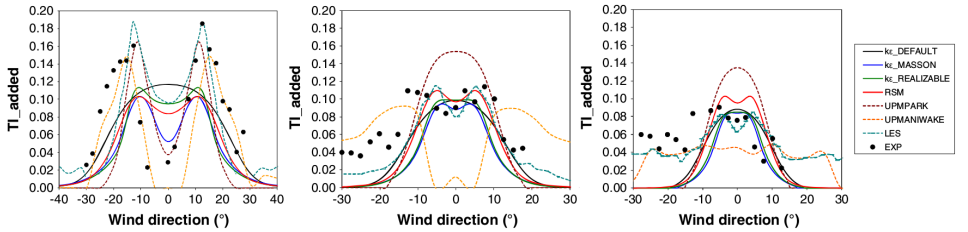
- **RSM models:** Cabezon et al. [12] tested a Reynolds Stress Model (RSM) for a single wake analysis and then compared the results to different two-equation RANS models, parabolic models, LES and experimental data. The more sophisticated turbulence model improved the prediction of the velocity deficit, in particular, in the near wake as compared to the standard and corrected k-ε model. Moreover, more physical turbulence intensity predictions were obtained. However, the agreement with LES data was not consistently improved since sometimes also some of the more simple RANS models gave better results. Comparison between the different models in terms of velocity deficit, turbulence intensity and stream-

wise shear stress are shown in Figure 2.3.

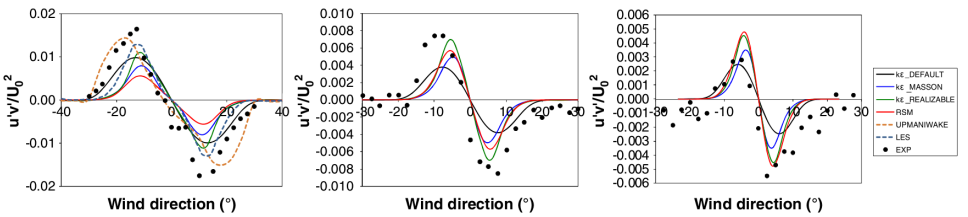
- **Parabolic codes:** This type of model neglects both the diffusion and pressure gradient in streamwise direction to reduce computational cost [77]. Cabezon et al. [12] showed that parabolic codes can yield surprisingly good results for the near wake below 3 diameters downstream of the rotor, because they inherently lack diffusive terms as compared to elliptic models which tend to be too diffusive in the near wake region. However, beyond the near wake region, this type of model leads to much less accurate results than elliptic models.



(a) Velocity deficit



(b) Turbulence intensity



(c) Reynolds stress

Figure 2.3: Wake deficit, turbulence intensity and plane Reynolds stress for different inflow directions, reproduced with permission from [12]. From left to right the plots are stream-wise slices at 2.5, 5.5, and 8 diameters downstream of the rotor. UPMPMARK and UPMANIWAKE are parabolic codes, the rest of the labels should be self-explanatory.

Antonini et al. [5] argue that the inaccuracy of the k-ω SST and the RSM model arise not only from modeling limitations but also from uncertainty in the inflow wind direction. They derived a method to consider **uncertainty in the inflow direction** by combining steady-state RANS simulations for different inflow angles. This yielded a more

favorable comparison to measurement results regarding the energy yield of turbines embedded in the Horn Rev wind farm.

Summarizing, **standard linear eddy viscosity turbulence models in combination with an actuator disk model suffer from structural shortcomings** when modeling wind turbine wakes and their interactions. For this application, the most frequently used RANS turbulence model to date is the k-ε model. However, in its standard form, this model results in an underprediction of the velocity deficit and an overprediction of the turbulence intensity in the wake. Various modifications have been proposed, most of which aim at reducing the eddy viscosity in the near wake region. The most successful ones rely on an eddy-viscosity limiter in areas of high shear. While these corrected models have shown some improvements over the original model, so far, no model has shown consistent improvement in both the prediction of the velocity field and the turbulence anisotropy in the near- and far-wake. The shortcomings of these models in predicting the wake of a free-standing turbine will no doubt be further propagated once interacting turbines are investigated. While some improvements over these models could be obtained using a Reynolds Stress Model, still a large deviation between averaged LES and RANS results was observed [12]. Moreover, a Reynolds Stress Model requires solving five additional transport equations compared to the standard two-equation linear eddy viscosity models.

It should be mentioned that unsatisfactory results in predicting the mean flow features in wind farms are most likely not solely due to the structural limitations of LEVMs, but possibly also due to limitations in the rotor modeling. Current actuator disk models only indirectly account for the effect of the rotor on the turbulent structures, which can lead to large errors in the flow properties close to the rotor plane. This will be explained in detail in Section 2.3.

2.2.4. ADDING STRATIFICATION TO RANS MODELS

So far, none of the aforementioned publications have considered **atmospheric stability**. In literature, there are only a few examples of RANS simulations for wind energy where stratification is modeled [8, 92, 97, 48]. However, previous work exists in the field of wind engineering and industrial aerodynamics [62]. Under unstable atmospheric conditions, the air is warmer close to the ground than at higher altitudes. Under stable conditions, the situation is reversed. The resultant buoyancy forces influence the undisturbed velocity and tke profiles. For stable stratification, the turbulence intensity is lower and the shear in the velocity profile over rotor area is larger as compared to neutral stratification. For unstable stratification, again the situation is reversed. Higher turbulence intensity leads to faster wake recovery, whereas a larger shear in the velocity profiles leads to more fatigue loading on the rotor. As mentioned in the previous subsection, in Eq. (2.2), the effect of local temperature gradients on the mean flow is modeled through the Boussinesq approximation, $(\rho - \rho_0)/\rho_0 << 1$. Due to this linearization, the incompressible NS formulation can be used by adding one correction term. Additionally, a transport equation for potential temperature (energy conservation) needs to be solved, and a model for the turbulent heat flux is also necessary. The effect of buoyancy on the turbulent properties can be modeled by adding a turbulence production term in the transport equations for the turbulence model. The formulation of these source terms for the k-ε model were

introduced in the previous subsection in (2.18).

Assuming that the stratification mainly affects the undisturbed velocity and tke profiles, and only indirectly interacts with the turbine wake, simpler models that do not require the addition of the energy equation have been found in the literature. Laan et al. [48] have derived correction terms for the k-ε turbulence equation based on the Monin Obukhov Similarity Theory (MOST). The correction to the turbulence equations is chosen such that inflow profiles chosen according to MOST will not further develop even in a large domain. MOST theory assumes that the flow is in equilibrium and only changes in the vertical direction such that the only input parameter relating to the stratification is the Monin-Obukhov length

$$L = -\frac{u_*^3 \theta_0}{\kappa g \overline{\theta' w'}}.$$

This approach works well for weakly stratified flows, but it has not been tested sufficiently for strongly stratified flows yet. Later, Baungaard et al. [7] further simplified the model and simulated a single wind turbine wake under stratified conditions. Additionally, they combined the k-ε-f_P model with the MOST k-ε model from Laan et al. This required the introduction of an additional parameter to the limiter formulation for f_P as the original model is calibrated for neutral stratification. This gave improved results compared to the k-ε baseline model, but again tke and turbulence anisotropy predictions can be further improved. The underlying assumption for these simplified stratification models is that the undisturbed flow is in equilibrium, i.e. it will not further develop from the inflow. This is of course nonphysical as the heat flux from the ground into the domain will continuously modify the energy in the domain, and thus, also the flow. However, the important thing is the relevant time and length scales of this process. Quite possibly, the undisturbed profiles will not change much over the length of a wind farm, even if the unsteadiness due to buoyancy is modeled. Hence, if one is interested only in a snapshot analysis of the wind farm for a specific wind profile, this is a valid approach. However, if one wants to observe a wind farm over several hours, with stratification changing due to night/day time, then an equilibrium approach is not valid.

2.2.5. DATA-DRIVEN MODELS FOR WIND ENERGY

A handful of **data-driven RANS models have been derived specifically for wind farm wake modeling** to counter these shortcomings [34, 41, 2]. Iungo et al. [34] employed LES results for a single free-standing turbine to find an optimal mixing length distribution for a mixing length LEVM model in combination with an actuator disk turbine model. Checking the magnitude of the different strain rate-dependent terms contributing to the eddy viscosity revealed that the radial gradient of the streamwise velocity $\frac{\partial U_x}{\partial r}$ component is the most important term. Comparison with the corresponding Reynolds stress term $\overline{u'_x u'_r}$ gave way to find an optimal mixing length distribution using

$$-\overline{u'_x u'_r} = l_m^2 \frac{\partial U_x}{\partial r}^2.$$

They assumed that the resulting mixing length distribution is mainly sensitive to the streamwise position. The best fit to the LES Reynolds stress and mean velocity revealed

that up to one rotor diameter downstream of the rotor, it remains constant and beyond that distance starts to increase linearly. The slope of this increase is related to the tip speed ratio of the turbine. Assuming measurements of both streamwise and radial velocity components are available at a location downstream of a rotor, this is a simple way to correct some of the shortcomings of LEVM models for predicting the wake development of a wind turbine.

A similar but more sophisticated approach was taken by Adcock and King [2] who used Lidar data for different atmospheric stability conditions with the aim to better capture the effects of atmospheric stability using a RANS turbulence model. However, the flow field was only a horizontal slice of the flow field at hub height. The turbine was modeled using an actuator disk. An adjoint-based optimization was used to find an optimized mixing length distribution for stable, unstable, and neutral stratification. The measured turbulence intensity was used to assign the measurements to different atmospheric stability classes. The optimization objective was formulated to minimize both the deficit between the mean velocity and thrust coefficient of the turbine as compared to measurements. The results for a single turbine are shown in Figure 2.4. Subsequently, a rotor-centric correction term was calibrated using the linear combination of three Gaussian distributions. Vast improvements as compared to the uncorrected simulations were obtained even for wind farm layouts that the model was not calibrated on.

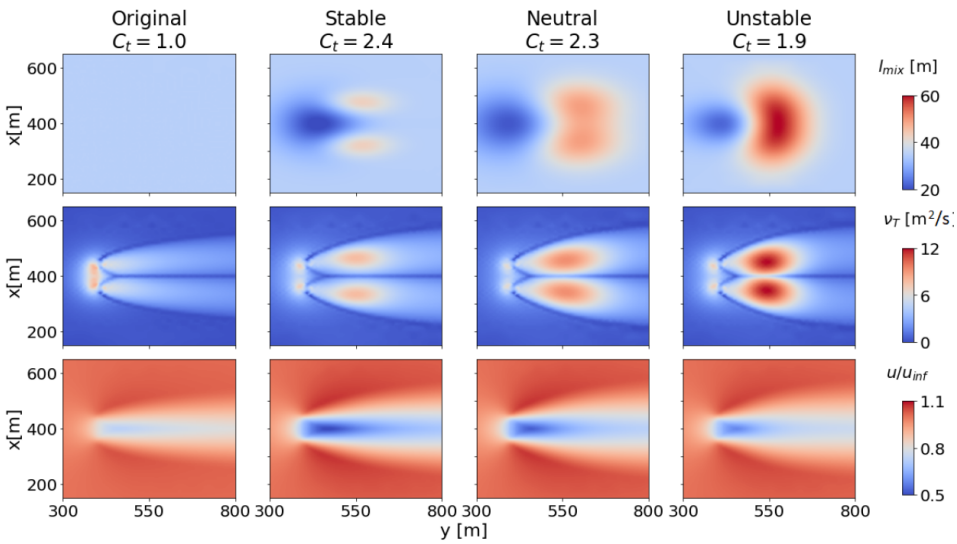


Figure 2.4: Original and optimized mixing length field, as well as, the resulting eddy viscosity and velocity field. Reproduced with permission from [2].

Later, King et al. [41] extended the previously mentioned framework to 3D. Again an adjoint-based optimization was carried out to find an optimized eddy viscosity distribution for a RANS simulation given averaged LES reference data. Subsequently, this optimized eddy viscosity field was sampled, and Gaussian Process Regression was employed to construct a data-driven turbulence model, including uncertainty of the predictions.

The inputs to the machine-learning algorithm were based on velocity, pressure, and spatial gradients thereof. Figure 2.5 shows the results of an entire wind plant simulation unrelated to the training set. The flow field is improved compared to the original mixing length model and produces structured changes in the wake shear layers.

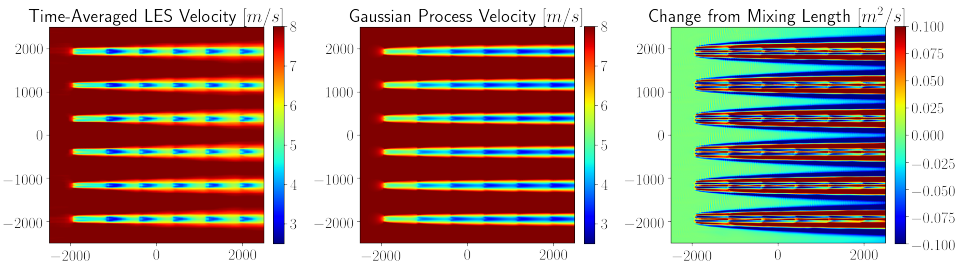


Figure 2.5: Plant flow field prediction from LES, the calibrated Gaussian Process model and the original mixing length model. Reproduced with permission from [41].

More recently, Eidi et al. [24] have developed a combined approach where machine learning is used to estimate uncertainty bounds for velocity and tke in the wake of multiple interacting turbines. The approach is based on perturbation of the realizable k-ε base model towards limiting states on the barycentric map. The amount of perturbation toward one/two/three-component turbulence is determined by a reference LES simulation; perturbation towards two-component turbulence is omitted because it is mostly zero compared to LES. In the next step, a gradient-boosted decision-tree technique (XGBoost) is used to learn the perturbations using an extended integrity basis that also includes pressure and tke gradients fields as input. A mutual information script reduces the feature input set from 54 to 13. The generated models also generalized well between training and test datasets which all included multiple interacting turbine wakes. Figure 2.6 shows the LES, the baseline RANS, the perturbed RANS, and the learned perturbed RANS simulation results.

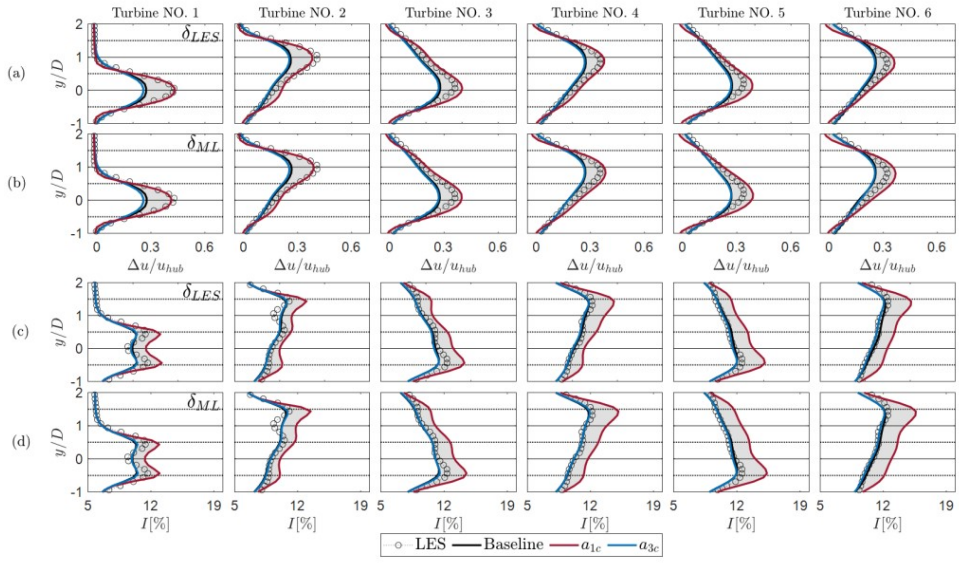


Figure 2.6: Lateral profiles of the normalized velocity deficit (a,b) and turbulence intensity (c,d) at 5D downstream of each turbine at the hub height in Case C. Here, for the perturbed RANS (toward one- and three-component turbulence), δ calculated based on LES data (a,c) and δ predicted by the ML model (b,d) are utilized. Picture reproduced with permission from Eidi et al. [24].

2.3. SPECIFIC SHORTCOMINGS OF $k-\varepsilon$ MODEL FOR WIND ENERGY

The previous subsection presented a general analysis of the shortcomings of RANS turbulence models for wind farm modeling. In this section, we go more in detail on the specifics of these shortcomings for the most frequently used k-ε model.

In his dissertation, Réthoré made a detailed comparison between an LES simulation using the Smagorinsky model and a RANS simulation using the k-ε model for a single wind turbine under neutral conditions [72]. This setup allows the evaluation of specific shortcomings of the RANS model as compared to LES. There are two primary error sources for the RANS model: (i) the eddy viscosity assumption, and (ii) the simplified turbulence transport equations themselves. The validity of the eddy viscosity assumption, i.e.

$$\tau_{ij} = \frac{2}{3}k\delta_{ij} - 2\nu_t \overline{S}_{ij} \text{ with } \nu_t = C_\mu \frac{k^2}{\varepsilon} \tag{2.28}$$

can be evaluated in terms of the magnitude (i.e. eddy viscosity ν_t) or the normalized anisotropy tensor $b_{ij} = \tau_{ij}/2k$. Réthoré carried out both of these analyses and found that the eddy viscosity assumption was violated in particular in the near wake, both in magnitude and directionality.

Further, realizability constraints are also violated in the near wake region. Pope [68] reformulates this constraint for simple shear flow as follows

$$C_\mu \leq \frac{4/9}{\mathscr{P}/\varepsilon}. \tag{2.29}$$

For the k-ε model, C_μ is a constant whose default values vary between $0.03 < C_\mu < 0.09$ and in the near wake of a wind turbine $\mathscr{P}/\varepsilon >> 1$ as the flow is not in equilibrium. If the ratio of turbulence production and dissipation is also only $\mathscr{P}/\varepsilon \approx 20$, the constraint will not hold anymore. Thus, realizability constraints do not hold for a constant C_μ.

Additionally, in the derivation of the simplified transport equations for the tke and the dissipation from the exact equations, two terms were left out in both the tke and dissipation equation

$$A_k = \overline{u_i' F_i} - \overline{u_i' \frac{\partial p}{\partial x_i}} \tag{2.30}$$

$$A_\epsilon = 2\nu \overline{\frac{\partial F_i}{\partial x_j} \frac{\partial u_i'}{\partial x_j}} - 2\nu \overline{\frac{\partial^2 p}{\partial x_i \partial x_j} \frac{\partial u_i'}{\partial x_j}}, \tag{2.31}$$

compare to Eqs. (2.17) and (2.18) of the k-ε model.

The first terms in Eqs. (2.30) and (2.31) capture the effect of the volume forcing F on the turbulence, and this is left out in the derivation of the turbulence transport equations because for most flows volume forcing is not applied. The second term refers to the pressure-velocity coupling, which is assumed to be negligible. However, this does not hold in the presence of an actuator disk, which introduces a strong adverse pressure gradient at the rotor disk, and in the near wake. Hence, the terms that are neglected in the derivation of the RANS transport equations describe the effect of the actuator forcing on the turbulence. In fact, from measurement and LES simulations, it is clear that the turbine can extract energy from the turbulent fluctuations [94]. Since a RANS model cannot capture this, tke is overpredicted in the near wake and the wake recovers too fast. Compared with the LES simulations, Réthoré was able to show that the A_k term is up to two orders of magnitude larger than the local dissipation, and the A_ε term was much smaller and is negligible.

2.4. DATA-DRIVEN TURBULENCE MODELING

Given the complexity of turbulent flows and the shortcomings of the currently available RANS turbulence models, not only for wind farm applications but also for more general applications, more recently, **data-driven models** have been used to formulate closure models based on calibration data from higher fidelity models such as DNS and LES [44, 18]. Ideally, these data-driven models exceed the accuracy of classical RANS closure models in the domain in which they are trained, approaching the accuracy of LES or DNS at only a fraction of the cost. Data-driven modeling uses existing baseline LEVM models and modifies them to predict the same Reynolds stresses as the high-fidelity data on which they are trained. The corrections to the baseline model are usually calculated from local parameters that are used as inputs to train the data-driven model.

While classical RANS closure models are also calibrated using either numerical or experimental reference data, this is only done for special flow scenarios, such as for ex-

ample isotropic turbulence, and is restricted to constant tuning parameters only. Recently, Menter et al. [60] have developed the Generalized k-ε model (GEKO), which is a consolidation of different two-equation models into one model. Depending on the application, the user has the ability to tune the model using six coefficients relating to separation, non-equilibrium flow at walls, mixing strength of free shear layers, free jet flows, curvature correction, and secondary corner flows. The coefficients are designed such that validity of the models for the flat plate and simple mixing layer calibration cases still holds. This goes in the same direction as data-driven models but still falls back on existing models, whose flaws have been described previously in this chapter.

However, in practice, data-driven turbulence models are a relatively new field of research, and their limitations are not well defined yet [44, 18]. Intrinsically such models depend on training data which may only be available for a very limited application range. For example, DNS data are only available at low Reynolds numbers. How far extrapolation outside of the training range still yields improved results is unclear. Additionally, as compared to traditional turbulence models, it may be more difficult to physically interpret what the model does, especially if black box models are used. In addition to this, there remain challenges concerning implementation, as outlined in the following along with some mitigation strategies.

First, given ground-truth data from a higher fidelity model such as DNS or LES, the "accurate" Reynolds stresses are a second-order statistic whereas the mean velocity field is a first-order statistic. Consequently, the mean velocity field is more converged than the mean Reynolds stresses and errors in the prediction of the latter will be propagated to the data-driven model [89, 69]. These publications showed that when computing RANS channel flow solutions using DNS Reynolds stresses with relatively small errors, significant deviations occurred between the resulting RANS mean velocity profile and the DNS mean velocity profile. Repeating this procedure for available DNS data for channel flow at different friction Reynolds numbers showed that the error amplification becomes more significant as the Reynolds number increases [69]. Wu et al. [103] tried to explain this observation by formulating a local, mesh-independent condition number for the RANS equations to quantify the sensitivity of the solved mean velocity at a given location for a perturbation of the Reynolds stress field. For channel flow, they showed an increase of this local condition number with the friction Reynolds number. Further, they show that an implicit treatment of the linear part of the Reynolds stresses leads to a reduction of the condition number by at least an order of magnitude as compared to the explicit treatment of the linear Reynolds stresses. Thus, according to this limited evidence, **consistency between different turbulence modeling approaches** such as RANS, DNS, and LES is not a given.

Second, the mapping between the local mean flow quantities and local mean Reynolds stresses is not unique due to **non-local and non-equilibrium physics**. One of the assumptions behind linear and nonlinear eddy-viscosity models is that the mean Reynolds stresses depend only on the local mean strain and rotation tensor. This implies equilibrium turbulence meaning the turbulence production balances the dissipation everywhere in the flow field [68]. Of course, this does not hold true in general. Through machine learning, to some extent, this can be modeled by including the kinetic energy gradient and the pressure field in the model input [38, 104]. A more complete picture of the

influence of the flow history can be obtained by solving transport equations for the flow anisotropy or the turbulent kinetic energy [30, 21]. However, suppose such a transport equation with, for example, one scalar field parameter of interest is to be included in a machine learning model. In that case, an inversion problem needs to be solved to obtain an optimal distribution of this parameter. One example of this is the local scaling of the turbulence production term in the Spallart-Allmaras turbulence model to better predict the stall onset on wind turbine airfoils [83].

Third, **physical constraints** such as Galilean and rotational invariance, as well as, realizability are not enforced by default by machine-learning algorithms, neither in their inputs nor outputs. Hence, the input to an algorithm needs to be formulated such that physical constraints are automatically fulfilled. Alternatively, if this is not possible, penalizing non-physical terms in the objective function is also an option. Galilean invariance can be ensured through the smart selection of the form of the input variables. Given a local tensor, such as the spatial gradient of the velocity field, or a vector, such as the spatial gradient of the pressure, the rotational invariance of any tensor constructed as a linear combination of this basis can be enforced through the use of a minimal integrity basis [68]. Such a tensor basis has already been successfully employed in machine learning for the mean strain and rotation rate, as well as the mean gradient of the pressure and the kinetic energy of a given flow field [53, 95, 38, 104]. Realizability is more difficult to parameterize and at this point the author is not aware of any frameworks where this is automatically fulfilled.

Finally, the **resulting Reynolds stress field needs to be smooth**. Since $\nabla \cdot \tau$ appears in the RANS equations, to ensure that the robust convergence of the solver is not lost, the addition of a data-driven turbulence model should be continuous. So far, two techniques have been used to stabilize machine-learning-enhanced turbulence models. Kaandorp and Dwight [38] employed under-relaxation of the anisotropic tensor against the prediction from the LEVM model and subsequent substitution of this term into the turbulent kinetic energy transport equation. Wu et al. [104] use the already mentioned implicit treatment of the linear Reynolds stress components through an optimized eddy viscosity.

This gave an overview of the requirements that data-driven models have to fulfill. Now, **an overview of the most frequently explored approaches is given**.

Initial approaches to data-driven turbulence modeling aimed at explicit substitution of terms such as the Reynolds stresses and turbulence production through a data-driven model [53, 18]. However, due to the previously mentioned ill-conditioning of the resulting system of equations, Duraisamy et al. [18] suggest that learning discrepancy functions between existing physics-based models and reference data may be the more successful approach to incorporate machine learning into turbulence modeling. However, depending on the methodology the calculation of those discrepancy terms may not be as straightforward as the explicit substitution of fields obtained from a higher fidelity model.

A critical aspect of data-driven modeling is the location in the governing equations at which the baseline model is modified. Simply scaling terms in the transport equation of the baseline model can be effective for some applications [90, 18, 33]. Alternatively, introducing correction terms in the formulation of the Reynolds stresses, with and with-

out additional source terms in the turbulence transport equations, can capture more complex phenomena [79, 53, 39, 99, 35, 106]. Variations of these approaches exist.

Another important aspect is how the correction fields are derived. Field inversion approaches are a possibility [33, 18], but they require an adjoint formulation of the problem. Implementing an adjoint approach can be labour intensive and is not always numerically stable. An alternative is to use a frozen approach, which does not need an adjoint because it solves for the correction terms based on the frozen mean flow and turbulence properties by algebraically manipulating the turbulence transport equations and the definition of the turbulence quantities [79, 99, 106]. However, a frozen approach needs full-field data. Both of these approaches are suitable for offline learning. Alternatively, if online learning is possible, then cost functions that are evaluated at every step and penalize deviation from the high-fidelity reference data may also be considered [110]. This does not necessarily require the availability of gradients.

Once it is clear where the corrections should be applied and how they will be calculated, it remains to determine how they will be learned. Two broad model categories can be distinguished based on the underlying regression model: (i) blackbox models such as neural networks and random forests [39, 90, 66, 82, 83, 53, 106, 33], and (ii) explicit models built using symbolic algorithms such as sparse regression and Gene Expression Programming (GEP) [79, 99, 100, 35, 109, 110].

Independent of the choice of algorithm is the need for an objective function. For scalar, frame-independent fields, an appropriate formulation will be obvious. However, for frame-dependent fields such as the anisotropy components, evaluating the objective function in terms of deviation from the reference anisotropy tensor eigenvalues and eigenvectors may be worth considering. A few publications use a frame-independent objective function for the anisotropy correction in the form of deviation of the eigenvalues and eigenvectors of the Reynolds stress tensor [95, 104, 96].

Lastly, while most authors use deterministic models, some authors have also used uncertainty quantification to obtain bounds on predictions. Bayesian inference can be used to obtain bounds on the coefficients of turbulence models. When adding Bayesian Model Averaging, predictive models, including uncertainty quantification, can be obtained [22, 23, 21]. The approach from Eidi et al. [24] that was mentioned in Section 2.2 does something similar by perturbing the eigenvalues of the Reynolds stress tensor towards limiting turbulence states.

Finally, a brief introduction to **classifiers** in traditional and data-driven turbulence modeling is presented, as this could potentially help develop data-driven models. A classifier is a function that yields values between zero and one, and could hence be used to switch closure corrections off and on, so they are active where needed, whilst elsewhere the unmodified base model is used. This is analogous to sensors in traditional closures, which detect specific physical effects and active, relevant terms locally. Classifiers also have the potential to reduce the training cost for the algorithms as corrections only need to be trained in specific areas. Of course, this only works if the training cost of the classifier is manageable. In the literature, there are only a handful of examples of classifiers used within the context of data-driven turbulence modeling. Gorlé et al. [29] developed a simple nonlinear marker for RANS simulations to identify regions where the flow field deviates from parallel shear flow. Ling et al. [54] defined three separate markers

that pertain to different ways in which the Boussinesq hypothesis fails: (i) the negativity of the eddy viscosity, (ii) turbulence anisotropy, and (iii) the difference between a linear and nonlinear eddy-viscosity model prediction. The markers were derived by solving a classification problem using different supervised machine-learning approaches: Support Vector Machines (SVMs), Adaboost decision trees, and Random Forests (RFs). Longo et al. [55] used the marker from Gorlé et al. and blended a LEVM model with an NLEVM in regions of non-parallel shear flow around buildings. A blending function was used to smooth the marker properties further, because the marker itself can have very sharp gradients.

To finish this chapter, **some concrete examples of data-driven turbulence** are given below.

In one of the most well-known publications on data-driven turbulence modeling, Ling et al. [53] developed tensor basis neural networks to predict improved anisotropy tensors for standard low Reynolds benchmark cases. These neural networks were specifically developed to be used with Pope's integrity basis. This is important because this allows learned quantities such as an anisotropy tensor to be rotationally invariant. The models were trained to learn not a correction to the anisotropy tensor but the full anisotropy tensor. Their training database consisted mainly of DNS data available from literature such as duct flow, channel flow, perpendicular jet in cross-flow, inclined jet in cross-flow, flow around a square cylinder, and flow through a converging-diverging channel. The test cases were duct flow at Re = 2000, and flow over a wavy wall at Re = 6850. The results were compared to a simple LEVM and a cubic NLEVM, and the learned models outperformed both baseline models. They were able to resolve secondary flow structures, such as corner vortices for the duct flow and flow separation for the wavy wall case, but still, they did not fully match with DNS, and there is room for improvement. The authors also mention that predicting the anisotropy tensor perfectly does not guarantee a matching velocity field. Hence they suggest including an inference step in their methodology to further improve their results.

In 2020, Zhao et al. [110] use Gene-Expression Progamming (GEP) to improve the predictions of the k-ω SST model with γ-Re_Θ transition model for two-dimensional simulation of turbomachinery blades. The Reynolds number of the cases is around $\mathrm{Re} \approx 500,000$. The correction to the baseline model is introduced as a simple anisotropy correction. The GEP algorithms input list consists of the first three tensors and the first two invariants of Pope's integrity basis. In this publication, the learning is done online, unlike earlier versions of the authors, where they used a frozen approach [99]. The cost function for the learning penalizes deviations from the reference anisotropy tensor obtained from LES simulations and is evaluated for every model of every generation of the genetic algorithm. The resulting models are significantly simpler than the ones obtained from the frozen procedure, and this makes them more numerically robust. Compared to the baseline model and reference LES data, marked improvement is observed on both training and test cases. The authors conclude that the framework works well for deriving simple, accurate, and robust corrections to the baseline model, the only drawback being the prohibitive cost of the online training routine.

The last example from Edeling et al. [23] used Bayesian inference to assess the parameter uncertainty of five frequently used LEVM models on flat plate boundary layer

flows at varying pressure gradients. The training dataset consisted of experimental data. The Markov-Chain Monte-Carlo method was used to obtain probability distributions. A large variance in the coefficients for all the LEVM models was observed. To obtain bounded predictions, Bayesian model-scenario averaging was applied, but the obtained bounds were very large and hence not useful. To remedy this, the authors also developed a smart sensor that weights the different scenarios in the Bayesian model averaging based on similarity to the test case at hand. This yielded more useful bounds on the predictions. The authors propose to extend their methodology to more complicated test cases.

2.5. CONCLUSION

The purpose of this Section is to put the previous Sections into context and motivate the methodology presented in Chapter 3.

In Section 2.1 a brief introduction to different turbulence modeling approaches was given and it was reasoned that, in general, increased accuracy comes at the cost of computational time and possibly also numerical robustness. Then, in Section 2.2, the state of the art in CFD for wind energy was presented: LES simulations are the state-of-the-art, but RANS simulations are most frequently employed as LES is too computationally expensive for most users. Additionally, different RANS models have been trialed in literature for modeling wind turbine wakes, but there is no model that gives consistent improvement while being numerically robust. In Section 2.3, a study in the literature highlighted the specific shortcomings of the k-ε model as compared to and LES model applied to a wind turbine wake: (i) the eddy viscosity assumption is violated in both magnitude and directionality especially in the near wake, (ii) the Reynolds stresses are unrealizable in areas of high shear, (iii) terms relating to the pressure gradient and volume forcing are dropped in the derivation of the transport equation for the tke and the dissipation and are introducing errors at the actuator disk. Finally, in Section 2.4, the field of data-driven turbulence modeling is introduced highlighting both the difficulties and the potential of this young research field.

In summary, for wind turbine wakes, it would be beneficial to develop RANS models which are more accurate, numerically robust and which address the structural shortcomings of the existing models. So far, traditional approaches to turbulence modeling have failed to produce such models and it makes sense to investigate the potential of data-driven approaches.

Then, the review showed that there are different classes of RANS turbulence models. The most accurate ones being RSM models, then NLEVM and finally LEVM. The usage of an RSM model is appropriate if there are either strong curvature effects, strong hysteresis effects or strong adverse pressure gradients at work. On the other hand, LEVM are indicated if none of these aspects apply. Wind turbine wakes with exception of the actuator disk do not display any of these effects very strongly, but they are present. In standard RSM models the effect of the actuator disk on the turbulence is not modeled, hence they do not present an advantage over NLEVM in this aspect. The author thus suggests investigating a data-driven approach based on a NLEVM model that can better model anisotropy and the interaction between the actuator disk and the turbulence model. Such an approach will be presented in the next chapter.

3

METHODOLOGY[1]

This chapter gives an overview of the methodology that underlies the work presented in this thesis. The complete data-driven turbulence modeling chain consists of three main steps. First, we define a set of cases and perform LES simulations of those cases, to provide training and validation data. This data serves as a target and ground-truth of subsequent RANS modeling efforts, as documented in Section 3.1. Second, we solve for RANS corrective fields, as explained in Section 3.2. The frozen approach used here delivers corrective fields which, when injected in the RANS simulations of the training-cases, reproduce very accurately the LES mean-fields. Note that it is not sufficient to merely use the LES-obtained Reynolds-stress tensor (RST) to correct the momentum equation, as established by Thompson et al. [89], as this does not necessarily lead to the correct mean flow. Rather, our procedure serves the same purpose as *field inversion* in the work of Parish & Duraisamy [66], but does not require an adjoint or an optimization. Section 3.3 presents the requirements for the training data when a classifier is used. A classifier can turn models "on" and "off" by weighting the corrections in the turbulence model, such that they are only active in flow regions where the corrections obtained from the frozen method are large. Finally, in Section 3.4, we use sparse symbolic regression to discover a concise algebraic expression approximating these corrective fields, using only local flow quantities available in the RANS simulations. The result is a new turbulence closure model, customized to the training cases, which can be used to make predictions for similar setups outside the training set. The author will refer to this methodology, without classifier, as the base methodology. The base methodology is used in the results section, but additionally a second slightly modified methodology was tested as well to deal with some shortcomings of the base methodology. On the cases tested in this work, the base methodology leads to rather complex corrections. Hence, in an updated methodology, a classifier is used and obtained through logistic sparse regression, analogously to the turbulence correction terms. This is presented in Section 3.4 as well.

[1] The writing presented in this chapter is heavily based on the two most recent publications of the author, they are listed in B.2

3.1. CASE DEFINITION AND LES DATABASE GENERATION

The first step in the proposed methodology is to set up a database of cases that serve as a ground-truth, to both train and validate new closure models. In this work, the database consisted of two different cases. The same surface roughness and hub-height velocity were used for both cases, but the turbine configuration was changed, as visualized in Figure 3.1. The turbine and inflow properties correspond to the wind-tunnel experiment from Chamorro and Porté-Agel [14]. The most important parameters are listed in Table 3.1. As a consequence, variations in the inflow conditions and the turbine operational conditions are not considered in this work, however this would be interesting to do in the future. The author chose to look at a variety of configurations of turbines only since it is a more challenging generalization task, compared to parametric variation that might be tackled with surrogate modeling.

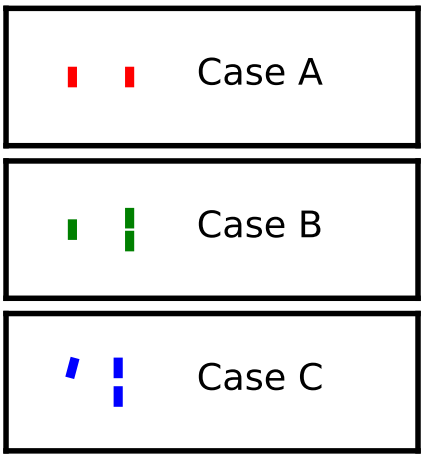


Figure 3.1: Case configurations, turbine diameter is to scale.

For the CFD model, OpenFOAM-6.0 is used in conjunction with the SOWFA-6 toolbox [15]. For the RANS solver, a modified $k-\varepsilon$ model [78] is the baseline closure. For the LES solver, the WALE model is used to model the unresolved scales [64, 78]. The closure coefficients used here for the two models can be found in Table 3.2. Validation of both turbulence models is carried out on the benchmark case from Chamorro and Porte-Agél [14]. Additionally, Xie and Archer's results [107] are used to determine an appropriate mesh resolution for the LES simulations. SOWFA's actuator disc model with the same turbine geometry, rotational speed and force projection parameter is used in both the RANS and LES simulations. The turbine diameter, hub height and rotational speed are given in table 3.1, Stevens et al. [74] give a more detailed description of the rotor geometry. No controller is used in the simulations, and the turbine is run at a fixed rotational speed Ω. For the force projection, the Gaussian width is chosen to be twice the largest cell size in the rotor area $\epsilon = 0.03$m [57]. For simplicity and to avoid interpolation errors, the same mesh resolution was used for both RANS and LES throughout the majority of the work, though in practice the RANS simulations could potentially be run at a

Turbine	
Diameter	$D = 0.15$m
Hub height	$h_{hub} = 0.125$m
Rotational speed	$\Omega = 1190$rpm
Inflow boundary	**layer**
Velocity	$U(h_{hub}) = 2.2$m/s
Turbulence intensity	$\sigma_U(h_{hub}) = 1.0\%$
Mesh	
Domain size	$5.4 \times 1.8 \times 0.46$m^3
Resolution	$360 \times 120 \times 64$

Table 3.1: Case setup parameters

WALE model	
C_e	0.93
C_k	0.0673
C_w	0.325
$k-\varepsilon$ **model**	
C_μ	0.03
$C_{\varepsilon 1}$	1.42
$C_{\varepsilon 2}$	1.92
σ_ε	1.3
σ_k	1.3

Table 3.2: Turbulence model parameters

slightly lower resolution, at least in the wall normal direction. To clarify, while a coarser mesh would help reduce the computational cost of RANS, the main cost reduction as compared to LES comes from the fact that the RANS simulation is steady-state and no time-averaging is necessary. In the next chapter, in Section 4.1.6, a mesh convergence study is carried out by varying the mesh density of the baseline and the corrected RANS simulations. The ABL is modelled in the LES by means of a precursor simulation with doubly periodic boundary conditions, and a uniform body-force applied to achieve the desired hub height velocity. A zero-flux condition was used at the top of the domain for both the precursor and the simulations with turbines. In the latter, periodic boundary conditions were used at the sides, a zero-gradient boundary condition at the outlet, and at the inlet plane, instantaneous fields from the precursor are applied. At the ground, standard boundary conditions for a rough wall are used, see Section 3.2. For both RANS and LES, second-order discretization schemes are used in space with the exception of the convection terms in the turbulence transport equations for the RANS model where a first-order upwind scheme is used for numerical stability. The temporal discretization for the LES simulations is a second order Crank-Nicolson scheme.

Figure 3.2 shows the validation of the models on the benchmark case in terms of mean velocity and turbulence intensity. As expected, RANS over-predicts turbulence intensity and wake recovery as compared to LES. Nevertheless, neither of the models perfectly matches the experiment, possibly also due to the relatively low Reynolds number of the wind tunnel setup whilst the wall functions and the RANS turbulence model are derived for higher Reynolds numbers. The Reynolds number based on the boundary layer height δ of the wind tunnel experiment is $\mathrm{Re}_\delta = U_\infty \delta / \nu \approx 930{,}000$ [14]. Further, the LES simulations show an unphysical overshoot in the turbulent kinetic energy close to the wall. The peak in the turbulent kinetic energy in the LES simulations is a well documented problem for LES simulations with wall functions for rough walls [10]. This is something that can be improved in future work. It is worth noting that the aim of this work is not to perfectly reproduce the experiments, but to showcase the potential of a methodology that systematically improves RANS based predictions using time-averaged

LES data.

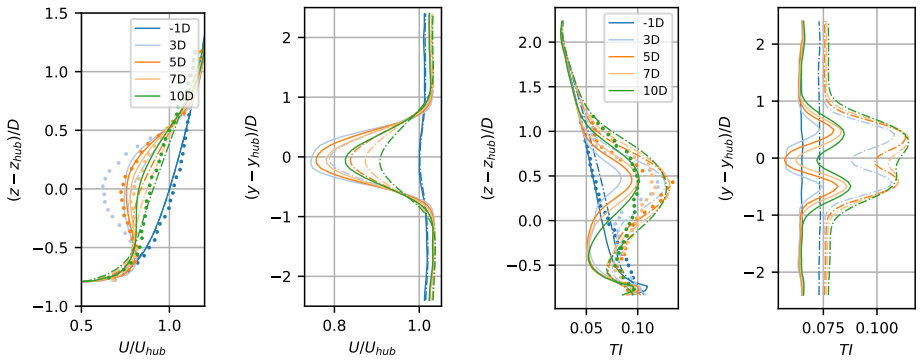


Figure 3.2: Validation of case setup and turbulence models through vertical and horizontal slices of the flow field up and downstream of the rotor plane in terms of velocity and turbulence intensity, solid line corresponds to LES, dash dotted line corresponds to RANS $k-\varepsilon$, and dots refer to experimental results.

3.2. DISCOVERY OF OPTIMAL CORRECTIVE FIELDS

Given LES reference data for a given setup, this work aims to find corrections to the RANS equations in the form of frozen fields for that same setup, such that RANS matches time-averaged LES in terms of mean velocity and turbulence intensity. The core approach presented here was developed by Schmelzer et al. [79], to which we add two modifications specific to the wind-farm application.

The method is similar to the "frozen approach" for estimating turbulence dissipation rates from LES results. Namely variables that are known from LES are injected into model equations, and the values of the remaining variables are deduced. Specifically, let LES quantities be denoted by a $\star$, so the LES mean velocity is $U^\star$, turbulent kinetic energy $k^\star$ and Reynolds stresses $\tau_{ij}^\star$, whereby both resolved and SGS modeled turbulence quantities are implied. Let the baseline $k-\varepsilon$ model be modified in two places: (i) in the momentum equation, a correction to the normalized anisotropy tensor, denoted $\tilde{b}_{ij}^{\Delta}$, is added, and (ii) in the equation for k, a scalar correction term $\tilde{R}$ is added and accounts for errors in the turbulent kinetic energy production and other inconsistencies in the transport equation for the turbulent kinetic energy. These correction terms are both spatially varying fields (tensor and scalar, respectively), and are embedded in the model as:

$$\frac{Dk^\star}{Dt} = \mathscr{P}_k^\star + \tilde{R} - \varepsilon + \frac{\partial}{\partial x_j}\left[(\nu + \nu_t/\sigma_k)\frac{\partial k^\star}{\partial x_j}\right], \tag{3.1}$$

$$\frac{D\varepsilon}{Dt} = \left[C_{\varepsilon 1}\left(\mathscr{P}_k^\star + \tilde{R}\right) - C_{\varepsilon 2}\varepsilon\right]\cdot\frac{\varepsilon}{k^\star} + \frac{\partial}{\partial x_j}\left[(\nu + \nu_t/\sigma_\varepsilon)\frac{\partial \varepsilon}{\partial x_j}\right] \tag{3.2}$$

where the production term is defined as

$$\mathscr{P}_k^\star := 2k^\star b_{ij}^\star S_{ij} \tag{3.3}$$

with

$$b^{\star}_{ij} := \frac{\tau^{\star}_{ij}}{2k^{\star}} - \frac{1}{3}\delta_{ij} = -\frac{\nu_t}{k^{\star}} S^{\star}_{ij} + \tilde{b}^{\Delta}_{ij}. \tag{3.4}$$

The equation to calculate the eddy viscosity is the same as for the baseline model, namely:

$$\nu_t := C_{\mu} \frac{k^{\star 2}}{\varepsilon}. \tag{3.5}$$

Given an initial guess for ε (e.g. from the baseline $k-\varepsilon$ model or just a reasonable constant value), ν_t can be calculated from Eq. (3.5). Then, $\tilde{R}$ can be computed directly from (5.1). Subsequently ε can be updated by solving (5.2) with the most recent $\tilde{R}$, and we iterate back and forth until convergence. Then $\tilde{b}^{\Delta}_{ij}$ can be computed directly from (5.5).

The resulting fields satisfy the modified $k-\varepsilon$ equations, with the LES data as a solution.

In practice, two adjustments are made to this procedure to address issues specific to the wind farm application: (i) blending of the correction terms to zero at the bottom and the top of the domain, and (ii) an atmospheric boundary-layer correction which only varies in the direction perpendicular to the wall.

BLENDING OF THE TURBULENCE CORRECTION TERMS

The blending term at the top and the bottom of the domain is introduced to avoid interaction between the correction terms and the boundary conditions. The blending term F_{β} employed in this work is a simplified version of the one used by Menter [59] for the blending of the $k-\varepsilon$ model and the $k-\omega$ model in the $k-\omega$ *SST* model. It is formulated as

$$F_{\beta}(z) = \begin{cases} \tanh\left[\left(\frac{z}{z_{\text{lower},\beta}}\right)^{\alpha}\right] & \textbf{for } z \leq z_{\text{mid}} \\ \tanh\left[\left(\frac{z_{\text{max}}-z}{z_{\text{max}}-z_{\text{upper},\beta}}\right)^{\alpha}\right] & \textbf{for } z > z_{\text{mid}} \end{cases} \tag{3.6}$$

where the exponent α determines how fast the blending transitions between 0 and 1, $\beta \in \{\text{ABL}, \text{wake}\}$ is used to distinguish between the different blending applied to the correction terms for the ABL precursor and for the main simulation, z_{mid} and z_{max} are related to the domain dimensions, and finally, $z_{\text{lower},\beta}$ and $z_{\text{upper},\beta}$ are domain specific threshold parameters. In [59] the lower bound for the blending is chosen according to the nondimensional wall distance. This is not necessary here, since a relatively simple case with uniform surface roughness and flat terrain is used. Generally, different blending terms can be used for all the correction terms. However, in this particular case, using two different blending functions between the ABL and the wake correction worked well. The parameters used here are found in Table 3.3. The wall blending for the ABL corrections was chosen such that the correction is zero in the first cell center.

MATCHING RANS BOUNDARY-LAYER PROFILES TO LES

In the undisturbed ABL, LES and the baseline RANS model give different profiles for mean-velocity and turbulent kinetic energy. Even though the LES precursor profile is

Parameter	Value
α	4
z_{mid}	0.23m
z_{max}	0.46m
$z_{\text{upper,all}}$	0.4m

Parameter	Value
$z_{\text{lower},ABL}$	0.01m
$z_{\text{lower,wake}}$	0.05m

Table 3.3: Blending parameters for the blending function F_β.

set as the RANS inflow, it evolves before contact with the turbines. As such, if the profiles are not matched, the RANS corrective fields that are discovered will necessarily include some component that corrects the ABL mismatch, and some other component to correct the turbine wake. We prefer to separate these corrections, and so first match the ABL profiles. To achieve this, two modifications are applied: (i) the boundary condition representing the ground for the two simulations is made consistent, and (ii) the velocity profiles away from the boundaries are adjusted through a one-dimensional RANS closure correction varying as a function of wall normal distance only.

Matching the boundary condition at the wall is complicated by the use of wall models in both LES and RANS. In particular, both use equilibrium assumptions and the log-law for a rough wall to determine skin-friction. They assume that first cell is in the log-layer, so that e.g.

$$\tau_{xz} \simeq -\rho u_\star^2$$

but they estimate $u_\star$ differently. In the LES, the time averaged velocity at the first cell U_1 at height z_1 above the wall, is used to estimate an average friction velocity, using

$$u_\star^{\text{LES}} \simeq \frac{\kappa U_1}{\log(z_1/z_0)},$$

where κ is the von Karman constant and z_0 is the surface roughness length. The local instantaneous wall friction is then computed using a Schumann boundary condition. On the other hand, RANS relates the local friction velocity to the turbulent kinetic energy in the first cell

$$u_\star^{\text{RANS}} \simeq \sqrt[4]{C_\mu}\sqrt{k_1}$$

and then uses the log-law to determine an expression for the eddy viscosity there. For consistency, we require that in the RANS boundary-condition

$$C_\mu = \left(u_\star^{\text{LES}}\right)^4 / k_1^2. \tag{3.7}$$

The model parameter C_μ appears also in the definition of the eddy viscosity and it has a large influence on the turbulent kinetic energy. In fact, in this role it can be used to regulate the turbulence intensity at hub height. In the standard $k-\varepsilon$ model [50], the recommended value is $C_\mu = 0.09$, but for atmospheric boundary layers a value of $C_\mu = 0.03$ is often suggested [84].

In addition, in RANS, we use the standard equilibrium-assumption boundary-condition for epsilon [50]:

$$\varepsilon = \frac{C_\mu^{3/4} k^{3/2}}{\kappa z_0}.$$

Having matched boundary conditions between RANS and LES, the profiles of U and k may still not match sufficiently well. In this case, the frozen approach described above can be applied using the LES precursor as a data source, and a RANS simulation of a flat-plate with fully periodic boundary conditions on the sides of the domain (a 1D domain). Two corrections for the ABL, $b_{ij}^{\Delta,\mathrm{ABL}}$ and R^{ABL}, are obtained that can then be used to eliminate remaining differences almost everywhere.

FULL FORMULATION OF CORRECTION TERMS

Finally, now the full formulation for the correction terms can be written as

$$R = \sigma \cdot F_{\mathrm{wake}} \cdot R^{\mathrm{wake}} + F_{\mathrm{ABL}} \cdot R^{\mathrm{ABL}} \tag{3.8}$$

$$b_{ij}^{\Delta} = \sigma \cdot F_{\mathrm{wake}} \cdot b_{ij}^{\Delta,\mathrm{wake}} + F_{\mathrm{ABL}} \cdot b_{ij}^{\Delta,\mathrm{ABL}} \tag{3.9}$$

with blending terms F_β, ABL correction terms R^{ABL}, $b_{ij}^{\Delta,\mathrm{ABL}}$, wake correction terms R^{wake}, $b_{ij}^{\Delta,\mathrm{wake}}$, and classifier σ. In case no classifier is used, σ is one everywhere, otherwise it varies between zero and one. **Henceforth, unless explicitly mentioned, the ABL terms are included in the simulation and the wake superscript will be dropped, such that R^{wake} becomes R and $b_{ij}^{\Delta,\mathrm{wake}}$ becomes b_{ij}^{Δ}.** The next section has a closer look at the classifier. Then, in Section 3.4, generalized expressions for the wake correction terms are inferred. Contrary to the wake correction terms, the ABL correction terms can be used as is. However, this means that they are not general and need to be recomputed if one of the following parameters changes: surface roughness, inflow velocity, and - depending on how strong Coriolis effects are - wind direction.

3.3. SPECIFICATION OF THE CLASSIFICATION TARGET

The corrective fields $\tilde{b}_{ij}^{\Delta}(\mathbf{x})$ and $\tilde{R}(\mathbf{x})$ defined above are non-zero everywhere, but negligible in large regions of the flow. We define a single *classification target* $\tilde{\sigma}(\mathbf{x}) : \Omega \to \{0,1\}$, a function of the spatial coordinate $\mathbf{x}$ taking the value 0 when no model correction is required, and 1 where correction is required. This assessment is based on the values of the corrective fields. Specifically we define

$$\tilde{\sigma}(\mathbf{x}) := \begin{cases} 1 & \text{if } \left(\frac{\left|\tilde{\mathscr{P}}_k^{\Delta}\right|}{\left|\mathscr{P}_k^{\star}\right| + \epsilon} > 0.2 \right) \cup \left(\frac{|\tilde{R}|}{\left|\mathscr{P}_k^{\star}\right| + \epsilon} > 0.2 \right), \\ 0 & \text{otherwise,} \end{cases} \tag{3.10}$$

where $\mathscr{P}_k^{\star}$ is the LES tke production, and

$$\tilde{\mathscr{P}}_k^{\Delta} := 2k^{\star} \tilde{b}_{ij}^{\Delta} \frac{\partial U_i}{\partial x_j},$$

is the extra tke production (beyond Boussinesq) due to $\tilde{b}_{ij}^{\Delta}$. This marker is thus active when the correction due to either $\tilde{R}$ or $\tilde{b}_{ij}^{\Delta}$ exceeds 20% of the reference tke production. Note that $\tilde{P}_k^{\Delta}$ may be zero, even when other measures of the tensor-valued correction $\tilde{b}_{ij}^{\Delta}$ are significant. This choice of classification target was made based on the observation that $\tilde{b}_{ij}^{\Delta}$ tends to have the biggest effect on the mean flow via the production term.

A small threshold, $\epsilon := 0.01$, is added to avoid division by zero. Finally, multi-dimensional Gaussian smoothing with a filter width of two cells has been applied to all the fields in the marker expression before using them to calculate the marker. This mitigates noise in the input data. The *implied* frozen corrections with classification then become:

$$\tilde{R}^{\sigma}(\mathbf{x}) := \tilde{\sigma}(\mathbf{x}) \cdot \tilde{R}(\mathbf{x}), \qquad \tilde{b}_{ij}^{\Delta,\sigma}(\mathbf{x}) := \tilde{\sigma}(\mathbf{x}) \cdot \tilde{b}_{ij}^{\Delta}(\mathbf{x}), \tag{3.11}$$

i.e. the same classification is applied to both corrections, and to all components of $\tilde{b}_{ij}^{\Delta}$. The effect is simply to switch corrections off when they fall below the threshold given in (3.10).

3.4. MODELING THE CORRECTION TERMS AND THE CLASSIFIER

The objective of this section is to take the corrective fields $\tilde{b}_{ij}^{\Delta}(\mathbf{x})$ and $\tilde{R}(\mathbf{x})$, and the classification target $\tilde{\sigma}(\mathbf{x})$ (which are all functions of space), and make generalizable models for them in terms of local flow quantities available to RANS. This is the point at which the methods of supervised machine learning are valuable.

The input features we use are as comprehensive as we can achieve. In a later stage, sparse regression will eliminate features that are not informative. We closely follow Wu et al. [104], and use an integrity basis based on the set $\{\mathbf{S}, \mathbf{\Omega}, \mathbf{A}_p, \mathbf{A}_k\}$ where:

$$\begin{aligned}
\mathbf{S} &:= \frac{1}{2}\frac{k}{\varepsilon}(\nabla U + \nabla U^T), \\
\mathbf{\Omega} &:= \frac{1}{2}\frac{k}{\varepsilon}(\nabla U - \nabla U^T), \\
\mathbf{A}_p &:= -\frac{\sqrt{k}}{\varepsilon} I \times \nabla\left(\frac{p}{\rho}\right), \\
\mathbf{A}_k &:= -\frac{\sqrt{k}}{\varepsilon} I \times \nabla k,
\end{aligned}$$

all of which are non-dimensional. We obtain a generalization of the Pope basis [67] augmented with pressure- and k-gradients, resulting in 47 scalar invariants $\mathbf{I} := [I_1, \ldots, I_{47}]$. In addition, we supplement the feature set with 11 non-dimensionalized physical features such as actuator forcing, $\mathbf{q} := [q_1, \ldots, q_{11}]$. We use all these features when approximating both the corrective fields and the classification target. The full list of features can be found in [86].

Where approximating $\tilde{b}_{ij}^{\Delta}$, we employ the first four Pope basis tensors $T_{ij}^{(n)}$, and by construction guarantee Galilean and rotational invariance:

$$\hat{b}_{ij}^{\Delta}(\mathbf{I}, \mathbf{q}) := \sum_{n=1}^{4} T_{ij}^{(n)} \alpha_n(\mathbf{I}, \mathbf{q}). \tag{3.12}$$

An analogous modeling approach is taken for approximating $\tilde{R}$, i.e.

$$\hat{R}\left(\mathbf{I},\mathbf{q}\right) := 2k\frac{\partial u_i}{\partial x_j}\left[\sum_{n=1}^{4} T_{ij}^{(n)}\beta_n\left(\mathbf{I},\mathbf{q}\right)\right] + \varepsilon\cdot\gamma\left(\mathbf{I},\mathbf{q}\right), \tag{3.13}$$

In the above $\alpha_n(\cdot)$, $\beta_n(\cdot)$ and $\gamma(\cdot)$ are arbitrary scalar-valued functions of the features.

In (3.13) we allow two types of terms: (i) those that mirror a correction to the turbulence production, and (ii) those that represent a correction to the dissipation rate. Since the production term should already be correct (due to knowledge of the RST from LES), we expect most corrections to take the latter form. Nonetheless, the former form allows us to capture other model-form errors - in our case notably the omission of the effect of the rotor forcing on the turbulence.

For the modeling of the classification target $\tilde{\sigma}(\mathbf{x})$ the sigmoid function, $s(x) := 1/(1+\exp(-x))$, is used in conjunction with a scalar-valued function of the features, $\delta(\cdot)$:

$$\hat{\sigma}\left(\mathbf{I},\mathbf{q}\right) := s(\delta(\mathbf{I},\mathbf{q})) \tag{3.14}$$

where – by construction – $\hat{\sigma}\in[0,1]$ for arbitrary $\delta(\cdot)$.

SPARSE REGRESSION

Any functional representation can be used to parameterize the scalar-valued functions α_n, β_n, γ, and δ. We use a library approach with sparse regression [79].

The $47+11=58$ input features are used to build a large library of $L\in\mathbb{N}$ candidate (basis) functions $(\ell_1,\dots\ell_L)$. This is done by recombining features with each other (up to a maximum of three features), and applying exponentiation by $\frac{1}{2}$ and 2. This results in a library cubically larger than the feature set. Each scalar function is then represented as:

$$\alpha(\mathbf{I},\mathbf{q}) := \sum_{k=1}^{L}\theta_k\ell_k(\mathbf{I},\mathbf{q}), \tag{3.15}$$

i.e. a linear representation with coefficients $\boldsymbol{\theta}\in\mathbb{R}^L$. An elastic net is then used to identify an optimal regressor with sparsity (most of the coefficients are zero) [111].

Logistic regression is appropriate for the problem of discovering a classifier σ matching the classification target $\tilde{\sigma}$. In this case, $\delta(\cdot)$ takes the form (3.15), and we solve:

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^L}\left\{\sum_{k=1}^{N}\log\left[\tilde{\sigma}(\mathbf{x}_k)-\hat{\sigma}(\mathbf{I}(\mathbf{x}_k),\mathbf{q}(\mathbf{x}_k))\right]+\lambda\rho\left\|\boldsymbol{\theta}\right\|_1+0.5\lambda\left(1-\rho\right)\left\|\boldsymbol{\theta}\right\|^2\right\}. \tag{3.16}$$

where $1\le k\le N$ indexes the mesh-points of the training data, and $\rho\in[0,1]$ and $\lambda\in\mathbb{R}^+$ control the level of sparseness and coefficient magnitude, respectively.

Given a classifier $\sigma(\mathbf{I},\mathbf{q})$ for the anisotropy correction we solve

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^{4\times L}}\left\{\sum_{k=1}^{N}\hat{\sigma}(\mathbf{I}_k,\mathbf{q}_k)\|\tilde{b}_{ij}^{\Delta}(\mathbf{x}_k)-\hat{b}_{ij}^{\Delta}(\mathbf{I}_k,\mathbf{q}_k)\|_F^2+\lambda\rho\|\boldsymbol{\theta}\|_1+\lambda(1-\rho)\|\boldsymbol{\theta}\|^2\right\},$$

where $\|\cdot\|_F$ is the Frobenius norm, and we have $4\times L$ coefficients in total (due to the four functions $\alpha_1,\dots,\alpha_4$ in (3.12)). The presence of the multiplicative $\sigma(\cdot)$ term in the sum,

limits training to locations where the (previously trained) classifier is active. The model for R is trained similarly.

Note that, in the above, $\boldsymbol{\theta}$ is a placeholder for the regression coefficients and takes different values for R, b_{ij}^{Δ} and σ. Similarly, the values of λ and ρ may be different for each model. The final correction models are then:

$$R(\mathbf{I},\mathbf{q}) := \hat{\sigma}(\mathbf{I},\mathbf{q}) \cdot \hat{R}(\mathbf{I},\mathbf{q}), \qquad b_{ij}^{\Delta}(\mathbf{I},\mathbf{q}) := \hat{\sigma}(\mathbf{I},\mathbf{q}) \cdot \hat{b}_{ij}^{\Delta}(\mathbf{I},\mathbf{q}),$$

in an analogous fashion to (3.11).

The outline of the full procedure is:

1. **Preprocessing**: Use a mutual-information criterion to remove features that have no relationship to the target. Then build the library, and reduce it by *cliqueing* (identifying and removing clusters of multi-colinear functions).

2. **Train a classifier**: Match the classification target with a model expression $\sigma(\cdot)$.

3. **Data reduction**: Use the classifier σ from the previous step as a condition for inclusion of a point in the training dataset.

4. **Model discovery**: Use the elastic net to identify model forms. By varying regularization parameters λ and ρ, obtain an array of models with a variety of non-zero terms.

5. **Model re-calibration**: For each model from the previous step, recalibrate the non-zero model terms using Ridge regression (i.e. L_2 regularization only). Select a regularization parameter λ to encourage small coefficients.

The preprocessing step makes use of two probabilistic procedures: mutual information (MI) [63, 93] and cliqueing [3]. MI can identify nonlinear relations between input features and correction terms and can hence help reduce the input feature set. Cliqueing checks if there is multi-collinearity in the input library and is thus useful for discarding co-linear input functions. Both of these procedures are vital for making the learning procedure manageable for our dataset.

4

PRACTICAL APPLICATION[1]

This chapter presents the results of the step-by-step application of the methodology described in Chapter 3. The chapter has two sections. Section 4.1 shows how well the approach works when no classifier is used, while Section 4.2 introduces models that are combined with a classifier and compares the results to those obtained without classifier.

The models obtained from the regression procedure are explicit, hence they can be directly integrated into the RANS solver. However, in order to study the errors of the optimal corrections from the k-frozen approach, the errors introduced by the sparse regression, and the errors in the final coupled models separately, we consider three kinds of corrections. To avoid confusion short definitions are listed below:

- *Frozen (or optimal)* refers to correction terms obtained from the frozen procedure of Section 3.2. For frozen corrections, the terminology is $\tilde{R}$ and $\tilde{b}_{ij}^{\Delta}$. For the classifier, there is no optimal solution, there is only the one we define with the threshold criteria. That one will be denoted as $\tilde{\sigma}$.

- *Fixed* refers to the correction term that results from applying a trained model to the LES flow field without coupling said correction to the flow field. This is generally a good representation of the optimal correction, but includes errors due to the inability of the elastic-net to represent the optimal correction with the given features. For fixed corrections, the terminology is $\hat{R}$, $\hat{b}_{ij}^{\Delta}$, and $\hat{\sigma}$.

- *Coupled* refers to a correction term that is a function of the flow field, so that it changes as the flow field changes, e.g. at every iteration of the flow solver. In this sense, it is a genuine turbulence model, operating independently of LES data. For coupled correction terms, the terminology is R, b_{ij}^{Δ}, and σ.

- *Specific models* will be denoted via a subscript independent of whether they are fixed or coupled, for example $[R]_1$ or $\hat{b}_1^{\Delta}$.

[1] The writing presented in this chapter is heavily based on the two most recent publications of the author which are listed in B.2.

4.1. WIND TUNNEL SETUP WITHOUT CLASSIFIER

This section details the models derived when applying the regression procedure directly to the data obtained from the frozen step. Subsection 4.1.1 will show how the RANS model is matched to the LES precursor data, such that the inflow profiles are in equilibrium. Then, Subsection 4.1.2 will show how closely the RANS simulation matches the LES data when injected with the frozen correction for the turbine wakes. Subsection 4.1.3 will document the application of the regression algorithm to the frozen data. Following this, in Subsection 4.1.4, the robustness of the learned correction terms is tested and the results are used to aid the model selection. What happens when the RANS model is fully coupled with the correction terms is shown in Subsection 4.1.5. Finally, in Subsection 4.1.6, a mesh sensitivity study is done to show if the models still work with a coarser and finer mesh.

4.1.1. MATCHING RANS BOUNDARY-LAYER PROFILES TO LES

The matching procedure for the undisturbed boundary layer has been described in Section 3.2. The parameter C_μ is chosen according to relation 3.2 that is based on the wall friction and wall tke from the LES data. This yields $C_\mu = 0.055$. This will match the boundary conditions perceived by RANS to the ones used for the LES simulation. For the baseline RANS simulations, the standard value of $C_\mu = 0.03$ will be chosen [84].

Having matched boundary conditions between RANS and LES, the profiles of U and k still do not match sufficiently well. The frozen approach described in Section 3.2 is used to match the profile better. The LES precursor is used as a data source, and a RANS simulation of a flat-plate with fully periodic boundary conditions on the sides of the domain (a 1D domain) is performed. Two corrections for the ABL, $b_{ij}^{\Delta,ABL}$ and R^{ABL}, are obtained that eliminate remaining differences almost everywhere.

Figure 4.1 shows the resulting profiles from the frozen approach and then the profiles in the case where the corrections are propagated (referred to as a corrected simulation). For corrected cases, the domain forcing is chosen such that the hub height velocity matches. The velocity profiles between the frozen and the corrected RANS simulation match very well, but the turbulent kinetic energy profiles do not match well close to the wall. In fact, the unphysical overshoot in the turbulent kinetic energy is also observed in the corrected RANS simulations, even though the peak was removed from the LES reference data. However, the turbulent kinetic energy in the rotor wake matches well between LES and corrected RANS, and this is what is relevant for this work.

4.1.2. FLOW FIELD WITH OPTIMAL CORRECTION TERMS

The optimal correction terms are derived for the training case A. Subsequently, the (static) optimal corrections are integrated into the RANS turbulence models for this setup. The results obtained from this are referred to as “frozen” or “optimally corrected” RANS. Figures 4.2 and 4.3 show the wake development as predicted by the LES, the baseline RANS, and the frozen RANS simulations using vertical slices through the flow field. The horizontal slices can be found in the appendix in Figures B.1 and B.2. Optimally corrected RANS represent the best-case scenario that can be obtained when using this methodology. In the next subsection, the generalized models for the correction terms will intro-

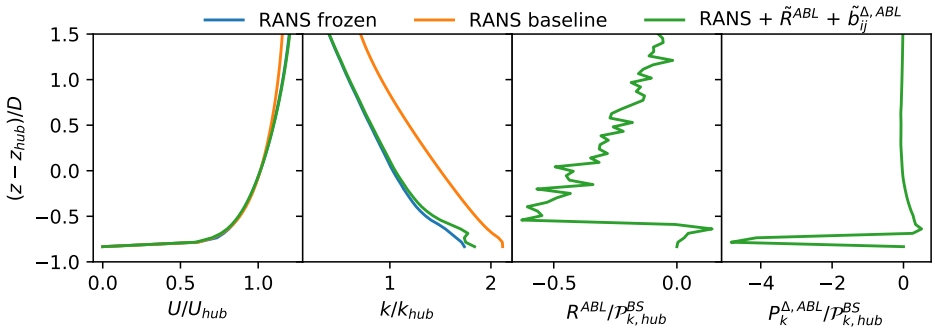


Figure 4.1: Matching of ABL profiles between the frozen RANS, the baseline RANS and, the corrected RANS simulations. The one-dimensional profiles for the velocity, the turbulent kinetic energy and the two correction terms are shown. The scalar correction term is normalized with the Boussinesq turbulent kinetic energy production at hub height.

duce additional errors. The results in the figure show that indeed the optimal correction terms lead to an almost perfect match between LES mean and frozen RANS velocity and turbulent kinetic energy fields.

The relative importance of the different frozen correction terms for the prediction of the velocity and turbulent kinetic energy field is also visible from Figures 4.2 and 4.3, respectively. Some conclusions can be drawn from the selective inclusion of the correction terms. The free-stream corrections, R^{ABL} and $b_{ij}^{\Delta,ABL}$, do not have much effect on the velocity field, but they slightly reduce the overprediction of the turbulent kinetic energy. Of course, this is closely tied to the choice of C_μ. The anisotropy correction term b_{ij}^{Δ} is more important than the scalar correction terms R. If only a correct prediction of the velocity field is necessary, then the scalar term R can be neglected. However, the scalar correction term R does yield some improvement for the prediction of the turbulent kinetic energy over the case where only the tensor correction term b_{ij}^{Δ} is used.

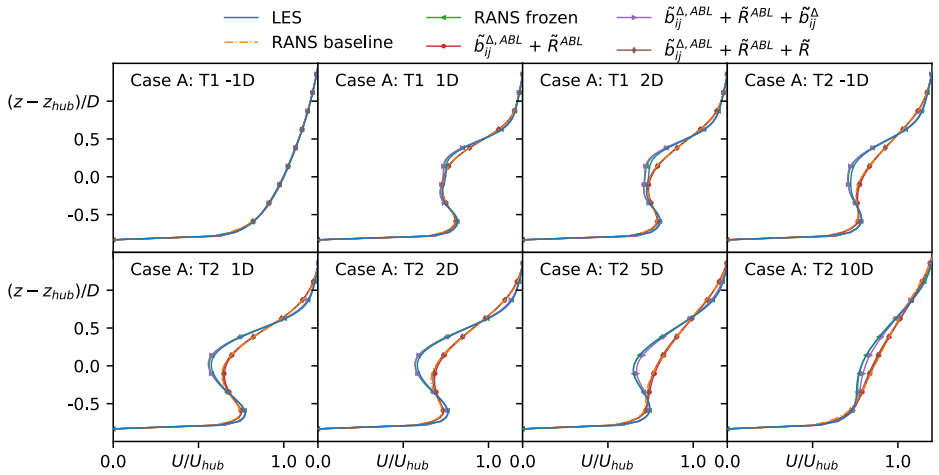


Figure 4.2: Comparison between LES, RANS baseline, and frozen RANS with selective inclusion of the different components of the (frozen) correction terms. Vertical slices of the velocity field up and downstream of the rotor plane of the two turbines of case A.

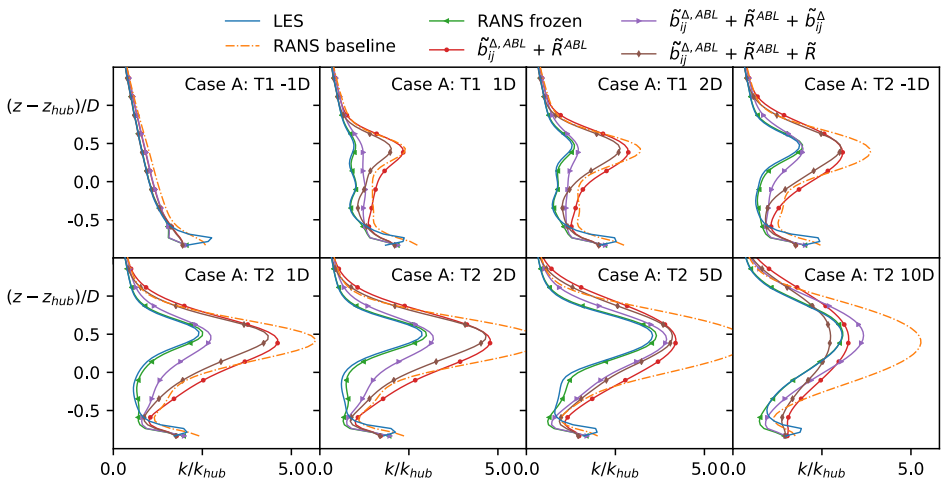


Figure 4.3: Comparison between LES, RANS baseline, and frozen RANS with selective inclusion of the different components of the (frozen) correction terms. Vertical slices of the turbulent kinetic energy field up and downstream of the rotor plane of the two turbines of case A.

4.1.3. LEARNING OF CORRECTION TERMS

The results presented in the following are based on the datasets for configuration A & B, as presented in Figure 3.1. Case A is used for the training of the models and case B is used to cross-validate the learned correction terms. Note that case B is more complex than case A because it includes one more turbine and one of the turbines is yawed with respect to the incoming flow. The training dataset does not include the entire dataset of case A, rather only entries centered around the turbines wake are used. This helps to avoid overfitting and reduces the dataset somewhat. The exact criteria for inclusion in the training dataset are $x_{rotor} - 1D < x < x_{rotor} + 20D$, $y_{rotor} - 1.5D < y < y_{rotor} + 1.5D$, and $0.05\text{m} < z < z_{rotor} + 1.5D$.

ID	Description	Raw feature	Normalization
q_γ	Shear parameter	$\left\|\frac{\partial U_i}{\partial x_j}\right\|$	$\frac{\varepsilon}{k}$
q_τ	Ratio of total to normal Reynolds stresses	$\|\|\overline{u_i' u_j'}_{Boussinesq}\|\|$	k
q_ν	Viscosity ratio	ν_t	100ν
$q_{TI}^\dagger$	Turbulence intensity	k	$\frac{1}{2}U_i U_i$
$q_F^\dagger$	Actuator forcing	$\|F_{cell}\|$	$\frac{1}{2}\rho_0 A_{cell}\|U\|^2$

Table 4.1: List of non-dimensionalized physical features used in the model discovery phase and their precise definition. The features that are not Galilean invariant are marked with †.

Following the methodology outlined previously, the feature set used to construct a library of basis functions is based on the results of the mutual information analysis between features and correction terms. A list of the input feature set divided into physical parameters and invariants that were obtained as a result of the preprocessing step can be found in Tables 4.1 and 4.2, respectively. The full list of physical features and invariants used as an input to the mutual information algorithm can be found in the Appendices A.2 and A.1. Additionally, only the first four tensors of the integrity basis are used where $T^{(1)} = S$, $T^{(2)} = S\Omega - \Omega S$, $T^{(3)} = \text{dev}(S^2)$, $T^{(4)} = \text{dev}(\Omega^2)$ and where dev is the deviatoric part of the tensor. Applying the cliqueing algorithm to the library of basis functions that was constructed from the reduced feature set, further reduced the size of the library by around a factor of 6.

Invariant ID	Definition
I_1	$\mathbf{S}^2$
I_2	$\mathbf{\Omega}^2$
I_{19}	$\mathbf{\Omega A}_k\mathbf{S}^2$
I_{25}	$\mathbf{A}_k^2\mathbf{S\Omega S}^2$
I_{35}	$\mathbf{A}_p\mathbf{A}_k\mathbf{S}^2$

Tensor ID	Definition	Normalization
$\mathbf{S}$	$\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$	$\frac{\varepsilon}{k}$
$\mathbf{\Omega}$	$\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i}\right)$	$\frac{\varepsilon}{k}$
$\mathbf{A}_k$	$-I \times \nabla p$	$\frac{\varepsilon}{\sqrt{k}}$
$\mathbf{A}_p$	$-I \times \nabla k$	$\rho_0\|u\nabla u\|$

Table 4.2: List of invariants used in the model discovery phase and their precise definition.

Subsequently, the three-step regularization methodology is applied to determine, firstly which candidate functions are important, and secondly what the magnitude of these coefficients should be. The result is a very large number of potential models. Figure 4.4 shows the results of this process for both the anisotropy correction b_{ij}^{Δ} and the scalar correction term R. The left side of the figure illustrates the trade-off between the anticipated robustness and the model accuracy by showing the influence of the Ridge regularization parameter λ_R on the mean and maximum error of the model on the training data set. The right side of the figure visualizes the trade-off between the model complexity and the model accuracy by highlighting the number of terms of the model. But the results are not straightforward and only limited trends can be identified. In general, more complex models are seen to give better predictions for both correction terms, but this is not always the case. The trend with respect to an increasing regularization parameter λ_R is different for the two correction terms. For the anisotropy correction b_{ij}^{Δ}, higher regularization correlates with a higher mean error but a lower maximum error. For the scalar correction term R, higher regularization generally leads to both higher mean and maximum error.

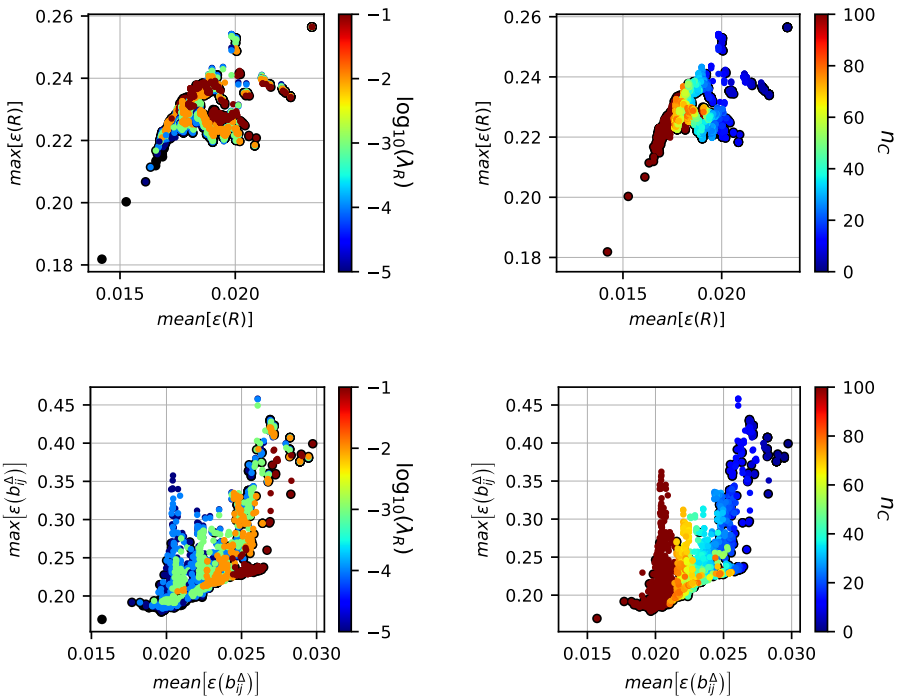


Figure 4.4: Scatter plot of all the models obtained for both correction terms. Members of the three-dimensional Pareto front with respect to mean and max error, as well as model complexity are highlighted in black. The coloring of the elements is according to the magnitude of the Ridge λ_R penalization parameter and the model complexity n_C .

Because the model discovery and calibration phase generate many models and because it was difficult to pick which models should be selected for further investigation, the three-dimensional Pareto front in terms of mean error, maximum error, and model complexity was computed. This is indicated in Figure 4.4 by black outlines. Going forward only the models which are a member of the Pareto front are investigated. Since the number of Pareto optimal models is still of the order of around 500, a further automated selection procedure is necessary. Cliqueing was again applied, this time to predictions of complete models, and models that were too similar were discarded.

The effect of this procedure is visualized in Figures 4.5 and 4.6 for R and b^{Δ} respectively. The figures show the spread of Pareto-optimal models, the selection of models obtained from the cliqueing, and finally the models selected for implementation in the CFD solver. The anisotropy correction is visualized by means of its effect on the turbulent kinetic energy production $\mathscr{P}_k^{\Delta} = 2kb_{ij}^{\Delta}\frac{\partial u_i}{\partial x_j}$. Our experience shows that this is a good indicator for the accuracy of the anisotropy correction term, and substantially easier to visualize.

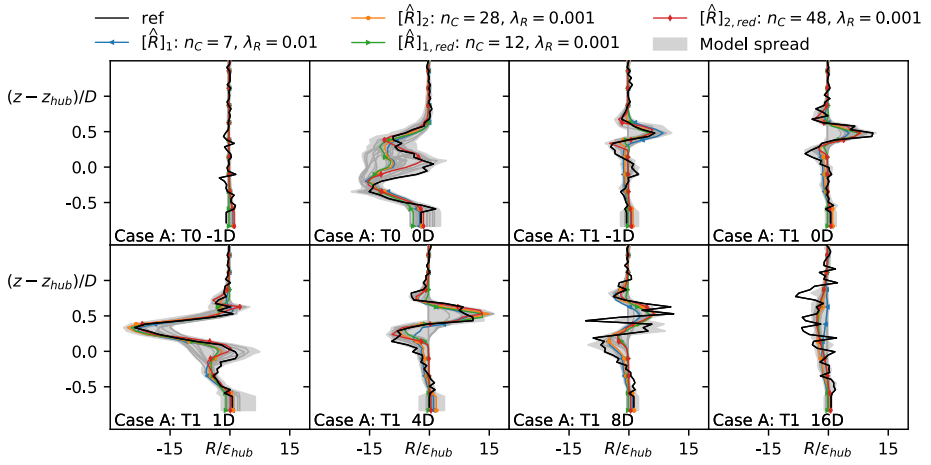


Figure 4.5: Spread of trained models for R for Case A. The subscript $R_{a,red}$ refers to models whose terms only contain positive powers. Vertical slices at the rotor plane at different streamwise stations. The model spread is for all models that are Pareto optimal. The models selected during the cliqueing post-processing step are shown explicitly either in color or in dark gray. The models selected for further investigation are highlighted in color. Finally, the optimal correction term is shown in black.

Figure 4.5 shows the predictions for the scalar correction term R. It can be seen that the entire selection of the spread of models can be reduced to about 20 models. The four models highlighted in color are the ones that will be implemented in the CFD solver in the next section. The four models were selected based on accuracy and complexity. Further, also the models named $[R]_1$ and $[R]_2$ contain terms with both negative and positive powers of the input features, whereas the models named $[R]_{1,red}$ and $[R]_{2,red}$ (i.e. reduced) only contain terms with positive powers. This was done because the negative powers negatively affected the convergence of the models once implemented in the CFD

model. For the R term this effect was not always present, but for the anisotropy correction term b_{ij}^{Δ} none of the models including negative powers lead to convergence. As such they are not discussed further.

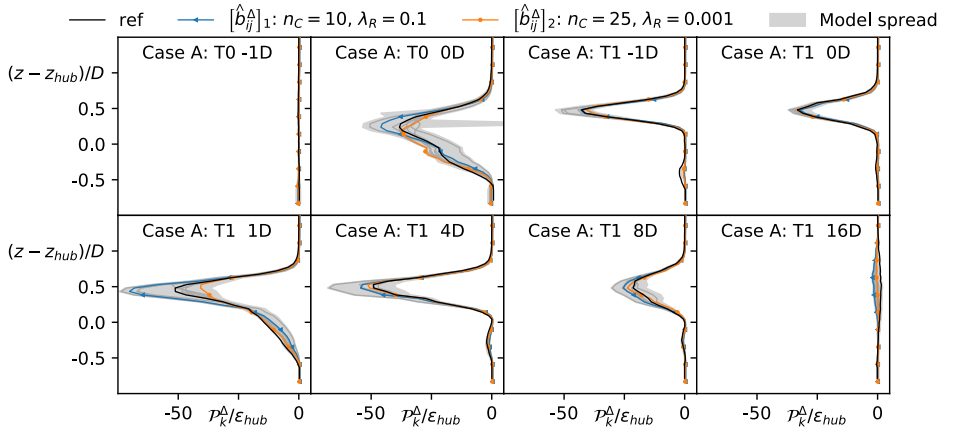


Figure 4.6: Spread of trained models for b^{Δ}, visualized using $\mathscr{P}_k^{\Delta}$ for case A. Vertical slices at the rotor plane at different streamwise stations. The model spread is for Pareto optimal models. The models selected during the cliqueing post-processing step are shown explicitly either in color or in dark gray. The models selected for further investigation are highlighted in color. Finally, the optimal correction term is shown in black.

Figure 4.6 shows the spread of the model prediction for the anisotropy prediction. Again, with a reduced set of about 10 models, the entire spread of results can be covered. Two models were selected for further investigation as a trade-off between accuracy and complexity. All of the models contain only positive powers of the input features. Going forward the two selected models will be referred to as $\left[b_{ij}^{\Delta}\right]_1$ and $\left[b_{ij}^{\Delta}\right]_2$.

4.1.4. ROBUSTNESS OF CORRECTION TERMS

The correction-learning methodology employed in this work is completely decoupled from the CFD model. This significantly simplifies the regression as compared to an online approach, where the terms are trained while coupled with the CFD model. However, this also means that once a coupling with the CFD solver is constructed, the correction terms may not be the same as predicted during the learning stage. Further, at this point, there is no clear criteria or methodology to determine the stability of a correction model *a priori*. Hence simple testing and cross-validation is the most immediate strategy.

Experience with the framework has shown that models that are very complex, i.e. above about 50 terms, tend not to converge for either one of the correction terms. Furthermore, in models trained on our data, the Ridge regression parameter should be at least $\lambda_R \geq 0.001$ to assure convergence not only on the training but also on test datasets.

To make the assessment of the models structured, the robustness of the two correction terms is assessed separately before implementing both terms simultaneously in the

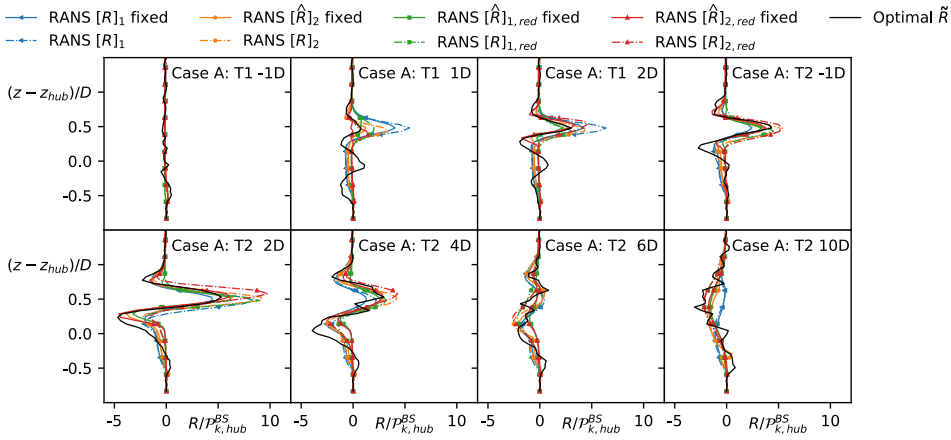


Figure 4.7: Spread of fixed and coupled corrections R for the training case A via vertical slices at the rotor plane up and downstream of the two turbines. The subscript $R_{a,red}$ refers to models whose terms only contain positive powers.

turbulence model. For example, the robustness of the model for R can be assessed by using the frozen correction for b_{ij}^{Δ}, and vice versa. In Figures 4.7 and 4.8 the robustness of the previously selected correction terms is shown for the scalar and the anisotropy correction models on the training setup, i.e. case A.

Figure 4.7 compares the optimal correction terms for R with the one obtained when coupled with the CFD model and the fixed one obtained during the learning phase with no coupling to the CFD solver. Ideally, the coupled and the fixed term would overlap perfectly. However, as visible from the figure, this is not the case and the effect is more or less pronounced for the different pictured correction models. The discrepancy between the coupled and the fixed terms is larger in regions where the optimal term has high gradients. If the discrepancy between the two terms is too large, the model not only becomes inaccurate but may also lead to an unstable coupling once both correction terms are introduced simultaneously.

Figure 4.8 shows the same analysis for the two selected model for the anisotropy correction b_{ij}^{Δ} in terms of the modified turbulent kinetic energy production term $\mathscr{P}_k^{\Delta} = 2kb_{ij}^{\Delta}\frac{\partial u_i}{\partial x_j}$. Again, in regions where the optimal correction term and its derivative are large, the disparity between the fixed and the coupled term is largest. Nevertheless, both terms lead to a converging simulation and hence will be further tested going forward.

4.1.5. FLOW FIELD WITH LEARNED CORRECTION TERMS

Finally, now that model selection and assessment of model robustness have been carried out, both correction terms can be implemented simultaneously while coupled to the RANS flow field. The models will be tested on case B, since they were trained on case A. The vertical profiles of velocity and k are shown in Figures 4.9 and 4.10 for case B. For comparison, also the baseline model, the frozen case, and the flow field with fixed

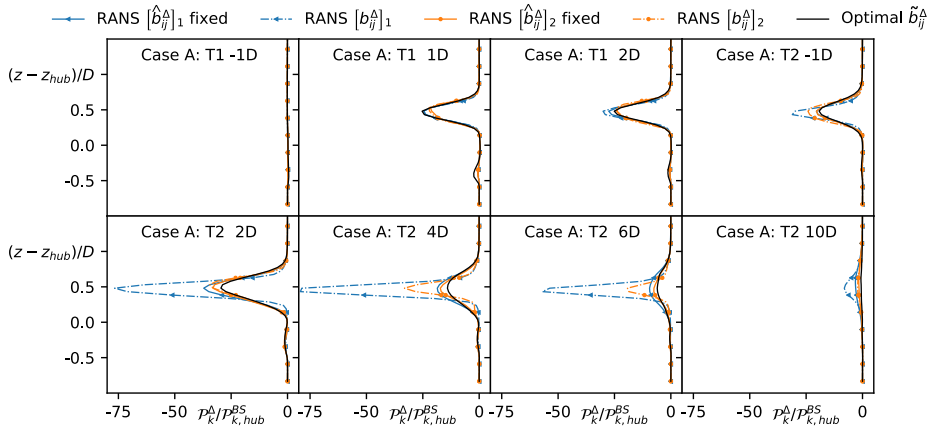


Figure 4.8: Spread of fixed and coupled corrections b_{ij}^{Δ} in terms of $\mathscr{P}_k^{\Delta}$ for the training case A via vertical slices at the rotor plane up and downstream of the two turbines.

learned correction terms are shown. For the cases with the fixed correction terms, only the spread between all possible combinations of the correction terms is shown.

As is visible from the figures, the spread between the simulations with the fixed correction terms is smaller than the spread for the simulations where the correction terms are coupled to the RANS velocity field. This is quite logical given the results from the robustness analysis. Nevertheless, all the shown models yield a solid improvement over the baseline model in the wake region. No results for the scalar correction term $[R]_{2,red}$ are shown because this term would lead to diverging simulations on both the test and training case, even when strong under-relaxation was used.

All the velocity profiles in Figure 4.9 from simulations with the coupled correction terms show significant improvement over the baseline model. In fact, the spread between the different models is minimal and the difference between the fixed and the coupled models is quite small. However, the discrepancy with respect to the reference profile increases further downstream akin to an error accumulation. Thus, it would be interesting to test the models on a case with more turbines to see how robust the enhanced models actually are.

In Figure 4.10 the tke profiles are shown for the baseline and the improved models. Here, the spread between the coupled and the fixed models is larger, especially in the near wake of the second rotor. Comparison between the results for the various enhanced models indicates that the scalar correction term R is what leads to the large spread between the models in the wake of the turbine. As compared to the velocity profiles, the discrepancy with respect to the reference does not increase downstream which is encouraging. There is also an unphysical underprediction of the tke close to the wall for the frozen case which is not present in the enhanced simulations: it seems the coupling with the flow field helps alleviate it. However, the tke close to the wall is still lower than the one for the reference time-averaged LES simulations, which show an unphysi-

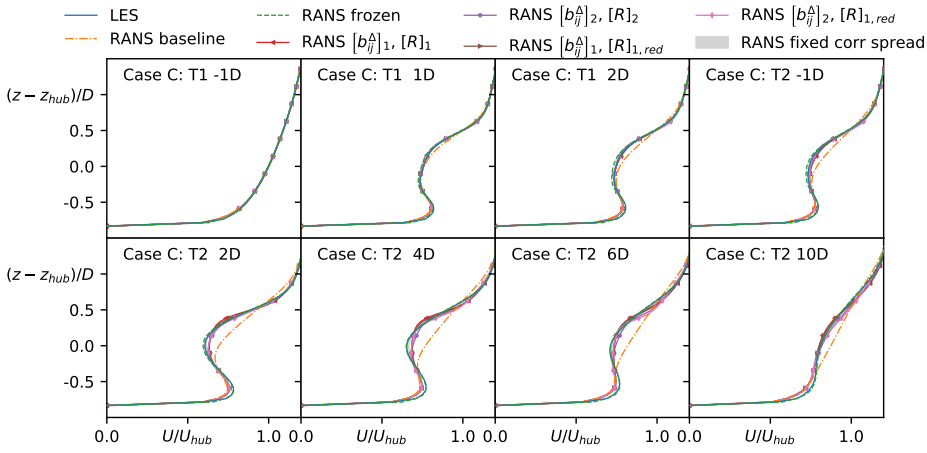


Figure 4.9: Comparison between LES, RANS baseline, frozen RANS and corrected RANS models via vertical slices of the velocity field up and downstream of the rotor plane for the three turbines of case B.

cal overshoot there, so this discrepancy is actually a positive. The peak in the tke in the LES simulations is a well documented problem for LES simulations with wall functions for rough walls [10]. This could be addressed in further work.

Overall, the combination of the correction terms $[R]_{1,red}$ and $\left[b_{ij}^{\Delta}\right]_2$ yielded the best results and hence the full formulation for these correction terms is:

$$\begin{aligned}
[R]_{1red} = 2kS_{ij}\,[\quad & 1.4771\cdot10^{-4}\cdot I_1^{0.5}\cdot q_v^{3.0}\cdot\mathbf{T}_{ij}^{(1)} - 1.9183\cdot q_{TI}^{0.5}\cdot q_F^{1.5}\cdot\mathbf{T}_{ij}^{(4)}] \\
+\varepsilon\,[\quad & 1.0970\cdot10^{1}\cdot q_{TI}^{0.5}\cdot q_F\cdot I_1^{0.5} + 6.1657\cdot10^{-5}\cdot q_{TI}\cdot I_1^{2.0}\cdot I_{34} \\
& + 8.3864\cdot10^{-3}\cdot q_{TI}^{1.5}\cdot I_{25} - 1.7888\cdot10^{2}\cdot q_{TI}^{2.0}\cdot I_{25} \\
& - 1.3956\cdot10^{1}\cdot q_F\cdot q_\gamma^{0.5} + 2.5231\cdot10^{-7}\cdot q_{TI}^{2.5}\cdot I_{25}^{2.0} \\
& - 2.2330\cdot q_F\cdot q_\gamma - 5.2367\cdot10^{-6}\cdot I_1^{2.0}\cdot q_v^{4.0} \\
& - 5.5597\cdot10^{-2}\cdot q_v^{3.0}]
\end{aligned} \tag{4.1}$$

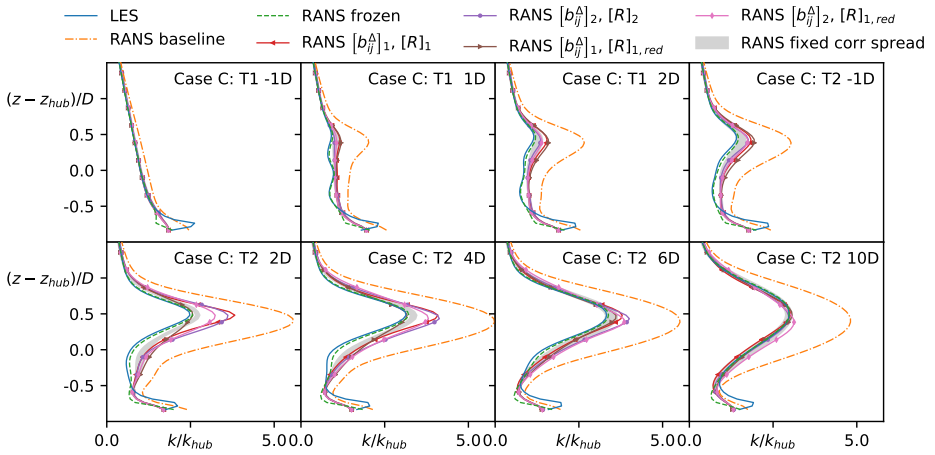


Figure 4.10: Comparison between LES, RANS baseline, frozen RANS and corrected RANS models via vertical slices of the turbulent kinetic energy field up and downstream of the rotor plane for the three turbines of case B.

and

$$
\begin{aligned}
\left[b_{ij}^{\Delta}\right]_2 = [\quad & 2.5095\cdot10^{-2}\cdot q_{TI}^{0.5}\cdot I_1^{0.5} + 1.090\cdot10^{-5}\cdot q_{TI}\cdot q_F^{0.5}\cdot I_1^{2.0} \\
& + 3.4089\cdot10^{-4}\cdot q_{TI}^{2.0}\cdot q_F^{0.5}\cdot I_1^{2.0} - 4.0175\cdot10^{-6}\cdot q_{TI}^{2.0}\cdot I_1^{2.0}\cdot q_v \\
& - 3.6356\cdot10^{-5}\cdot q_{TI}^{2.0}\cdot I_1^{2.5} + 9.6825\cdot10^{1}\cdot q_{TI}^{3.0}\cdot q_v^{2.0} \\
& - 2.8904\cdot10^{3}\cdot q_{TI}^{3.5} + 6.1482\cdot10^{-2}\cdot q_F^{0.5} \\
& - 9.4482\cdot10^{-5}\cdot q_F^{0.5}\cdot I_1\cdot q_v^{2.0} - 2.1767\cdot10^{-3}\cdot q_v^{2.5} \\
& + 8.6126\cdot10^{-4}\cdot I_1^{0.5}]\cdot\mathbf{T}_{ij}^{(1)} \\
+ [& - 9.4932\cdot10^{-2}\cdot q_{TI}^{0.5}\cdot q_F + 1.0716\cdot10^{-2}\cdot q_{TI}^{0.5}\cdot q_F^{1.5} \\
& + 6.3229\cdot10^{-4}\cdot q_{TI}^{0.5}\cdot q_v^{2.5} + 6.3233\cdot10^{-5}\cdot q_{TI}^{0.5}\cdot q_v^{3.0} \\
& + 3.7871\cdot10^{-4}\cdot q_{TI}\cdot I_{34} + 7.5746\cdot10^{-4}\cdot q_{TI}^{2.5}\cdot I_{18} \\
& - 1.7673\cdot10^{3}\cdot q_{TI}^{4.5} + 4.8578\cdot10^{-3}\cdot q_F \\
& - 4.1741\cdot10^{-8}\cdot I_1^{0.5}\cdot I_2 + 1.3261\cdot10^{-6}\cdot I_1]\cdot\mathbf{T}_{ij}^{(2)} \\
+ [& - 1.3262\cdot10^{-3} - 2.7248\cdot10^{-6}\cdot I_1^{0.5}\cdot q_v^{4.0} \\
& + 6.5684\cdot10^{-7}\cdot I_1\cdot q_v^{2.5}]\cdot\mathbf{T}_{ij}^{(3)} \\
& - 3.5887\cdot10^{-5}\cdot q_v^{4.5}\cdot\mathbf{T}_{ij}^{(4)}
\end{aligned}
\tag{4.2}
$$

The anisotropy correction term $\left[b_{ij}^{\Delta}\right]_2$ consists of 25 terms of which 11 are multiplied by the first tensor of Pope's invariant basis, $\mathbf{T}^{(1)} = \mathbf{S}$, i.e. the correction tensor is linear. Thus, this part of the correction tensor is implemented in the turbulence model

in a semi-implicit form, and the remaining non-linear terms are incorporated in a fully explicit manner. It is expected that this further increases the stability of the numerical implementation.

Some of the coefficients for the two correction models have a very small magnitude, so it may seem that they are not necessary. However, the influence of neglecting each coefficient was checked and the shown coefficients all result in a change in the relative mean or maximum error of at least three percent as compared to the full formulation shown above. Hence, all the shown terms have a non-negligible contribution. Nevertheless, the models are quite complex and there is also at least partial cancellation between the different terms.

4.1.6. MESH CONVERGENCE WITH LEARNED CORRECTION TERMS

As pointed out by van der Laan [91], nonlinear eddy viscosity models can be prone to numerical instability when a fine mesh is used. To check whether the results of the developed model correction terms are actually grid-independent, a mesh convergence study is carried out both for the baseline, as well as, the corrected model. The results for the vertical velocity and the turbulent kinetic energy profiles are shown in Figures 4.11 and 4.12 for case B. The mesh properties are shown in Table 4.3.

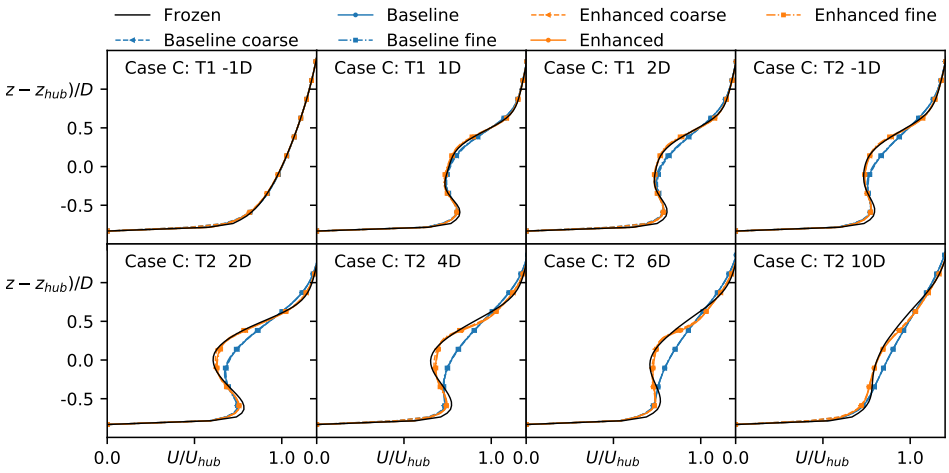


Figure 4.11: Mesh convergence study for the baseline and the corrected model. Shown are the vertical slices of the velocity field upstream and downstream of the rotor plane for the three turbines of case B.

The velocity profiles in Figure 4.11 are insensitive to the mesh for both the baseline and the corrected models. Hence, in terms of velocity the results are close to mesh independent at the presented refinement levels. There is more variation in the tke, see Figure 4.12, and the baseline model shows less sensitivity than for the corrected model. However, even for the corrected model the difference between the reference and the fine mesh is small indicating that the mesh is fine enough and that results are largely mesh independent.

Name	Density $n_x \times n_y \times n_z$
Coarse	$240 \times 120 \times 48$
Reference (same as for LES)	$360 \times 120 \times 64$
Fine	$540 \times 240 \times 64$

Table 4.3: Mesh convergence parameters.

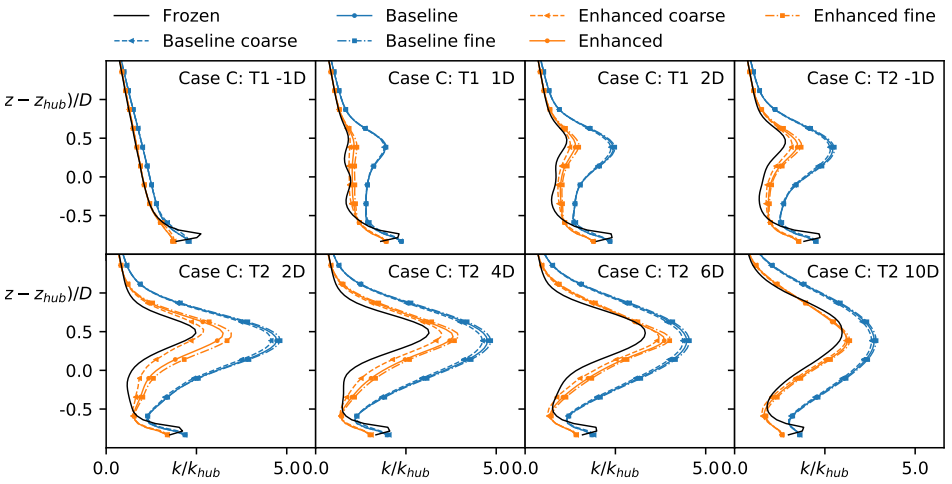


Figure 4.12: Mesh convergence study for the baseline and the corrected model. Shown are the vertical slices of the turbulent kinetic energy field up and downstream of the rotor plane for the three turbines of case B.

Overall, these results are encouraging. The correction terms depend on the normalized rotor forcing which in turn depends on the actuator model, so there is an additional coupling loop in the prediction. Despite this there is little variation between the results.

4.2. WIND TUNNEL SETUP WITH CLASSIFIER

In this section, the regression procedure will not be directly applied to the corrections obtained from the frozen procedure. A classifier will be used as an intermediate step to discard data points for which the corrections are below a certain threshold. This dramatically reduced the training data set for the turbulence corrections. However, the classifier needs to be trained on the full data set which adds computational cost. Due to this modified training process, the resulting models will look different. Here, the results are compared to the final model from the previous section. The exact model formulation can be found in equations 4.2 and 4.1 and will be referred to as the reference model.

The section is structured as follows. First, in Section 4.2.1, the results from injecting the frozen corrections in the RANS simulations are again shown, but additionally also the effect of the ideal classifier will be pictured. Then, in Sections 4.2.2 and 4.2.3, the results from the training process for the classifier and the two correction terms are shown,

respectively. In Section 4.2.4 two limiter functions are introduced to increase the robustness of the models. Finally, in Section 4.2.5 the obtained models are coupled with the CFD solver and they are compared to the models obtained without using the classifier.

4.2.1. INJECTION OF FROZEN CORRECTION TERMS

We first inject the frozen corrections of (3.11) into a RANS simulation. The fields $\tilde{R}^{\sigma}(\mathbf{x})$ and $\tilde{b}_{ij}^{\Delta,\sigma}(\mathbf{x})$ are calculated using the procedure of Section 3.2 for Case A, and are thereafter injected into a RANS simulation of the same case. This is therefore not a predictive test, but is useful for assessing the best-case scenario that can be obtained using our methodology. In the next subsection, the models for the correction terms will introduce additional errors.

Figure 4.13 shows the wall-normal profiles of the flow velocity and the turbulent kinetic energy (tke) (non-dimensionalised by their values at the turbine hub height) as a function of non-dimensional height, at different stream-wise locations in the domain: from a distance of $-1D$ upstream of the first turbine T1 to a distance of $10D$ downstream of the second turbine T2, where D is the turbine diameter. The locations of the turbines and the locations of the profiles are indicated in the figure. Wake profiles are shown for the LES, the baseline RANS ($k-\varepsilon$), and the frozen RANS simulations.

The results in the figure show that – in terms of these quantities – indeed the frozen correction terms lead to an almost perfect match between the LES mean and frozen RANS velocity, and only localized mismatch of k near the wall. In particular we see that restricting the corrective fields based on the threshold of (3.10) has not significantly harmed the quality of the fit. On this basis we progress to discovering models for the corrections.

Note that we can assess the relative importance of the two corrective fields by injecting only one or the other, see Figure 4.13. The anisotropy correction term $\tilde{b}_{ij}^{\Delta,\sigma}$ is significantly more important than the k-equation correction $\tilde{R}^{\sigma}$. In fact, if only a correct prediction of the velocity field is necessary, then R could be neglected completely. However, it does yield a significant improvement in the prediction of the turbulent kinetic energy, suggesting that in a predictive context it might become important.

4.2.2. TRAINING OF THE CLASSIFIER

Based on the mutual information analysis, the feature set for the classifier was reduced to only four variables, listed in the first four rows of Table 4.4. Other features did not show significant correlation with the classification target. Varying the regularization parameters of the elastic net resulted in the identification of a large number of classifiers, of which five were selected for further testing based on complexity and achieved fit. These are denoted σ_1 to σ_5 in the following. The complexity of the chosen classifiers ranges from one to nine terms; and notably much more complex models do not show a significant increase in accuracy. Figure 4.14 is a visualization of the terms used. There is significant overlap between the terms used by the classifiers, notably turbulence intensity, velocity shear and eddy viscosity ratio are dominant in all. Note that not all features used are Galilean invariant, notably turbulence intensity q_{TI} and actuator forcing q_F rely on a reference-frame fixed with the ground. As such, the developed models cannot be

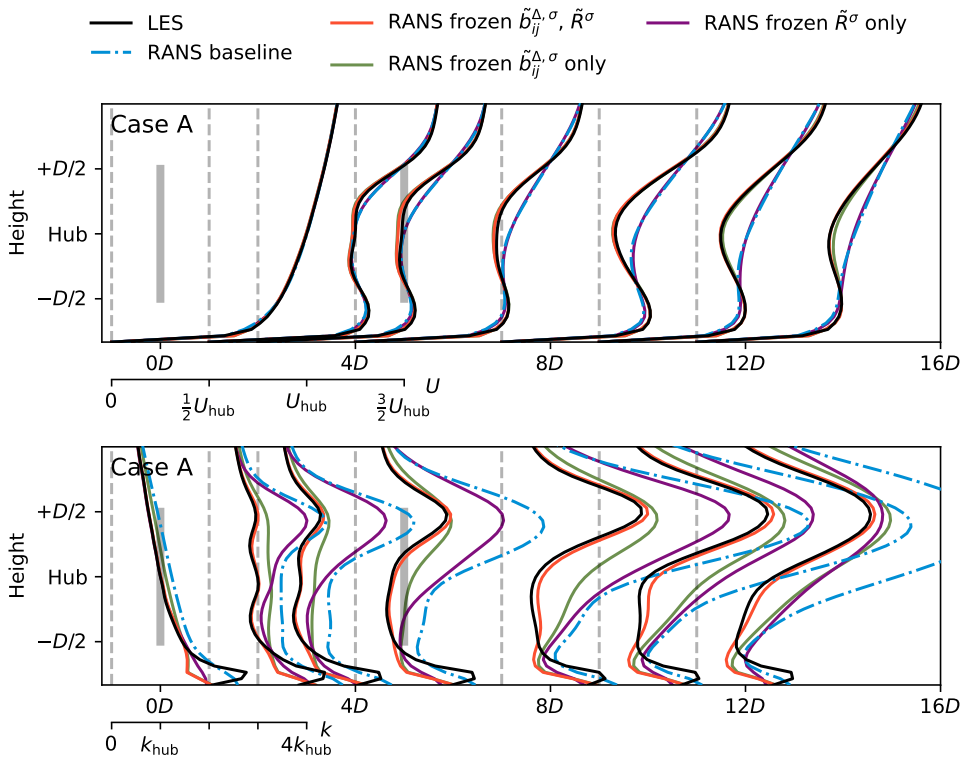


Figure 4.13: Vertical velocity and tke profiles for Case A; positions of the two turbines, and the profile locations are shown in gray. Comparison of LES; RANS baseline; RANS with injected frozen corrections; and RANS with injection of $\tilde{R}^{\sigma}$ and $\tilde{b}_{ij}^{\Delta,\sigma}$ separately.

considered general-purpose, but specific to wind farms.

In order to investigate the effect of the classifier models in isolation from the correction models, all five were implemented in the RANS solver in combination with the frozen correction terms. In particular, the corrections implemented were:

$$R(\mathbf{I},\mathbf{q}) := \sigma_l(\mathbf{I},\mathbf{q}) \cdot \tilde{R}, \qquad b_{ij}^{\Delta}(\mathbf{I},\mathbf{q}) := \sigma_l(\mathbf{I},\mathbf{q}) \cdot \tilde{b}_{ij}^{\Delta}, \qquad l \in \{1,\ldots,5\},$$

and the classifier was updated at every iteration of the flow solver until the system converged.

Figure 4.15 shows tke profiles and the classifier fields (velocity profiles are almost identical for all classifiers and are not shown in the figure). There is minimal variation in the tke profiles, except close to the wall. The classifier values themselves show significant spread in the bottom part of the wake and towards the wall which does not seem to affect the mean fields significantly. This could be a consequence of the corrections being generally small in the lower part of the wake. On this basis, classifier σ_1 is considered the most promising candidate as it is the simplest, with a single term. We also retain σ_3, of intermediate complexity and slightly higher accuracy.

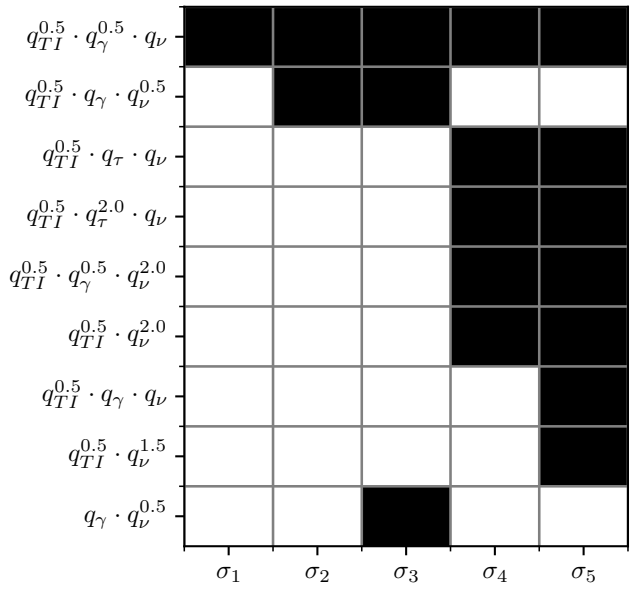


Figure 4.14: Visualization of non-zero terms in the five discovered classifier models, σ_1 to σ_5. Model coefficients are not shown.

ID	Description	Expression	Normalization
q_γ	Shear parameter	$\left\|\frac{\partial U_i}{\partial x_j}\right\|$	$\frac{\varepsilon}{k}$
q_τ	Ratio of total to normal Reynolds stresses	$\|\|\overline{u_i' u_j'}_{\text{Boussinesq}}\|\|$	k
q_ν	Viscosity ratio	ν_t	100ν
q_{TI}	Turbulence intensity	k	$\frac{1}{2}U_i U_i$
q_F	Actuator forcing	$\|F_{cell}\|$	$\frac{1}{2}\rho_0 A_{cell}\|U\|^2$
$q_\perp$	Nonorthogonality of U and ∇U	$\|U_i U_j \frac{\partial U_i}{\partial x_j}\|$	$\sqrt{U_l U_l U_i \frac{\partial U_i}{\partial x_j} U_k \frac{\partial U_k}{\partial x_j}}$
I_1	–	$\text{tr}\mathbf{S}^2$	–
I_2	–	$\text{tr}\mathbf{\Omega}^2$	–
I_{19}	–	$\text{tr}\mathbf{\Omega}\mathbf{A}_k\mathbf{S}^2$	–
I_{25}	–	$\text{tr}\mathbf{A}_k^2\mathbf{S}\mathbf{\Omega}\mathbf{S}^2$	–
I_{35}	–	$\text{tr}\mathbf{A}_p\mathbf{A}_k\mathbf{S}^2$	–

Table 4.4: Non-dimensional features used in the discovered models of the classifiers and corrective fields. Other features not listed, while included in the model discovery process, were ultimately not part of any model.

Note that the classifiers themselves might not be considered a particularly good fit to the classification target. There are two main reasons for this:

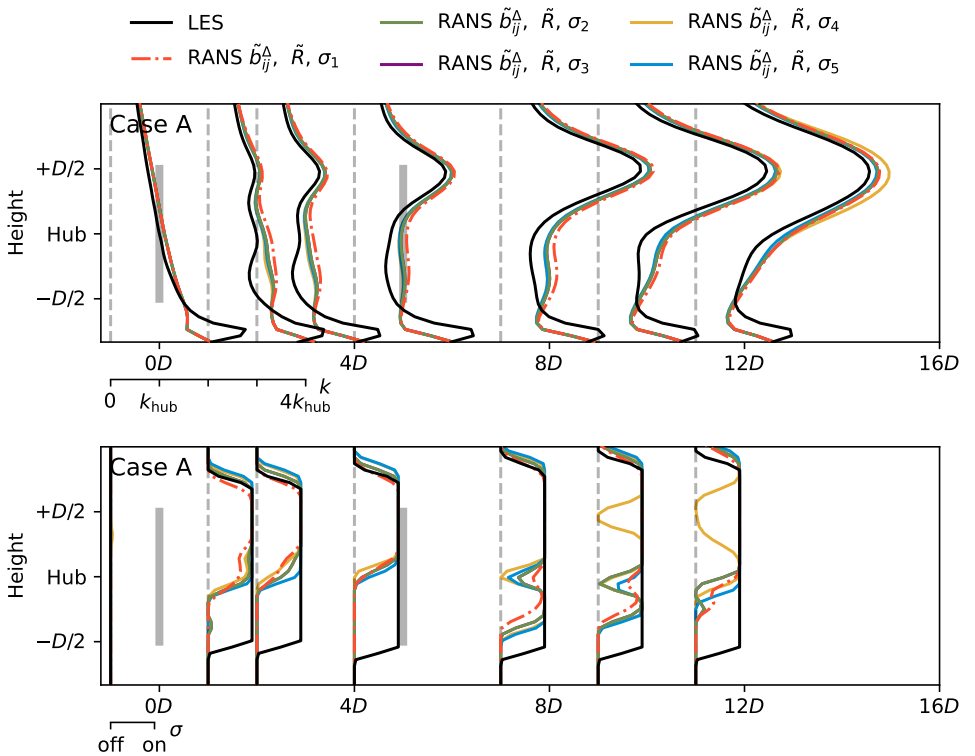


Figure 4.15: LES; RANS baseline; and corrected RANS simulations of Case A. Correction terms are frozen and classifiers are coupled with the solver. The figures depict vertical slices the tke (top) and classifier (bottom).

1. The classification target contains global information such as transport within the flow, whereas the classifiers are a function of local features only. As such there are regions in the lower part of the wake that cannot be effectively distinguished from regions in the upstream boundary layer. Thus the classification problem is harder than it appears.

2. The objective is to obtain good models that are as simple as possible, i.e. a better fit (of the classifier or the corrections) must be traded-off with model complexity. Dramatically increasing the number of terms involved in the symbolic classifier did not significantly improve the fit, and was therefore not considered. It is perhaps the case that by using a random-forest, ANN or other very highly parameterized model, a better fit may be possible – but at the cost of significant model complexity.

4.2.3. TRAINING OF CORRECTION MODELS

The correction models are now trained for regions of the flow where the classifier is active. The classifier is always trained first (see previous section), and used to discard

points from the training data-set. As such the correction model only has to reproduce the frozen correction where the classifier is active, potentially allowing for a better fit with simpler models. Experiments with σ_1 and σ_3 showed no benefits of using the more complex σ_3, therefore, all models here are trained with data selected using σ_1.

ANISOTROPY CORRECTION

All features were used in the model discovery phase resulting in a large number of candidate models. The procedures described in Section 3.4 was used to select a small set of promising models, and in addition we preferentially selected models without $T^{(2)}, \dots, T^{(4)}$ terms – i.e. linear eddy-viscosity models. Table 4.4 lists the features that were ultimately part of the resulting models.

Figure 4.16 visualizes the nonzero terms of the selected model formulations for the anisotropy correction term. We denote the models b_1^Δ to b_6^Δ in order of increasing model complexity, which ranges from 6 to 15 terms. Models b_1^Δ to b_4^Δ use only $T^{(1)}$ – making them *linear* EVMs – whereas b_5^Δ and b_6^Δ also use higher-degree base tensors, so we refer to them as non-linear EVMs. Four terms are used by all models (the first four rows of Figure 4.16): these use $T^{(1)}$ combined with q_{TI} and q_F, and/or the invariant I_1. In the remaining terms, the physical features q_ν and $q_\perp$ are most often represented, followed by the invariant I_2. The most frequently used feature overall is I_1. For the nonlinear eddy viscosity models a large overlap between the terms with nonlinear tensors is seen. The reference SpaRTA model without classifier is shown in the figure as "ref", and can be seem to be significantly more complex.

We assess the models for b_{ij}^Δ independently of the classifier and R models, by using the frozen $\tilde{R}$ and $\tilde{\sigma}$, and coupling only the model for b_{ij}^Δ. Figure 4.17 shows the effect of these partially coupled models for Case A. Again, the spread between the models is significantly larger for the k profiles than for the velocity profiles. Further, there is no significant spread between the models for the first turbine's wake, while the models differ for the second turbine – though all represent a significant improvement over the baseline model. Remarkably, it is the simplest linear model b_1^Δ and the simplest nonlinear model b_6^Δ which yield the most consistent improvement over the baseline model, with the latter having a slight edge in the wake of T2.

TURBULENT KINETIC ENERGY EQUATION CORRECTION

The nonzero terms of the correction models discovered for R are shown in Figure 4.18, and in this case the models are denoted $R_1, \dots, R_7$. The first observation is that terms describing a correction of the dissipation rate ε dominate these models, compared to terms describing a modification of production. Two terms are used in all models (the first two rows of Figure 4.18): (i) a modified linear production and (ii) a dissipation associated with actuator forcing. When compared to the anisotropy correction models, the tke production correction models use a wider range of features, most likely because this correction term is more complex and not as strongly related to velocity shear.

Once more, in Figure 4.19 we visualize the effect of coupling these models with the flow solver using frozen $\tilde{b}_{ij}^\Delta$ and $\tilde{\sigma}$. The models are not distinguished in the plot since their predictions are largely consistent with each other, with very little spread. Indeed, since this correction term mainly affects the turbulent kinetic energy, there is no visible

difference in the velocity profiles. There is some spread in the turbulent kinetic energy profiles which is largest in the near wake of the second turbine. The most complex model yields the most consistent improvement over the baseline model, but the differences between models is not significant, so again we select the simplest model for further investigation.

4.2.4. ROBUSTNESS OF CORRECTION TERMS

We observe the introduction of the classifier makes the data-driven SpaRTA models more prone to instabilities. NLEVMs are typically less stable than linear EVMs, and models produced by SpaRTA are no exception. However the classifier exacerbates this problem. This is probably because models derived in conjunction with classification are not required to be zero in non-wake regions. As such the models have become more sensitive to changes in the input features and tensors.

The instabilities manifest themselves as divergence of the anisotropy correction in the near wake of the turbines, and the R correction close to the rotor disk. Both manifestations are based on the same underlying effects:

1. The baseline $k-\epsilon$ model tends to over-predict k in the wake, and the actuator disc model does not remove turbulence kinetic energy from the flow. As a result, in the near-wake, the correction terms act mainly to remove energy.

2. The discovered models depend on the shear strain invariant I_1, which increases in magnitude as shear increases.

As a consequence, a model removing too much energy in the near wake results in larger velocity gradients there (not seen during training), which leads to larger I_1 and increasing amounts of energy removed.

To break this positive feedback loop, in as flow-agnostic a way as possible, two limiters are proposed for the two corrective terms:

- **Eddy viscosity limiter:** Inspired by the k-ε-f_P model [47], the linear components of the present anisotropy models were limited to a proportion of Boussinesq:

$$\alpha_1 = \min\left(0.8 \cdot \frac{\varepsilon}{k^2} \cdot \nu_t, \alpha_1\right). \tag{4.3}$$

- **Form error limiter:** This addresses the tendency of the correction models to aggressively remove energy near the actuator discs. This limiter is only active in areas where actuator forcing is applied, and is chosen based on the Boussinesq turbulent kinetic energy production as:

$$R = \mathrm{sgn}(R) \cdot \min\left(0.5 P_k^{\mathrm{Boussinesq}}, |R|\right). \tag{4.4}$$

The thresholds were derived based on an analysis of the available data-set and are chosen sufficiently low that limiters are only active during the convergence of the solver.

4.2.5. PREDICTIVE SIMULATIONS

Any meaningful test of these models must be in a predictive setting on unseen cases, in this work Cases B & C. Based on the partially coupled results of Sections 4.2.2 and 4.2.3, models composed of the following components are selected:

- the simplest σ_1,
- the simplest linear b_1^Δ and nonlinear b_5^Δ models,
- the simplest R_1 and medium complexity R_4.

All four model combinations were implemented in a fully coupled manner in the simulation code, and predictions compared with LES data for Cases B & C. Figure 4.20 shows the predictions for Case C. Additional figures for Case B (and Case A, not predictive) are presented in Appendix B.2 and give broadly similar results. All the results, except those from the simplest model, are not distinguished in the figure, for the benefit of visualization.

Examining the predictions, firstly all the correction models yielded a significant improvement over the baseline $k-\varepsilon$ model. The variability in the predictions of the four models did not increase significantly beyond the variability observed in the partially coupled results. This suggests that there is no apparent strong interaction between the two correction terms which is reassuring. Given the similar performance of all four models, we again use our bias for parsimony to select a single "best" model consisting of b_1^Δ, R_1 and σ_1. Explicitly this model is:

$$\begin{aligned} b_1^\Delta := \; & [1.62\cdot 10^{-1}\cdot q_{TI}^{1/2}\cdot q_F^{1/2} \\ & +4.84\cdot 10^{-3}\cdot q_{TI}^{1/2}\cdot I_1^{1/2} \\ & -1.90\cdot 10^{-11}\cdot q_{TI}\cdot I_1^4 \\ & +2.51\cdot 10^{-2}\cdot q_F^{1/2} \\ & +2.00\cdot 10^{-3}\cdot I_1^{1/2} \\ & +1.49\cdot 10^{-15}\cdot I_1^{9/2}]\cdot T_{ij}^{(1)} \end{aligned} \tag{4.5a}$$

$$\begin{aligned} R_1 := \; & 8.06\cdot 10^{-5}\cdot I_1^{1/2}\cdot q_v^3\cdot k\cdot T_{ij}^{(1)}\frac{\partial u_i}{\partial x_j}+ \\ & [-2.91\cdot 10^{1}\cdot q_{TI}^{1/2}\cdot q_F\cdot I_1^{1/2} \\ & +4.28\cdot 10^{-1}\cdot q_\perp^2\cdot q_F\cdot I_1^{1/2} \\ & -1.22\cdot q_F\cdot q_\gamma \\ & +2.30\cdot q_F^2\cdot I_2]\cdot\epsilon \end{aligned} \tag{4.5b}$$

$$\sigma_1 := 1/\left(1+\exp\left(-205.041112\cdot q_{TI}^{1/2}\cdot q_\gamma^{1/2}\cdot q_v+9.01862802\right)\right) \tag{4.5c}$$

The magnitude of the coefficients of the terms can be misleading, because the range of magnitude of the features is quite large. For example, although the first term of R_1 has a small coefficient, but it is one of the largest terms in the near wake. Similarly the tiny

coefficient of the last term of b_1^Δ is counteracted by the high power of I_1, leading to this term having an effect in regions of high shear. Of course it is to be supposed that such high powers and small coefficients may not generalize to other cases well – although they generalize to our Cases B & C here.

The ability of our procedure to produce many diverse models, all with similar performance, suggests on the one hand that procedure is reasonably robust, but on the other hand that perhaps the model space is not being sufficiently explored. For example none of the discovered models come close to matching the frozen correction results (which themselves match LES well), even for the training Case A. It would be expected that with sufficiently many terms, an overfit model for Case A could be developed – but this was not achievable with less than 50 terms. There may be a case to be made for very highly parameterized models such as neural networks in this context.

In these results, the eddy viscosity limiter was active on average in about 1500 cells mainly in the upper part of the near wake, and the form error limiter was active on average in about 4000 cells mainly in the center of the rotor disk. Given that the total number of cells in the domain is around 3 million, the limiter is seldom used, which is deemed acceptable.

COMPARISON WITH CORRECTIVE MODEL WITHOUT CLASSIFIER

Figure 4.21 presents a comparison between the simplest model with classifier above, a the reference SpaRTA model discovered without classifier in [86], whose expression is given in equations 4.1 and 4.2. In both cases the training data is Case A only, and prediction is for Case C.

There are some minor differences between predictions of the reference and the model with classifier, although overall the reference model performs slightly better. However, the reference model contains a total of 36 terms, 25 for b_{ij}^Δ and 11 for R, including terms involving $T^{(2)}, \ldots, T^{(4)}$. On the other hand, our simple classifier model of (4.5) contains only 12 terms in total – including the classifier, and does not exploit base tensors beyond $T^{(1)}$. If our premise is accepted, that – all else being equal – simple models should be preferred to complex ones, then the use of the classifier has been seen to have the potential to simplify the resulting models for similar predictive accuracy.

Admittedly, it is indeed the case that without classifiers, we were able to find models that worked without limiters, and that here (with classifiers) that was not possible. A possible reason for this is as follows. When training corrections without a classifier, the correction model is encouraged to predict zero, or very small, corrections in large regions of the flow. As a result, it is often less aggressive everywhere. When training with a classifier, all regions of small correction have been filtered out by the classifier, and the corrector only has to match what remains. It can therefore be more aggressive, and matches the correction better with fewer terms. This aggressive fitting is what leads to the instabilities we observe. We expect to see a similar correspondence between goodness-of-fit and stability whenever the power of the fitting method increases – as seen in other work with random forests [39], which also fit training corrections very well, and are often unstable in prediction.

It is worth noting that, in general, issues of instability are not restricted to the present classifier-correction framework. In our experience, they are a problem for data-driven

RANS modeling in general and the SpaRTA approach without classifier in particular. Although it was possible to find stable SpaRTA models, by selecting them from the multiple models we generated, the more complex the cases studied, and in particular the more significant the corrections needed, the more difficult it becomes to find stable models. This is one of the main motivating factors behind so-called CFD-consistent modeling approaches [32, 98], in which the CFD code is inside the training loop, automatically making unstable models unfit. Our work attempts to achieve as much as possible *without* CFD-in-the-loop [110] in order to build methods that are more scalable to expensive simulations and work without adjoints.

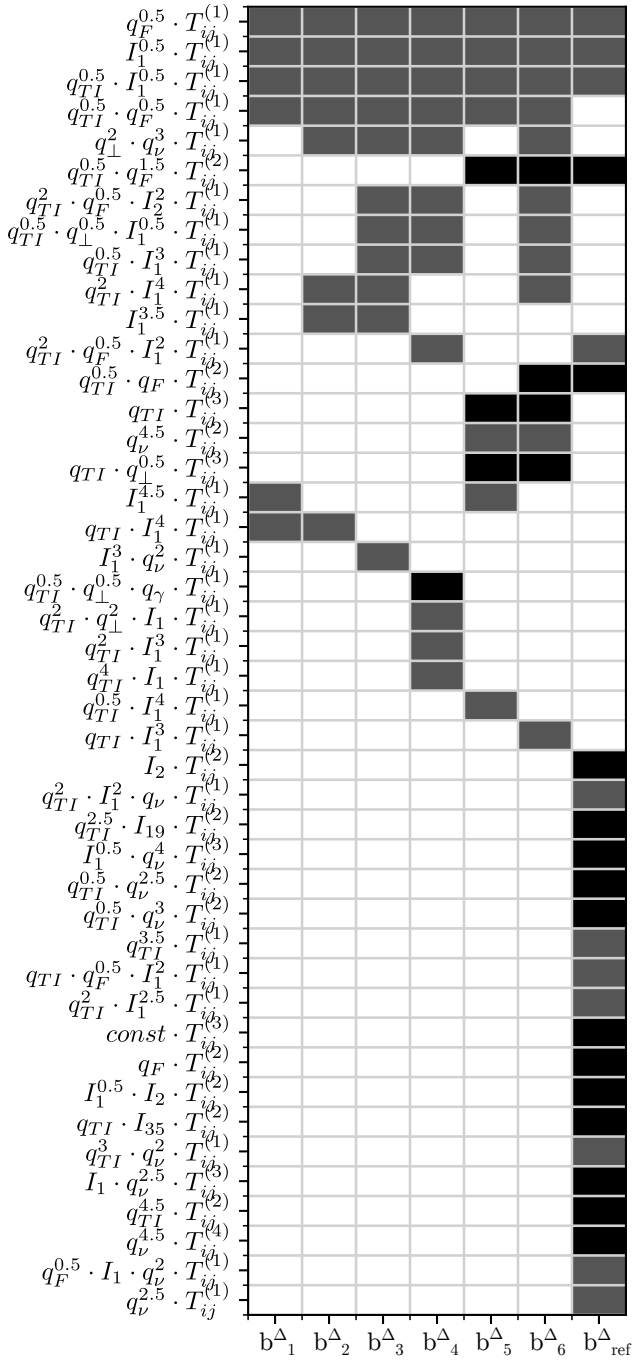


Figure 4.16: Visualization of non-zero terms in the six discovered anisotropy-correction models, b_1^Δ to b_5^Δ, as well as b_{ref}^Δ from [86]. Terms based on $T^{(1)}$ only are gray, while terms involving $T^{(2-4)}$ are black. The model coefficients are not shown.

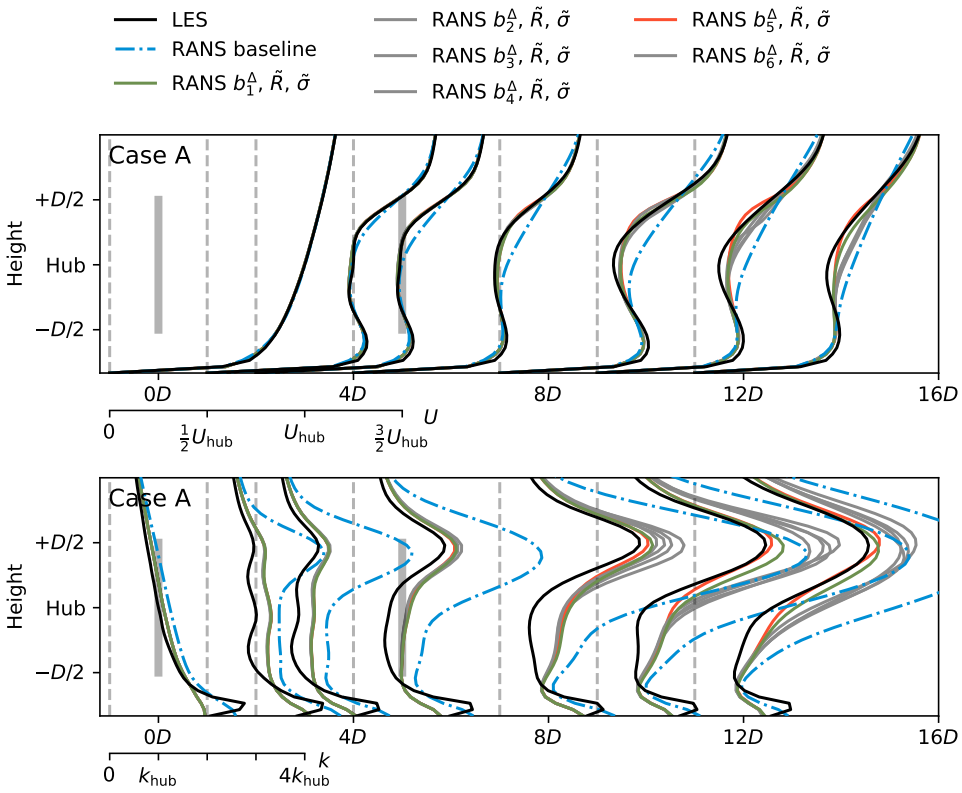


Figure 4.17: LES, RANS baseline, and RANS coupled with different anisotropy correction models for Case A. The tke production correction $\tilde{R}$ and the classifier $\tilde{\sigma}$ are frozen. Only two of the six b^Δ models are distinguished with color; the remaining are all gray (not distinguished) in order to visualize the spread of predictions.

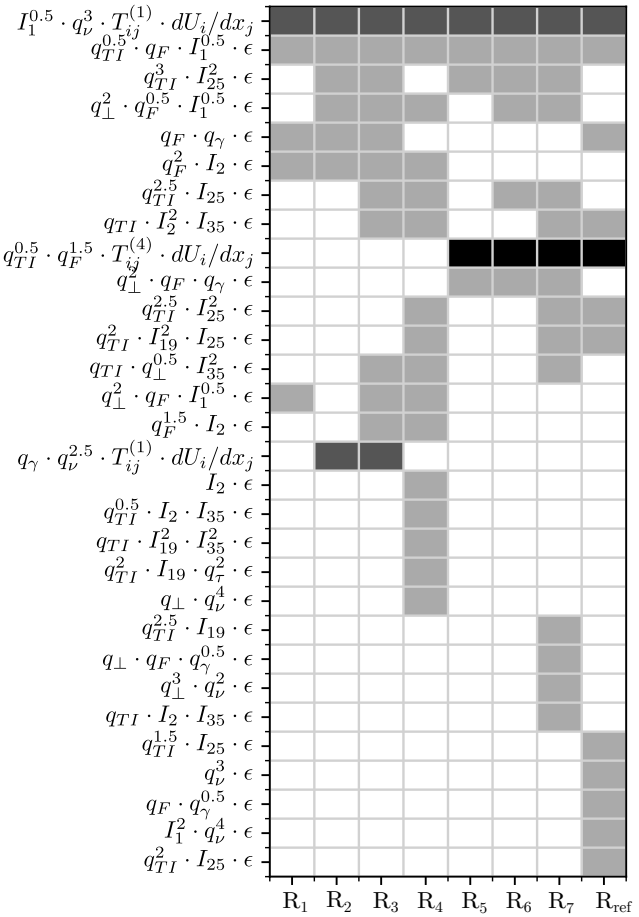


Figure 4.18: Visualization of non-zero terms in the seven discovered R models, R_1 to R_7. Light-gray indicates a dissipation-rate correction; gray indicates the use of $T^{(1)}$; and black the use of $T^{(2-4)}$.

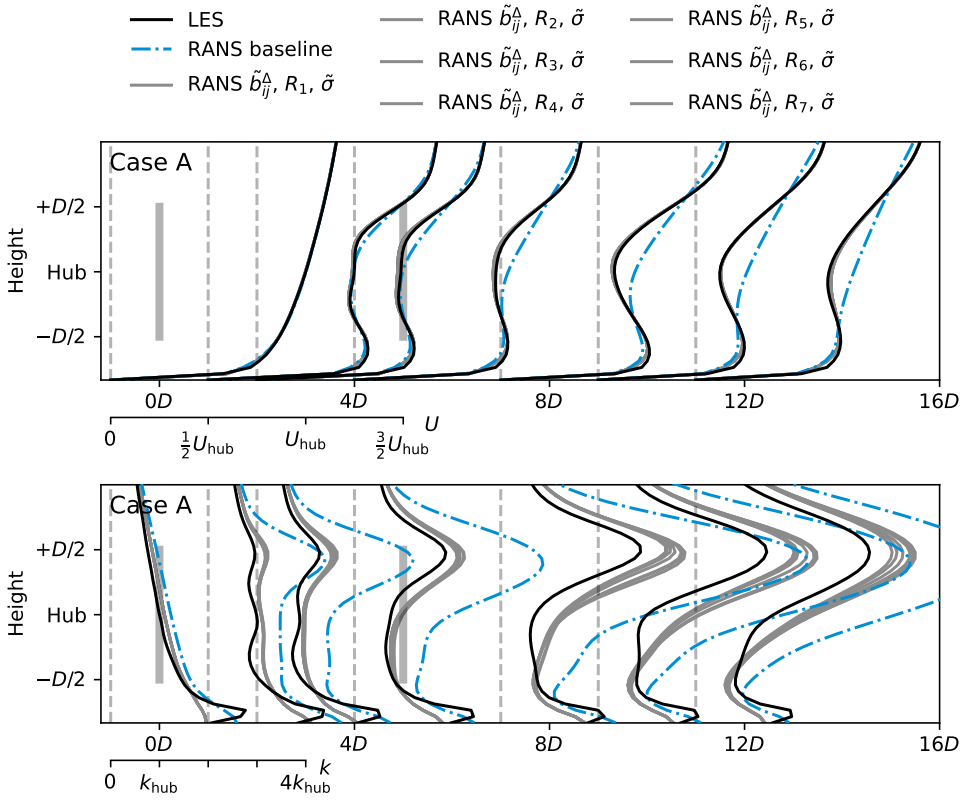


Figure 4.19: LES, RANS baseline, and RANS coupled with seven different models for R for Case A. The tke production correction $\tilde{R}$ and the classifier $\tilde{\sigma}$ are frozen. The seven R models are all gray (not distinguished) in order to visualize the spread of predictions.

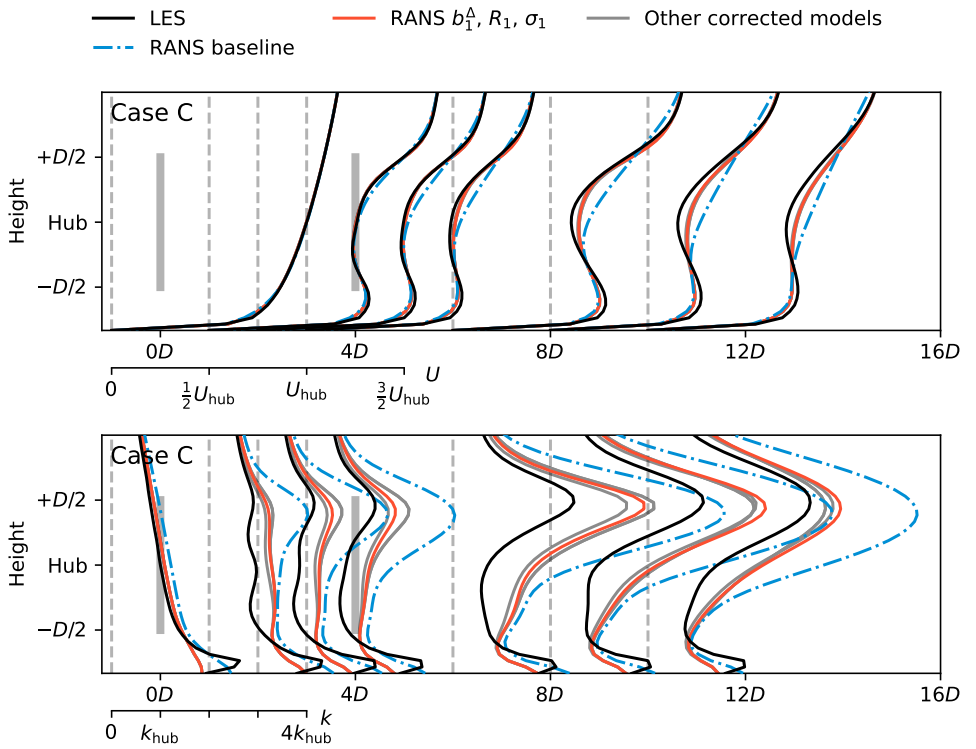


Figure 4.20: LES, RANS baseline, and SpaRTA RANS models for Case C.

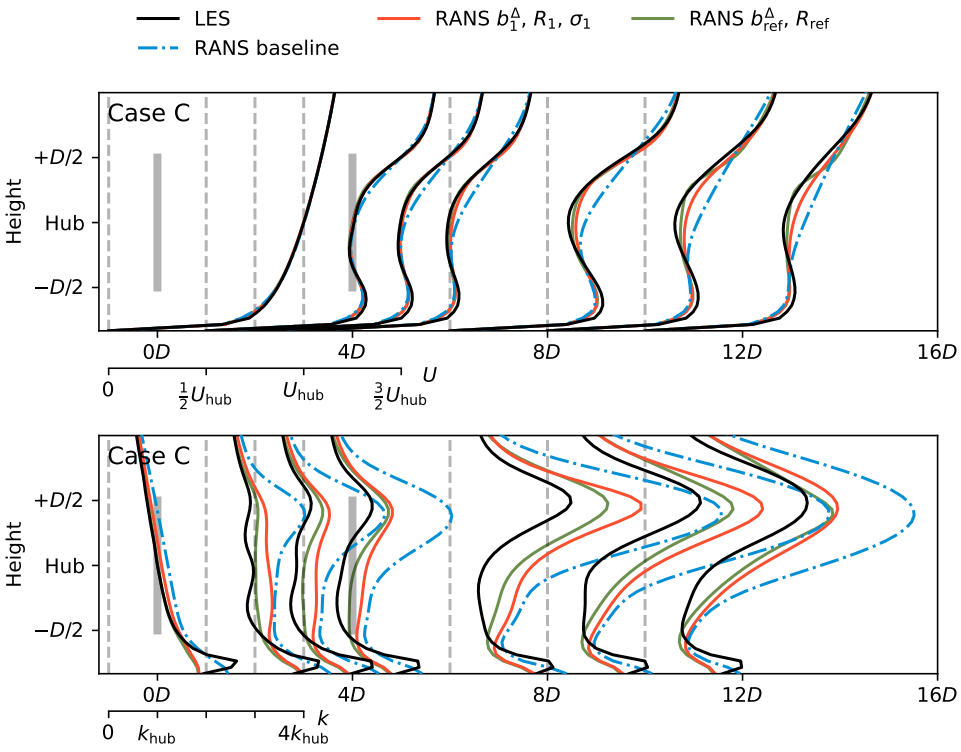


Figure 4.21: LES, RANS baseline, SpaRTA model with classifier, and a reference SpaRTA model without classifier for Case C.

/

5

CONCLUSIONS

This work uses a methodology combining a k-corrective frozen-RANS approach with sparse symbolic regression to derive data-driven RANS turbulence models for flow around scaled wind turbines. The resulting models give significantly better predictions than the baseline k-ε model for both velocity and turbulent kinetic energy (tke) in the near and far wake of the turbines, with an almost perfect match between reference LES and corrected RANS data. The dataset includes three test cases with between two and three turbines. The training was done on test case A and then validated on the training cases B and C.

The approach was initially developed by Schmelzer et al. [79] - who referred to it as SpaRTA - and applied it to 2D test cases at moderate Reynolds numbers. For this approach to become useable for wind energy applications, it needed some extensions: (i) a more comprehensive feature set that includes physical features such as normalized actuator forcing; using only invariants of Pope's extended integrity basis was found to be inadequate, (ii) the introduction of blending functions to allow integration with wall functions and suppression of corrections at the wall, (iii) the splitting of corrections into ABL and wake corrections, and (iv) algorithmic developments for the sparse regression procedure to allow the processing of larger datasets.

The field of data-driven turbulence modeling is rather young. It is a reaction to the field of traditional turbulence modeling, where further improvements to RANS for industrial applications have been limited. With the boom of artificial intelligence in general, there have been many recent examples of data-driven turbulence modeling as summarized by Duraisamy et al. [18]. However, most of these cases consider 2D benchmark cases with low to moderate Reynolds numbers. In this work, results from 3D cases with $\mathrm{Re} \approx 100,000$ were presented for more industrially relevant cases. Furthermore, the different case settings considered show what improvements are necessary for currently existing algorithms to take the field of data-driven modeling to industrial practice. Finally, the work also combines corrections models with a classifier. There are no examples of this in the literature.

DISCUSSION

The remaining methodology of this work will now be assessed to draw conclusions and highlight shortcomings in each step of the process.

Summary of the Frozen Approach: The k-corrective frozen formulation relies on an anisotropy correction term $\tilde{b}_{ij}^{\Delta}$ and a source correction term $\tilde{R}$. Expressions for these terms are obtained by solving the k-ε turbulence model transport equations using frozen time-averaged LES data for the velocity and tke. The correction terms are included in the equations as follows:

$$\frac{Dk^{\star}}{Dt} = \mathscr{P}_k^{\star} + \tilde{R} - \varepsilon + \frac{\partial}{\partial x_j}\left[(\nu + \nu_t/\sigma_k)\frac{\partial k^{\star}}{\partial x_j}\right], \tag{5.1}$$

$$\frac{D\varepsilon}{Dt} = \left[C_{\varepsilon 1}\left(\mathscr{P}_k^{\star} + \tilde{R}\right) - C_{\varepsilon 2}\varepsilon\right]\cdot\frac{\varepsilon}{k^{\star}} + \frac{\partial}{\partial x_j}\left[(\nu + \nu_t/\sigma_{\varepsilon})\frac{\partial \varepsilon}{\partial x_j}\right] \tag{5.2}$$

where the production term is defined as

$$\mathscr{P}_k^{\star} := 2k^{\star} b_{ij}^{\star} S_{ij} \tag{5.3}$$

with

$$b_{ij}^{\star} := \frac{\tau_{ij}^{\star}}{2k^{\star}} - \frac{1}{3}\delta_{ij} = -\frac{\nu_t}{k^{\star}} S_{ij}^{\star} + \tilde{b}_{ij}^{\Delta}. \tag{5.4}$$

The k-corrective frozen RANS approach is numerically stable, and the propagation of the obtained correction terms in a baseline simulation leads to a very close match between reference LES and corrected RANS simulation except near the wall, where some issues must be appropriately managed.

Making sure that the boundary conditions for the velocity and the tke, as well as the domain forcing, match between RANS and LES is essential for matching the undisturbed inflow profiles (including the ABL corrections). An analysis of the wall boundary conditions for LES and RANS was carried out, and to perfectly match them, the constant C_{μ} had to be modified to a non-standard value. Because these profiles did not match away from the wall, horizontally constant corrections were also introduced to the inflow profiles. Both the LES and RANS also showed an unnaturally large peak in the tke near the wall, which is a numerical artifact that is difficult to remove but does not significantly influence the profile in the rotor area [10]. Wall blending was therefore introduced such that no corrections were applied in the cells closest to the wall, as the corrections were not accurate there due to the numerical artifact.

While this procedure for matching the undisturbed profiles from the precursor worked, it could potentially be simpler. Not trying to establish a perfect match from the wall up and just trying to get a good match in the rotor area only could be a better solution. This could be achieved by modifying the forcing and wall roughness. If the said matched profile is not accurate enough, then again corrections could be introduced for the undisturbed profiles as well. However, it would be more difficult to determine those corrections due to the deviations at the wall.

Once the undisturbed profiles were matched, the injection of the corrections for the wakes was found to be much less complicated. Selective injection of the wake source and anisotropy correction lead to some interesting conclusions. If only the anisotropy correction was injected, then the velocity profiles matched well, but the tke fields did not match completely. The source term R was then also injected and both velocity and tke fields matched well. This indicates that modeling approaches focusing solely on improving the anisotropy term will not succeed in getting a good match for both velocity and tke fields. The source term can account for the direct interaction between the actuator forcing, and the turbulence model - the rotor will extract energy from the tke as well - the anisotropy correction cannot. However, the source term is not only large in areas where there is actuator forcing, but also further downstream in the wake. One possible explanation is that it captures hysteresis effects in the flow. It would thus be interesting to explore the frozen approach in combination with a lag turbulence model [52]. A lag model solves an additional transport equation called the lag equation which is supposed to account for non-equilibrium effects. Preliminary results from this approach have been obtained internally [65].

Summary of the Learning Procedure: The correction terms are formulated using a combination of an extension of Pope's integrity basis and physical features. Sparse regression is used to determine the scalars α_n, β_n, and γ as a function of the input features. The tensors $T_{ij}^{(n)}$ that are used for the regression are determined a priori. The first of these is always the strain rate tensor $T_{ij}^{(n)} = S_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$. The correction tensors are then formulated as

$$b_{ij}^{\Delta} = \sum_n \alpha_n T_{ij}^{(n)}$$

and

$$R = \sum_n \beta_n T_{ij}^{(n)} \frac{\partial u_i}{\partial x_j} + \gamma \cdot \varepsilon.$$

The learning of the frozen terms using the elastic net (essentially a least-squares fit with regularization to penalize complex and redundant models) required modifications to Schmelzer's original setup. Normally, the input features are combined to give a list of basis functions. The regression procedure then delivers a list of models that are a linear combination of the input basis functions. Using only the first few invariants of Pope's integrity basis as input features did not yield good results. The inclusion of more invariants and tensors from the extended integrity basis was attempted, but the input feature set became too big resulting in several thousand basis functions. Two preprocessing steps were introduced to remedy this: (i) a mutual information step to reduce the input feature set and (ii) a cliqueing step to remove colinear basis functions. These are common preprocessing steps for machine learning algorithms to reduce the dimensionality of the input feature set. This made the training process manageable but did not give significantly better models. Finally, the inclusion of physical features such as actuator forcing and a reformulation of the source term in the tke equation gave much better results.

Additionally, correction models were combined with a classifier in order to turn corrections off outside of the wake. Logistic regression was used for this purpose with the same input features as for the learning of the correction terms. This procedure simplified the learning procedure for the correction terms as the amount of data points in the input data set could be significantly reduced. It also produced simpler correction models. This indicates that in the models trained without the classifier some cancellation between terms occurs, requiring the models to be more complex.

Implementing the models obtained from the elastic net in OpenFOAM was straightforward, but the model selection was more complicated. The regression algorithm produced a large array of models. Cliqueing was applied in the model selection, and an array of models was selected based on accuracy and complexity. For some models, there was little difference between offline (or uncoupled) and online (or coupled) performance. For others, the coupling to the evolving flow field made them unstable. For the selection of the classifier formulation, the spread between proposed classifier models was much smaller, such that very simple classifier models could be used.

For most models the deviation between uncoupled and coupled values was largest in the near wake near the top of the wake. This is the area of the domain where the shear is largest (apart from the near wall area where the correction terms are blended out). Things that made the models less robust in this area - meaning a larger deviation between online and offline performance - were more complex models, higher-order invariants, a finer mesh, and combination with a classifier. For some models, a bit more energy was removed than intended. For others, too much energy was removed in a very local area leading to negative eddy viscosity and divergence of the simulation. A positive coupling loop was identified involving the velocity gradient and the corrections. This shortcoming is similar to what van der Laan [48] described for traditional nonlinear Eddy Viscosity models.

For the corrections obtained with the classifier, two limiters had to be introduced to avoid divergence: (i) an eddy viscosity limiter which ensured that the total tke production could be reduced to a maximum of 20 % of the Boussinesq production, and (ii) a form error limiter which ensured that at most 50 % of the Boussinesq turbulence production could be removed in the rotor wake area by the source term. These limiters were active only in a handful of cells but managed to stabilize the models. These limiters were also applied to the corrections that were obtained without the classifier and slightly improved the results there as well.

These results indicate that, unsurprisingly, traditional and data-driven turbulence models suffer from the same numerical instabilities. The difference is that for data-driven models, more flexibility is available in the formulation of the model terms. The author has some ideas on how this could be remedied besides the use of the limiters outlined above. First, online learning could potentially be used to avoid instabilities as the coupling loops are then included in the training process, but online learning is also much more expensive in terms of CPU time. Hence, using the results from the offline learning process and just tuning the parameters of the basis functions online could provide relief. Second, avoiding the combination of velocity gradient-based features and tensors could be helpful as this stops the positive coupling loop, although this might negatively affect accuracy. Third, the anisotropy correction term could be reformulated as

$$\frac{\tau_{ij}}{2k} - \frac{1}{3}\delta_{ij} = -\frac{\nu_t}{k}S_{ij} + b_{ij}^{\Delta} = -\frac{\nu_t}{k}S_{ij} + \frac{k}{\varepsilon}\sum_n \alpha_n T_{ij}^{(n)}$$
$$\rightarrow \frac{\tau_{ij}}{2k} - \frac{1}{3}\delta_{ij} = -\frac{\nu_t}{k}S_{ij} + b_{ij}^{\Delta} = \underbrace{\left(1 - \frac{\alpha_1}{C_\mu}\right)}_{\alpha_{new}} \cdot \frac{\nu_t}{k}S_{ij} + \frac{k}{\varepsilon}\sum_{n>1} \alpha_n T_{ij}^{(n)}, \tag{5.5}$$

where α_1 would then revert to a scaling function α_{new} for the linear part of the anisotropy tensor instead of a basis function for a tensor. This is similar to the formulation used in Wu et al. [104], in which the implicit treatment of the linear part and the explicit treatment of the nonlinear part was found to make the data-driven turbulence model more robust. Lastly, extending the data-set with data points that experience even more shear than the ones already in the data-set would most likely help a lot. The existing models remove too much energy as the velocity gradients increase due to the positive coupling loop explained before. Since the data-set does not contain any points with such high shear, then the models are used outside of the area where they were trained.

Finally, it should be noted that, the model for R contains terms resembling production (dimensioned by G_1, etc.) and terms resembling dissipation (dimensioned by ε). Arguably the former are inconsistent with the momentum equation, in the sense that they represent production which does not arise as a direct consequence of the modelled RST - unlike P_k^{Δ}, which is exactly the production resulting from the model for b_{ij}^{Δ}. Such inconsistencies have been referred to as a lack of energy conservation in the literature [19], though some successful models are also inconsistent in this respect, notably the Mentor SST model with vorticity source term [61]. For production corrections involving only G_1 the inconsistency can be fixed by modifying the eddy-viscosity [36]. Alternatively in SpaRTA, models for R could be restricted to ε terms only.

Outlook

One big shortcoming of the work presented in this dissertation was the limited dataset available. If one wants to further pursue a data-driven approach for wind energy purposes, the training dataset needs also to contain full-scale cases. Furthermore, the test dataset needs to contain more conditions in order to generalize the models as regression algorithms are prone to overfitting even when regularization is used. For example, it would be interesting to see if a wake correction is trained for one turbine only, how well will it perform when applied to multiple turbines in a row. The previously presented work from King et al. [41] suggests that such a model might already provide significant improvement.

The corrections were split into an ABL and a wake part. The wake part should also work for different ABL properties as long as C_μ is kept the same. However, the ABL part should be recalibrated for a different hub height velocity, wind direction, or to a lesser extent, surface roughness. Some preliminary work not included in this thesis on the neutral boundary layer at full scale suggests that simply tuning forcing and the surface roughness can already give a very reasonable match in the rotor area in many cases. Of course, this then comes at the cost of accuracy at the wall, further investigation would be necessary to quantify when this is a useful approximation.

Additionally, stratification has not been considered in this work. However, the work from Baungaard et al. [7] implies that at least for stably stratified flow and weakly convective flows, the effect of stratification on undisturbed ABL profiles is much more important than its effect on the development of the turbine wake. In fact, Baungaard et al. only model the effect of buoyancy on the flow through a Buoyancy term in the transport equation of the turbulence model. This Buoyancy term varies only in wall-normal direction and is formulated such that the inflow profiles for the velocity, turbulent kinetic energy and the turbulent dissipation are in accordance with MOST profiles. Other than that, the system of equation is the same as for neutral stratification. Hence for these specific conditions, it might be sufficient to neglect the interaction between wake and stratification as long as the undisturbed profiles are captured well by the (corrected) baseline model. Preliminary work from MSc students that the author supervised gave similar results [43, 42]. As stratified flows are inherently unsteady, the methodology presented in this work can not be directly applied. The mentioned MSc reports used short intervals (ten minutes to an hour) over which time-averaged results for both fields and boundary conditions were saved. Then, those fields were used for the frozen simulation; whatever time-dependent changes remained were unintentionally dumped in the source term R. The underlying idea is that the time scale of the turbulent fluctuations in the flow are much smaller than the time scales of the whole boundary layer, such that the unsteadiness due to stratification does not directly affect the modeling of Reynolds stresses due to turbulence fluctuations. This assumption is appropriate for stable and weakly convective boundary layers. For strongly unstable ABLs, the difference between time scales is less significant, making the assumption inappropriate. Corrections derived for a steady-state frozen simulation might also be applied to a URANS simulation, where the unsteadiness comes from the heat flux at the ground only.

While extending the dataset makes sense to get more relevant models, the current implementation of SpaRTA in scikit-learn is limited by the amount of memory that is necessary for building the basis function library. For running an elastic net procedure, a full matrix with dimensions (number of data points × number of linear basis functions) is built. Comparatively, for running the equivalent CFD simulations, a sparse matrix where the number of non-zero entries is (number of data points×number of equations, here five) is loaded. The number of linear basis functions obtained with the procedure presented in this thesis can be of the order of several thousand. Hence, the memory required for the full system matrix of the elastic net and/or the learning procedure can quickly become larger than the available memory even on an HPC node.

Simple methods like limiting the feature and basis function space using a feature selection algorithm, user knowledge and sub-sampling of the input dataset can all be helpful in this context. More integral or automated methods also exist: (i) bagging, where several models are trained on partitions of the dataset and the final model is an average of the models trained on the different partitions, (ii) chunking, where the dataset is also divided into partitions and then the model is trained sequentially on the different partitions, and (iii) batch gradient descent methods where the gradient is calculated several times on a subset of the data points per iteration of the algorithm. While these methods reduce memory requirements they require more computational steps, so, as always, a balance has to be struck between memory and power usage. Alternatively, other ma-

chine learning algorithms that do not suffer from these restrictions could be considered, but all algorithms come with their own drawbacks.

Overall, while data-driven methods have potential to improve turbulence modeling for wind turbines wakes, given the challenges outlined previously, human judgement will likely remain important in the selection of turbulence models, feature sets and data for quite some time.

BIBLIOGRAPHY

[1] M. Abkar and F. Porté-Agel. "Influence of atmospheric stability on wind-turbine wakes: A large-eddy simulation study". In: *Physics of Fluids* 27.3 (2015), p. 035104. DOI: 10.1063/1.4913695.

[2] C. Adcock and R. N. King. "Data-Driven Wind Farm Optimization Incorporating Effects of Turbulence Intensity". In: June 2018, pp. 695–700. DOI: 10.23919/ACC.2018.8431727.

[3] R. D. Alba. "A graph-theoretic definition of a sociometric clique". In: *The Journal of Mathematical Sociology* 3.1 (1973), pp. 113–126. DOI: 10.1080/0022250X.1973.9989826.

[4] J. D. Albertson and M. B. Parlange. "Natural integration of scalar fluxes from complex terrain". In: *Advances in Water Resources* 23.3 (1999), pp. 239–252. ISSN: 0309-1708. DOI: 10.1016/S0309-1708(99)00011-1.

[5] E. G. A. Antonini, D. A. Romero, and C. H. Amon. "Improving CFD wind farm simulations incorporating wind direction uncertainty". In: *Renewable Energy* 133 (2019), pp. 1011–1023. ISSN: 0960-1481. DOI: 10.1016/j.renene.2018.10.084.

[6] R. J. Barthelmie and L. E. Jensen. "Evaluation of wind farm efficiency and wind turbine wakes at the Nysted offshore wind farm". In: *Wind Energy* 13.6 (2010), pp. 573–586. DOI: https://doi.org/10.1002/we.408.

[7] M. Baungaard, M. P. van der Laan, and M. Kelly. "RANS modeling of a single wind turbine wake in the unstable surface layer". In: *Wind Energy Science* 7.2 (2022), pp. 783–800. DOI: 10.5194/wes-7-783-2022.

[8] J. Bleeg et al. "Modeling stable thermal stratification and its impact on wind flow over topography". In: *Wind Energy* 18 (2015), pp. 369–383. DOI: 10.1002/we.1692.

[9] Sanjeeb T. Bose and George Ilhwan Park. "Wall-Modeled Large-Eddy Simulation for Complex Turbulent Flows". In: *Annual Review of Fluid Mechanics* 50.1 (2018), pp. 535–561. DOI: 10.1146/annurev-fluid-122316-045241.

[10] J. G. Brasseur and T. Wei. "Designing large-eddy simulation of the turbulent boundary layer to capture law-of-the-wall scaling". In: *Physics of Fluids* 22.2 (2010), p. 021303. DOI: 10.1063/1.3319073.

[11] S.P. Breton et al. "A survey of modelling methods for high-fidelity wind farm simulations using large eddy simulation". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 375.2091 (2017), p. 20160097. DOI: 10.1098/rsta.2016.0097.

[12] D. Cabezon, E. Migoya, and A. Crespo. “Comparison of turbulence models for the computational fluid dynamics simulations of wind turbine wakesin the atmospheric boundary layer”. In: *Wind Energy* (2011). DOI: 10.1002/we.516.

[13] M. Calaf, C. Meneveau, and J. Meyers. “Large eddy simulation study of fully developed wind-turbine array boundary layers”. In: *Physics of Fluids* 22.1 (2010), p. 015110. DOI: 10.1063/1.3291077.

[14] L. P. Chamorro and F. Porté-Agel. “Effects of Thermal Stability and Incoming Boundary-Layer Flow Characteristics on Wind-Turbine Wakes: A Wind-Tunnel Study”. In: *Boundary-Layer Meteorology* 136.3 (Sept. 2010), pp. 515–533. ISSN: 1573-1472. DOI: 10.1007/s10546-010-9512-1.

[15] M. Churchfield and S. Lee. *Simulator for wind farm aerodynamics (SOWFA)*. https://nwtc.nrel.gov/SOWFA. Feb. 2019. URL: https://nwtc.nrel.gov/SOWFA.

[16] A. Crespo, J. Hernández, and S. Frandsen. “Survey of modelling methods for wind turbine wakes and wind farms”. In: *Wind Energy* 2.1 (1999), pp. 1–24. DOI: 10.1002/(SICI)1099-1824(199901/03)2:1<1::AID-WE16>3.0.CO;2-7.

[17] F. Ducros, Nicoud Franck, and Thierry Poinsot. “Wall-Adapting Local Eddy-Viscosity Models for Simulations in Complex Geometries”. In: *Numerical Methods for Fluid Dynamics VI* (Jan. 1998).

[18] K. Duraisamy, G. Iaccarino, and H. Xiao. “Turbulence Modeling in the Age of Data”. In: *Annual Review of Fluid Mechanics* 51.1 (2019), pp. 357–377. DOI: 10.1146/annurev-fluid-010518-040547.

[19] P. Durbin. *Advanced Approaches in Turbulence: Theory, Modeling, Simulation, and Data Analysis for Turbulent Flows*. 2021. DOI: https://doi.org/10.1016/C2019-0-01450-1.

[20] P.A. Durbin. “On the k-ε stagnation point anomaly”. In: *International Journal of Heat and Fluid Flow* 17.1 (1996), pp. 89–90. ISSN: 0142-727X. DOI: https://doi.org/10.1016/0142-727X(95)00073-Y.

[21] W. N. Edeling, G. Iaccarino, and P. Cinnella. “Data-Free and Data-Driven RANS Predictions with Quantified Uncertainty”. In: *Flow, Turbulence and Combustion* 100.3 (Apr. 2018), pp. 593–616. ISSN: 1573-1987. DOI: 10.1007/s10494-017-9870-6.

[22] W.N. Edeling et al. “Bayesian estimates of parameter variability in the k–ϵ turbulence model”. In: *Journal of Computational Physics* 258 (2014), pp. 73–94. ISSN: 0021-9991. DOI: https://doi.org/10.1016/j.jcp.2013.10.027.

[23] W.N. Edeling et al. “Bayesian estimates of parameter variability in the k–ϵ turbulence model”. In: *Journal of Computational Physics* 258 (2014), pp. 73–94. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2013.10.027.

[24] A. Eidi et al. “Data-driven quantification of model-form uncertainty in Reynolds-averaged simulations of wind farms”. In: (May 2022). DOI: 10.48550/arXiv.2205.14221.

[25] Bundesamt für Energie. *Neue Studie: Das Windenergiepotenzial in der Schweiz ist viel höher als bisher angenommen*. Aug. 2022. URL: https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-90116.html.

[26] Massimo Germano et al. "A dynamic subgrid-scale eddy viscosity model". In: *Physics of Fluids A: Fluid Dynamics* 3.7 (1991), pp. 1760–1765. DOI: 10.1063/1.857955.

[27] Sharath S. Girimaji and Sawan Suman. "Partially Averaged Navier Stokes (PANS) Method for Turbulence Simulations: Theory and Practice". In: *Progress in Hybrid RANS-LES Modelling*. Ed. by Song Fu et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 29–43. ISBN: 978-3-642-31818-4.

[28] T. Göçmen et al. "Wind turbine wake models developed at the technical university of Denmark: A review". In: *Renewable and Sustainable Energy Reviews* 60 (2016), pp. 752–769. ISSN: 1364-0321. DOI: 10.1016/j.rser.2016.01.113.

[29] C. Gorlé et al. "The deviation from parallel shear flow as an indicator of linear eddy-viscosity model inaccuracy". In: *Physics of Fluids* 26.5 (2014), p. 051702. DOI: 10.1063/1.4876577.

[30] P. E. Hamlington and W. J. A. Dahm. "Reynolds stress closure for nonequilibrium effects in turbulent flows". In: *Physics of Fluids* 20.11 (2008), p. 115101. DOI: 10.1063/1.3006023.

[31] F. H. Harlow. "Transport Equations in Turbulence". In: *The physiscs of fluid* (1970). DOI: 10.1063/1.1692845.

[32] J. R. Holland, J. D. Baeder, and K. Duraisamy. "Field Inversion and Machine Learning With Embedded Neural Networks: Physics-Consistent Neural Network Training". In: *AIAA Aviation 2019 Forum* (June 2019). DOI: 10.2514/6.2019-3200.

[33] Jonathan R. Holland, James D. Baeder, and Karthik Duraisamy. "Towards Integrated Field Inversion and Machine Learning With Embedded Neural Networks for RANS Modeling". In: *AIAA Scitech 2019 Forum*. DOI: 10.2514/6.2019-1884.

[34] G. V. Iungo et al. "Data-driven RANS for simulations of large wind farms". In: *Journal of Physics: Conference Series* 625 (2015), p. 012025. DOI: 10.1088/1742-6596/625/1/012025.

[35] J. Jack Weatheritt et al. "Data-driven scalar-flux model development with application to jet in cross flow". In: *International Journal of Heat and Mass Transfer* 147 (2020), p. 118931. ISSN: 0017-9310. DOI: 10.1016/j.ijheatmasstransfer.2019.118931.

[36] S. Jee, G. Medic, and G. Kalitzin. "Note on Turbulent-Kinetic-Energy Production for Reynolds-Averaged Navier-Stokes Models". In: *Journal of Fluids Engineering* 138 (June 2016). DOI: 10.1115/1.4033750.

[37] N.O. Jensen. *A note on wind generator interaction*. English. Risø-M 2411. Risø National Laboratory, 1983. ISBN: 87-550-0971-9.

[38] M. L. A. Kaandorp and R. P. Dwight. "Stochastic Random Forests with Invariance for RANS Turbulence Modelling". In: (Oct. 2018).

5

[39] M. L.A. Kaandorp and R. P. Dwight. “Data-driven modelling of the Reynolds stress tensor using random forests with invariance”. In: *Computers & Fluids* 202 (Apr. 2020), p. 104497. ISSN: 0045-7930. DOI: 10.1016/j.compfluid.2020.104497.

[40] A. El Kasmi and C. Masson. “An extended k–ϵ model for turbulent flow through horizontal-axis wind turbines”. In: *Journal of Wind Engineering and Industrial Aerodynamics* 96.1 (2008), pp. 103–122. ISSN: 0167-6105. DOI: 10.1016/j.jweia.2007.03.007.

[41] R. N. King et al. “Data-Driven Machine Learning for Wind Plant Flow Modeling”. In: *Journal of Physics: Conference Series* 1037 (June 2018), p. 072004. DOI: 10.1088/1742-6596/1037/7/072004.

[42] Louis Kokee. “Frozen-RANS Turbulence Model Corrections for Wind Turbine Wakes in Stable, Neutral and Unstable Atmospheric Boundary Layers”. MA thesis. Delft University of Technology and Technical University of Denmark, Aug. 2021. URL: http://resolver.tudelft.nl/uuid:0373e228-68a1-4706-8df1-cbe3c2b1f248.

[43] B. van de Krol. “Data-Driven Improvement of RANS Simulations of Wind Farms in Stable Atmospheric Boundary Layer Conditions”. MA thesis. Delft University of Technology, Aug. 2022. URL: http://resolver.tudelft.nl/uuid:e30ddc04-d433-48e1-99b4-25f29d686097.

[44] J. N. Kutz. “Deep learning in fluid dynamics”. In: *Journal of Fluid Mechanics* 814 (2017), pp. 1–4. DOI: 10.1017/jfm.2016.803.

[45] M. P. van der Laan and S. J. Andersen. “The turbulence scales of a wind turbine wake: A revisit of extended k-ϵ models”. In: *Journal of Physics: Conference Series* 1037 (June 2018), p. 072001. DOI: 10.1088/1742-6596/1037/7/072001.

[46] M. P. van der Laan et al. “An improved k-ϵ model applied to a wind turbine wake in atmospheric turbulence”. In: *Wind Energy* 18.5 (2015), pp. 889–907. DOI: 10.1002/we.1736.

[47] M. P. van der Laan et al. “Nonlinear eddy viscosity models applied to wind turbine wakes”. In: (Dec. 2013), pp. 514–525.

[48] P. van der Laan, M. C. Kelly, and N. N. Sørensen. “A new k-ϵ model consistent with Monin-Obukhov similarity theory”. English. In: *Wind Energy* 20.3 (2017), pp. 479–489. ISSN: 1095-4244. DOI: 10.1002/we.2017.

[49] B. E. Launder and B. I. Sharma. “Application of the Energy Dissipation Model of Turbulence to the Calculation of Flow Near a Spinning Disc”. In: *Letters in Heat and Mass Transfer* 1.2 (1974), pp. 131–138.

[50] B. E. Launder and D. B. Spalding. “The numerical computation of turbulent flows”. In: *Computer Methods in Applied Mechanics and Engineering* 3.2 (1974), pp. 269–289. ISSN: 0045-7825. DOI: 10.1016/0045-7825(74)90029-2.

[51] S. Lee et al. “Atmospheric and Wake Turbulence Impacts on Wind Turbine Fatigue Loading: Preprint”. In: (Jan. 2012). DOI: 10.2514/6.2012-540.

[52] Randolph Lillard et al. “The lagRST Model: a Turbulence Model for Non-Equilibrium Flows”. In: *50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*. 2012. DOI: 10.2514/6.2012-444.

[53] J. Ling, A. Kurzawski, and J. Templeton. “Reynolds averaged turbulence modelling using deep neural networks with embedded invariance”. In: *Journal of Fluid Mechanics* 807 (2016), pp. 155–166. DOI: 10.1017/jfm.2016.615.

[54] J. Ling and J. Templeton. “Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty”. In: *Physics of Fluids* 27.8 (2015), p. 085103. DOI: 10.1063/1.4927765.

[55] R. Longo et al. “Advanced turbulence models and boundary conditions for flows around different configurations of ground-mounted buildings”. In: *Journal of Wind Engineering and Industrial Aerodynamics* 167 (2017), pp. 160–182. ISSN: 0167-6105. DOI: 10.1016/j.jweia.2017.04.015.

[56] X. Mao and J.N. Sorensen. “Far-wake meandering induced by atmospheric eddies in flow past a wind turbine”. In: *Journal of Fluid Mechanics* 846 (2018), pp. 190–209. DOI: 10.1017/jfm.2018.275.

[57] L. Martinez Tossas et al. “A Comparison of Actuator Disk and Actuator Line Wind Turbine Models and Best Practices for Their Use”. In: Jan. 2012. ISBN: 978-1-60086-936-5. DOI: 10.2514/6.2012-900.

[58] D. Mehta et al. “Large Eddy Simulation of wind farm aerodynamics: A review”. In: *Journal of Wind Engineering and Industrial Aerodynamics* 133 (2014), pp. 1–17.

[59] F. R. Menter. “Two-Equation Eddy-Viscosity Turbulence Models for Engineering Applications”. In: *AIAA journal* 32.8 (1994).

[60] F. R. Menter, A. Matyushenko, and R. Lechner. *Development of a Generalized K-ω Two-Equation Turbulence Model*. Ed. by Andreas Dillmann et al. Cham: Springer International Publishing, 2020, pp. 101–109. ISBN: 978-3-030-25253-3.

[61] F.R. Menter. *NASA technical memorandum 103975: Improved Two-equation K-omega Turbulence Models for Aerodynamic Flows*. 1992.

[62] C. Montavon. “Validation of a non-hydrostatic numerical model to simulate stratified wind fields over complex topography”. In: *Journal of Wind Engineering and Industrial Aerodynamics* 74-76 (1998), pp. 273–282. ISSN: 0167-6105. DOI: https://doi.org/10.1016/S0167-6105(98)00024-5.

[63] Y.-I. Moon, B. Rajagopalan, and U. Lall. “Estimation of mutual information using kernel density estimators”. In: *Phys. Rev. E* 52 (3 Sept. 1995), pp. 2318–2321. DOI: 10.1103/PhysRevE.52.2318.

[64] F. Nicoud and F. Ducros. “Subgrid-Scale Stress Modelling Based on the Square of the Velocity Gradient Tensor”. In: *Flow, Turbulence and Combustion* 62.3 (Sept. 1999), pp. 183–200. ISSN: 1573-1987. DOI: 10.1023/A:1009995426001.

[65] Kyle Nieuwenhuisen. “Inferring algebraic stress models from high-fidelity data and applying machine learning to enhance and generalize turbulence lag models”. MA thesis. Delft University of Technology, June 2020. URL: http://resolver.tudelft.nl/uuid:edd61bf3-30b3-41ea-ab8b-096da65ebff3.

[66] E. J. Parish and K. Duraisamy. “A paradigm for data-driven predictive modeling using field inversion and machine learning”. In: *Journal of Computational Physics* 305 (2016), pp. 758–774. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2015.11.012.

[67] S. B. Pope. “A more general effective-viscosity hypothesis”. In: *Journal of Fluid Mechanics* 72.2 (1975), pp. 331–340. DOI: 10.1017/S0022112075003382.

[68] S. B. Pope. *Turbulent flows*. Cambridge ; Cambridge University Press, 2000.

[69] S. V. Poroseva, Colmenares F. J. D., and S. M. Murman. “On the accuracy of RANS simulations with DNS data”. In: *Physics of Fluids* 28.11 (2016), p. 115102. DOI: 10.1063/1.4966639.

[70] F. Porté-Agel et al. “Large-eddy simulation of atmospheric boundary layer flow through wind turbines and wind farms”. In: *Journal of Wind Engineering and Industrial Aerodynamics* 99.4 (2011). The Fifth International Symposium on Computational Wind Engineering, pp. 154–168. ISSN: 0167-6105. DOI: 10.1016/j.jweia.2011.01.011.

[71] J. M. Prospathopoulos et al. “Evaluation of the effects of turbulence model enhancements on wind turbine wake predictions”. In: *Wind Energy* 14.2 (2011), pp. 285–300. DOI: 10.1002/we.419.

[72] P.-E. Rethore. “Wind Turbine Wake in Atmospheric Turbulence”. Riso-PhD, No.53(EN). PhD thesis. Roskilde: Riso National Laboratory for Sustainable Energy, 2009.

[73] P.-E. M. Rethore et al. “Study of the atmospheric wake turbulence of a CFD actuator disc model”. English. In: *EWEC 2009 Proceedings online*. EWEC, 2009.

[74] J. A. M. S. Richard, L. A. Martínez-Tossas, and C. Charles Meneveau. “Comparison of wind farm large eddy simulations using actuator disk and actuator line models with wind tunnel experiments”. In: *Renewable Energy* 116 (2018), pp. 470–478. ISSN: 0960-1481. DOI: 10.1016/j.renene.2017.08.072.

[75] Rijksoverheid. *Why wind energy as an energy source?* 2022. URL: https://windopzee.nl/onderwerpen/wind-zee/waarom/#PagCls_1773602.

[76] W. Rodi. “A New Algebraic Relation for Calculating the Reynolds Stresses”. In: *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik* 56.S1 (1976), T219–T221. DOI: https://doi.org/10.1002/zamm.19765613093.

[77] B. Sanderse, S. P. Pijl, and B. Koren. “Review of computational fluid dynamics for wind turbine wake aerodynamics”. In: *Wind Energy* 14.7 (2011), pp. 799–819.

[78] J. Sanz Rodrigo, M. Churchfield, and B. Kosovic. “A methodology for the design and testing of atmospheric boundary layer models for wind energy applications”. In: *Wind Energy Science* 2.1 (2017), pp. 35–54. DOI: 10.5194/wes-2-35-2017.

[79] M. Schmelzer, R. Dwight, and P. Cinnella. “Discovery of Algebraic Reynolds-stress Models using Sparse Symbolic Regression”. In: *Flow Turbulence and Combustion* (Oct. 2019).

[80] T. Shih et al. “A new k-ε eddy viscosity model for high reynolds number turbulent flows”. In: *Computers & Fluids* 24.3 (1995), pp. 227–238. ISSN: 0045-7930. DOI: https://doi.org/10.1016/0045-7930(94)00032-T.

[81] M. Shives and C. Crawford. “Adapted two-equation turbulence closures for actuator disk RANS simulations of wind & tidal turbine wakes”. In: *Renewable Energy* 92 (2016), pp. 273–292. ISSN: 0960-1481. DOI: 10.1016/j.renene.2016.02.026.

[82] A. P. Singh and K. Duraisamy. “Using field inversion to quantify functional errors in turbulence closures”. In: *Physics of Fluids* 28.4 (2016), p. 045110. DOI: 10.1063/1.4947045.

[83] A. P. Singh, S. Medida, and K. Duraisamy. “Machine-Learning-Augmented Predictive Modeling of Turbulent Separated Flows over Airfoils”. In: *AIAA Journal* 55 (Aug. 2016). DOI: 10.2514/1.J055595.

[84] A. Sogachev, M. Kelly, and M. Y. Leclerc. “Consistent Two-Equation Closure Modelling for Atmospheric Research: Buoyancy and Vegetation Implementations”. In: *Boundary-Layer Meteorology* 145.2 (2012), pp. 307–327. DOI: 10.1007/s10546-012-9726-5.

[85] Philippe R. Spalart. “Detached-Eddy Simulation”. In: *Annual Review of Fluid Mechanics* 41.1 (2009), pp. 181–202. DOI: 10.1146/annurev.fluid.010908.165130.

[86] J. Steiner, R. P. Dwight, and A. Viré. “Data-driven RANS closures for wind turbine wakes under neutral conditions”. In: *Computers & Fluids* 233 (Jan. 2022), p. 105213. ISSN: 0045-7930. DOI: 10.1016/j.compfluid.2021.105213.

[87] R. J.A.M. Stevens and C. Meneveau. “Flow Structure and Turbulence in Wind Farms”. In: *Annual Review of Fluid Mechanics* 49.1 (2017), pp. 311–339. DOI: 10.1146/annurev-fluid-010816-060206.

[88] J. Thé and H. Yu. “A critical review on the simulations of wind turbine aerodynamics focusing on hybrid RANS-LES methods”. In: *Energy* 138 (2017), pp. 257–289. ISSN: 0360-5442. DOI: 10.1016/j.energy.2017.07.028.

[89] R. L. Thompson et al. “A methodology to evaluate statistical errors in DNS data of plane channel flows”. In: *Computers & Fluids* 130 (May 2016), pp. 1–7. ISSN: 0045-7930. DOI: 10.1016/j.compfluid.2016.01.014.

[90] B. D. Tracey, K. Duraisamy, and J. J. Alonso. “A Machine Learning Strategy to Assist Turbulence Model Development”. In: *53rd AIAA Aerospace Sciences Meeting*. DOI: 10.2514/6.2015-1287.

[91] P. van der Laan et al. “Efficient Turbulence Modeling for CFD Wake Simulations”. English. PhD thesis. Denmark, 2014.

[92] B.J. Vanderwende et al. “Simulating effects of a wind-turbine array using LES and RANS”. In: *Journal of Advances in Modeling Earth Systems* 8.3 (2016), pp. 1376–1390. DOI: 10.1002/2016MS000652.

[93] G. Ver Steeg and A. Galstyan. “Information-Theoretic Measures of Influence Based on Content Dynamics”. In: *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*. WSDM ’13. Rome, Italy: Association for Computing Machinery, 2013, pp. 3–12. ISBN: 9781450318693. DOI: 10.1145/2433396.2433400.

[94] L.J. Vermeer, J.N. Sørensen, and A. Crespo. “Wind turbine wake aerodynamics”. In: *Progress in Aerospace Sciences* 39.6 (2003), pp. 467–510. ISSN: 0376-0421. DOI: https://doi.org/10.1016/S0376-0421(03)00078-2.

[95] J.-X. Wang, J.-L. Wu, and H. Xiao. “Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data”. In: *Phys. Rev. Fluids* 2.3 (3 Mar. 2017), p. 034603. ISSN: 2469-990X. DOI: 10.1103/PhysRevFluids.2.034603.

[96] J.-X. Wang et al. “A Comprehensive Physics-Informed Machine Learning Framework for Predictive Turbulence Modeling”. In: (Jan. 2017).

[97] Y. Wang et al. “Impact of atmospheric stability on wind turbine wake velocity distribution”. In: *International Conference on Renewable Power Generation (RPG 2015)*. Oct. 2015, pp. 1–5. DOI: 10.1049/cp.2015.0456.

[98] F. Waschkowski et al. “Multi-objective CFD-driven development of coupled turbulence closure models”. In: *Journal of Computational Physics* 452 (Mar. 2022), p. 110922. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2021.110922.

[99] J. Weatheritt and R. Sandberg. “A novel evolutionary algorithm applied to algebraic modifications of the RANS stress–strain relationship”. In: *Journal of Computational Physics* 325 (Nov. 2016), pp. 22–37. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2016.08.015.

[100] J. Weatheritt and R. D. Sandberg. “The development of algebraic stress models using a novel evolutionary algorithm”. In: *International Journal of Heat and Fluid Flow* 68 (Dec. 2017), pp. 298–318. ISSN: 0142-727X. DOI: 10.1016/j.ijheatfluidflow.2017.09.017.

[101] J. Weatheritt et al. “A scalar-flux model for jets in cross flow from database analysis”. In: *DRAFT* (2019).

[102] D. C. Wilcox. “Reassessment of the scale-determining equation for advanced turbulence models”. In: *AIAA Journal* 26.11 (1988), pp. 1299–1310.

[103] J. Wu et al. “RANS Equations with Reynolds Stress Closure Can Be Ill-Conditioned”. In: (Mar. 2018).

[104] J.-L. Wu, H. Xiao, and E. Paterson. “Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework”. In: *Phys. Rev. Fluids* 3.7 (7 July 2018), p. 074602. ISSN: 2469-990X. DOI: 10.1103/PhysRevFluids.3.074602.

[105] Y.-T. Wu and F. Porté-Agel. “Large-Eddy Simulation of Wind-Turbine Wakes: Evaluation of Turbine Parametrisations”. In: *Boundary-Layer Meteorology* 138.3 (Mar. 2011), pp. 345–366. ISSN: 1573-1472. DOI: 10.1007/s10546-010-9569-x.

[106] H. Xiao and P. Cinnella. “Quantification of model uncertainty in RANS simulations: A review”. In: *Progress in Aerospace Sciences* 108 (2019), pp. 1–31. ISSN: 0376-0421. DOI: 10.1016/j.paerosci.2018.10.001.

[107] S. Xie and C. Archer. “Self-similarity and turbulence characteristics of wind turbine wakes via large-eddy simulation”. In: *Wind Energy* 18 (Aug. 2014). DOI: 10.1002/we.1792.

[108] Frederik Zahle et al. “Design of the LRP airfoil series using 2D CFD”. In: *Journal of Physics: Conference Series* 524.1 (June 2014), p. 012020. DOI: 10.1088/1742-6596/524/1/012020. URL: https://dx.doi.org/10.1088/1742-6596/524/1/012020.

[109] Y. Zhang et al. *Customized data-driven RANS closures for bi-fidelity LES-RANS optimization*. 2020. arXiv: 2004.03003 [physics.comp-ph].

[110] Y. Zhao et al. “RANS turbulence model development using CFD-driven machine learning”. In: *Journal of Computational Physics* 411 (June 2020), p. 109413. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2020.109413.

[111] H. Zou and T. Hastie. “Regularization and variable selection via the elastic net”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2 (Apr. 2005), pp. 301–320. ISSN: 1467-9868. DOI: 10.1111/j.1467-9868.2005.00503.x.

A

APPENDIX A1: FEATURES AND INTEGRITY BASES

A.1. POPE'S INTEGRITY BASIS

Invariants λ_1 to λ_5
$\{\mathbf{S}^2\}$ $\{\mathbf{\Omega}^2\}$ $\{\mathbf{S}^3\}$ $\{\mathbf{\Omega}^2\mathbf{S}\}$ $\{\mathbf{\Omega}^2\mathbf{S}^2\}$

Tensor number	Tensor expression
$\mathbf{T}^1$	$\mathbf{S}$
$\mathbf{T}^2$	$\mathbf{s\Omega} - \mathbf{\Omega S}$
$\mathbf{T}^3$	$\mathbf{S}^2 - \frac{1}{3}\mathbf{I}\{\mathbf{S}^2\}$
$\mathbf{T}^4$	$\mathbf{\Omega}^2 - \frac{1}{3}\mathbf{I}\{\mathbf{\Omega}^2\}$
$\mathbf{T}^5$	$\mathbf{\Omega S}^2 - \mathbf{S}^2\mathbf{\Omega}$
$\mathbf{T}^6$	$\mathbf{\Omega}^2\mathbf{S} + \mathbf{S}^2\mathbf{\Omega} - \frac{2}{3}\mathbf{I}\{\mathbf{s\Omega}^2\}$
$\mathbf{T}^7$	$\mathbf{\Omega S\Omega}^2 - \mathbf{\Omega}^2\mathbf{S\Omega}$
$\mathbf{T}^8$	$\mathbf{S\Omega S}^2 - \mathbf{S}^2\mathbf{\Omega S}$
$\mathbf{T}^9$	$\mathbf{\Omega}^2\mathbf{S}^2 + \mathbf{S}^2\mathbf{\Omega}^2 - \frac{2}{3}\mathbf{I}\{\mathbf{S}^2\mathbf{\Omega}^2\}$
$\mathbf{T}^{10}$	$\mathbf{\Omega S}^2\mathbf{\Omega}^2 - \mathbf{\Omega}^2\mathbf{S}^2\mathbf{\Omega}$

A.2. PHYSICAL FEATURES

A

Table A.1: Physics interpreted flow features. For each feature q_i the physical description is denoted including the raw feature with its normalization. The features that are not Galilean invariant are marked with †.

ID	Description	Raw feature	Normalization
q_Q	Ratio of excess rotation rate to strain rate (Q criterion)	$\frac{1}{2}(\|\mathbf{\Omega}\|^2 - \|\mathbf{S}\|^2)$	$\|\mathbf{S}\|^2$
$q_{TI}^{\dagger}$	Turbulence intensity	k	$\frac{1}{2}U_iU_i$
q_{ReD}	Wall distance based Reynolds number	$\frac{\sqrt{k}d}{50\nu}$	-
$q_{\partial p \partial s}^{\dagger}$	Pressure gradient along streamline	$U_k\frac{\partial P}{\partial x_k}$	$\sqrt{\frac{\partial P}{\partial x_j}\frac{\partial P}{\partial x_j}U_iU_i}$
q_T	Ratio of mean turbulent to mean strain time scale	$\frac{k}{\varepsilon}$	$\frac{1}{\|\mathbf{S}\|}$
q_ν	Viscosity ratio	ν_t	100ν
$q_{\perp}^{\dagger}$	Nonorthogonality between velocity and its gradient	$\|U_iU_j\frac{\partial U_i}{\partial x_j}\|$	$\sqrt{U_lU_lU_i\frac{\partial U_i}{\partial x_j}U_k\frac{\partial U_k}{\partial x_j}}$
$q_{\mathscr{C}_k/\mathscr{P}_k}^{\dagger}$	Ratio of convection to Boussinesq production of TKE	$U_i\frac{dk}{dx_i}$	$\|\overline{u_j'u_k'}S_{jk}\|$
q_τ	Ratio of total to normal Boussinesq Reynolds stresses	$\|\|\overline{u_i'u_j'}_{BS}\|\|$	k
q_γ	Shear parameter	$\left\|\frac{\partial U_i}{\partial x_j}\right\|$	$\frac{\varepsilon}{k}$
$q_F^{\dagger}$	Actuator forcing	$\|F_{cell}\|$	$\frac{1}{2}\rho_0 A_{cell}\|U\|^2$

A.3. EXTENDED INTEGRITY BASIS

Table A.2: Invariant bases, number of symmetric and antisymmetric tensors for each invariant are indicated by n_s and n_A, respectively. The invariant bases are the trace of the tensors listed. The asterisk on a invariant bases indicates that also the cyclic permutation of the antisymmetric tensors are included.

(n_S, n_A)	Feature index	Invariant bases
$(1,0)$	1-2	$\mathbf{S}^2$, $\mathbf{S}^3$
$(0,1)$	3-5	$\mathbf{\Omega}^2$, $\mathbf{A}_p^2$, $\mathbf{A}_k^2$
$(1,1)$	6-14	$\mathbf{\Omega}^2\mathbf{S}$, $\mathbf{\Omega}^2\mathbf{S}^2$, $\mathbf{\Omega}^2\mathbf{S}\mathbf{\Omega}\mathbf{S}^2$
		$\mathbf{A}_p^2\mathbf{S}$, $\mathbf{A}_p^2\mathbf{S}^2$, $\mathbf{A}_p^2\mathbf{S}\mathbf{A}_p\mathbf{S}^2$
		$\mathbf{A}_k^2\mathbf{S}$, $\mathbf{A}_k^2\mathbf{S}^2$, $\mathbf{A}_k^2\mathbf{S}\mathbf{A}_k\mathbf{S}^2$
$(0,2)$	15-17	$\mathbf{\Omega}\mathbf{A}_p$, $\mathbf{A}_p\mathbf{A}_k$, $\mathbf{\Omega}\mathbf{A}_k$
$(1,2)$	18-41	$\mathbf{\Omega}\mathbf{A}_p\mathbf{S}$, $\mathbf{\Omega}\mathbf{A}_p\mathbf{S}^2$, $\mathbf{\Omega}^2\mathbf{A}_p\mathbf{S}^*$, $\mathbf{\Omega}^2\mathbf{A}_p\mathbf{S}^{2*}$, $\mathbf{\Omega}^2\mathbf{S}\mathbf{A}_p\mathbf{S}^{2*}$
		$\mathbf{\Omega}\mathbf{A}_k\mathbf{S}$, $\mathbf{\Omega}\mathbf{A}_k\mathbf{S}^2$, $\mathbf{\Omega}^2\mathbf{A}_k\mathbf{S}^*$, $\mathbf{\Omega}^2\mathbf{A}_k\mathbf{S}^{2*}$, $\mathbf{\Omega}^2\mathbf{S}\mathbf{A}_k\mathbf{S}^{2*}$
		$\mathbf{A}_p\mathbf{A}_k\mathbf{S}$, $\mathbf{A}_p\mathbf{A}_k\mathbf{S}^2$, $\mathbf{A}_p^2\mathbf{A}_k\mathbf{S}^*$, $\mathbf{A}_p^2\mathbf{A}_k\mathbf{S}^{2*}$
$(0,3)$	42	$\mathbf{\Omega}\mathbf{A}_p\mathbf{A}_k$
$(1,3)$	43-47	$\mathbf{\Omega}\mathbf{A}_p\mathbf{A}_k\mathbf{S}$, $\mathbf{\Omega}\mathbf{A}_k\mathbf{A}_p\mathbf{S}$, $\mathbf{\Omega}\mathbf{A}_p\mathbf{A}_k\mathbf{S}^2$, $\mathbf{\Omega}\mathbf{A}_k\mathbf{A}_p\mathbf{S}^2$, $\mathbf{\Omega}\mathbf{A}_p\mathbf{S}\mathbf{A}_k\mathbf{S}^2$

B

APPENDIX A2: ADDITIONAL FIGURES FOR RESULTS CHAPTER

B.1. MODELS WITHOUT CLASSIFIER - HORIZONTAL SLICES

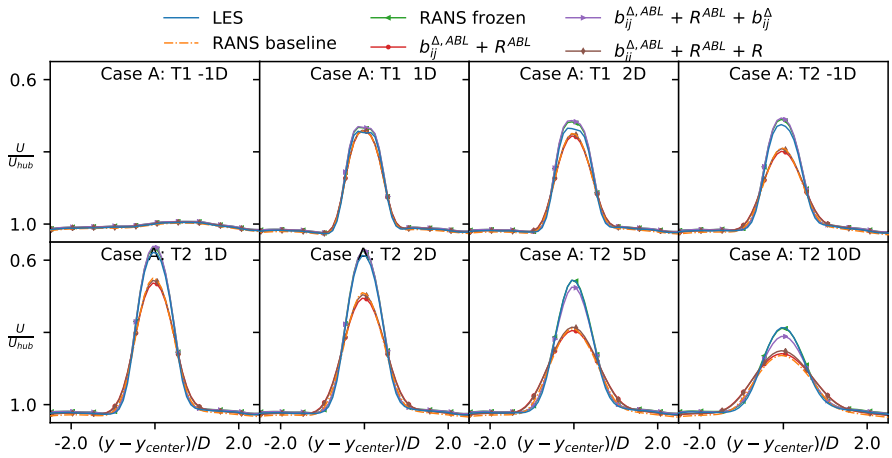


Figure B.1: Comparison between LES, RANS baseline and frozen RANS with selective inclusion of the different components of the correction terms via horizontal slices of the velocity field up and downstream of the rotor plane of the two turbines of case A.

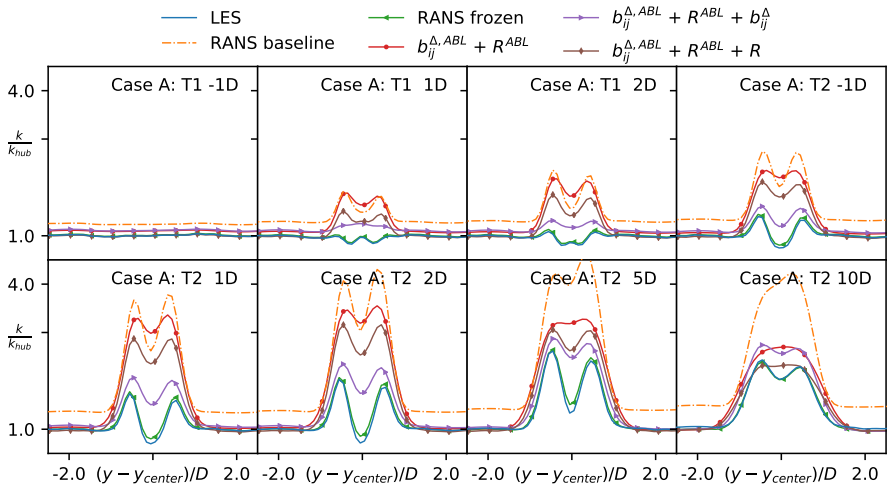


Figure B.2: Comparison between LES, RANS baseline and frozen RANS with selective inclusion of the different components of the correction terms via horizontal slices of the turbulent kinetic energy field up and downstream of the rotor plane of the two turbines of case A.

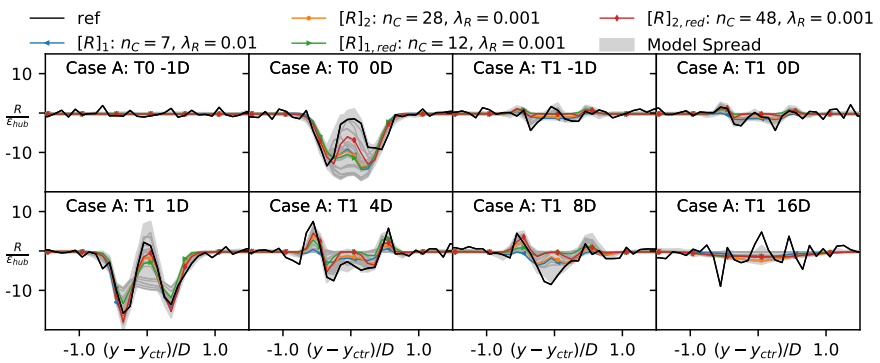


Figure B.3: Spread of learned model correction R for case A through horizontal slices at the rotor plane at different streamwise stations as labeled in the subplots. The subscript $R_{a,red}$ refers to models whose terms only contain positive powers. The model spread is for all models that are Pareto optimal as defined previously. The models selected during the cliqueing post-processing step are shown explicitly either in color or in dark gray. The models selected for further investigation are highlighted in color. Finally, the optimal correction term is shown in black.

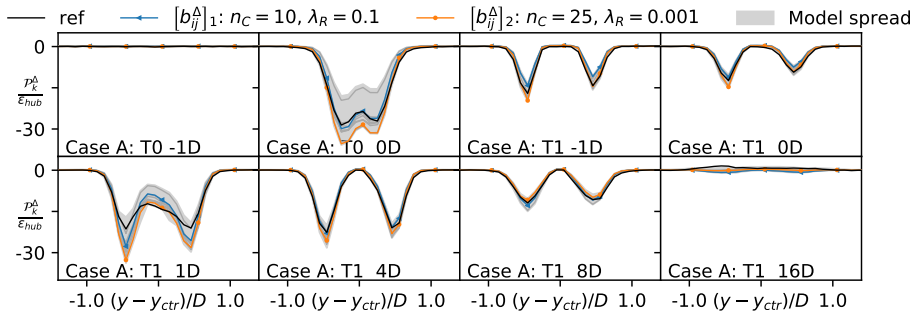


Figure B.4: Spread of learned model correction $\mathcal{P}_k^{\Delta}$ for case A through horizontal slices at the rotor plane at different streamwise stations as labeled in the subplots. The model spread is for all models that are Pareto optimal as defined previously. The models selected during the cliqueing post-processing step are shown explicitly either in color or in dark gray. The models selected for further investigation are highlighted in color. Finally, the optimal correction term is shown in black.

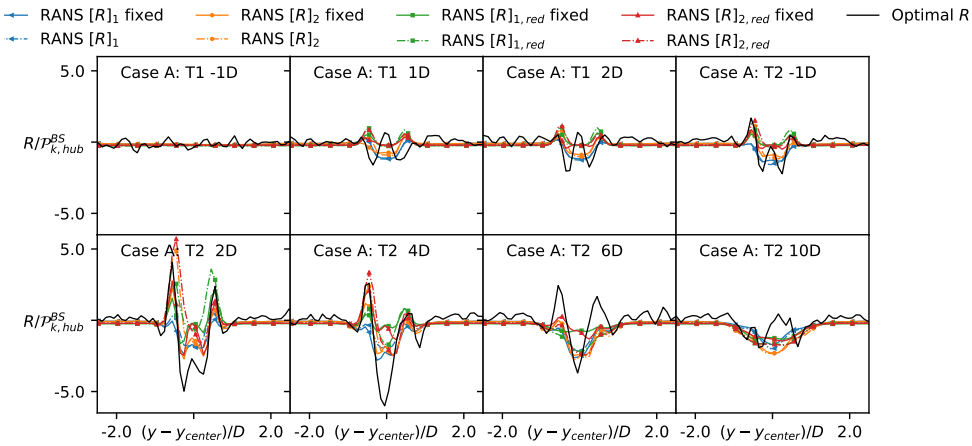


Figure B.5: Spread of fixed and coupled corrections R for the training case A via horizontal slices at the rotor plane up and downstream of the two turbines. The subscript $R_{a,red}$ refers to models whose terms only contain positive powers.

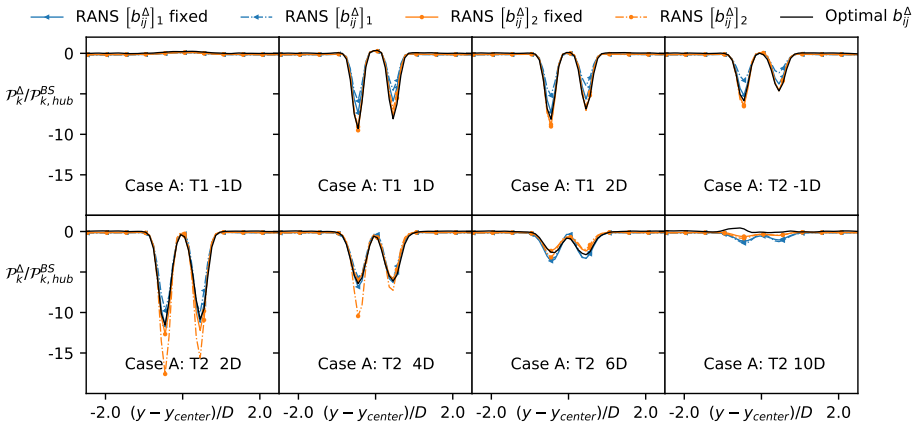


Figure B.6: Spread of fixed and coupled corrections b_{ij}^{Δ} in terms of $\mathscr{P}_k^{\Delta}$ for the training case A via horizontal slices at the rotor plane up and downstream of the two turbines.

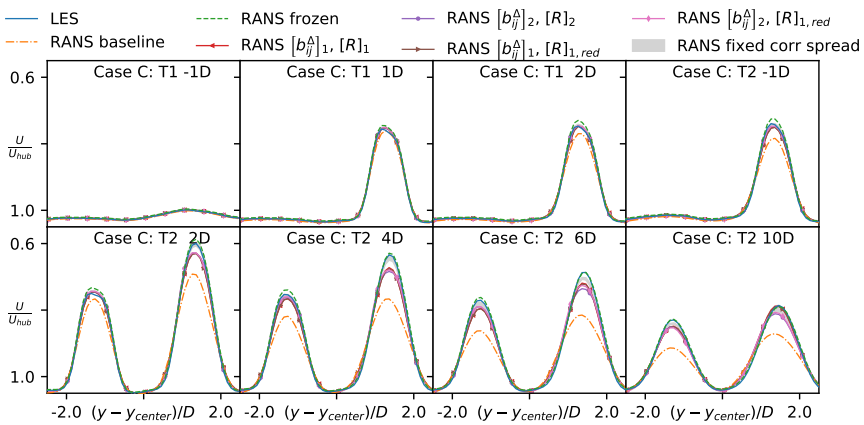


Figure B.7: Comparison between LES, RANS baseline, frozen RANS and corrected RANS models via horizontal slices of the velocity field up and downstream of the rotor plane for the three turbines of case B.

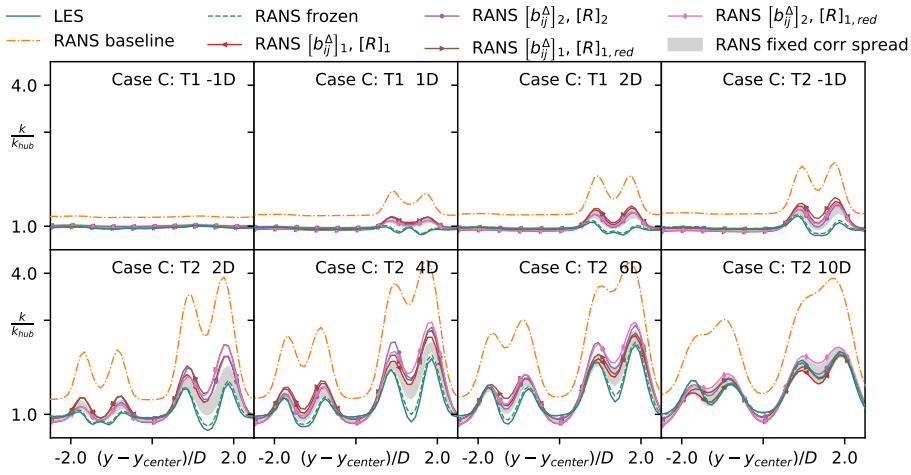


Figure B.8: Comparison between LES, RANS baseline, frozen RANS and corrected RANS models via horizontal slices of the turbulent kinetic energy field up and downstream of the rotor plane for the three turbines of case B.

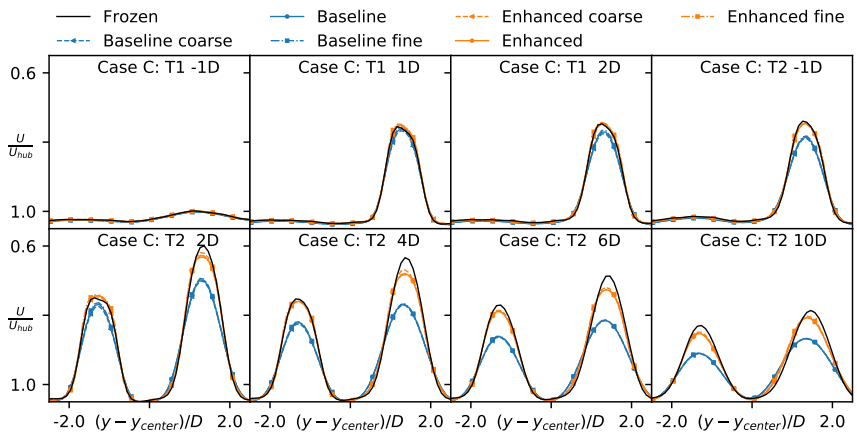


Figure B.9: Mesh convergence study for the baseline and the corrected model. Shown are the horizontal slices of the velocity field up and downstream of the rotor plane for the three turbines of case B.

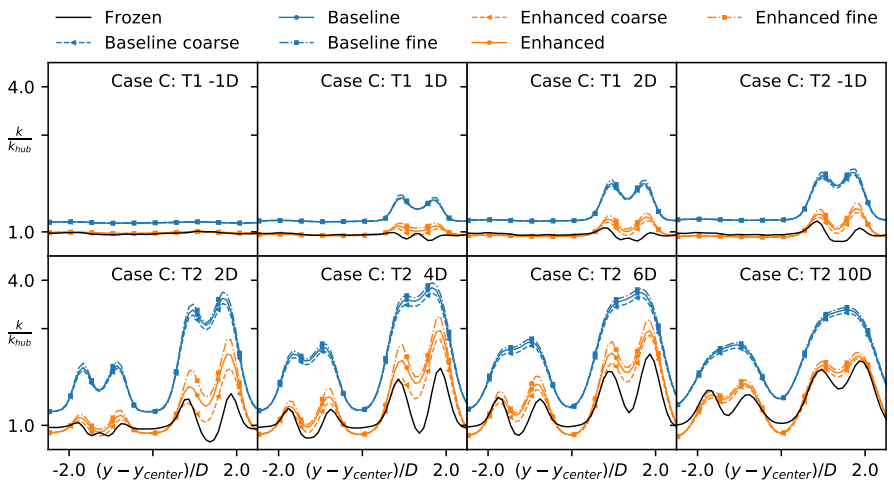


Figure B.10: Mesh convergence study for the baseline and the corrected model. Shown are the horizontal slices of the turbulent kinetic energy field up and downstream of the rotor plane for the three turbines of case B.

B.2. MODELS WITH CLASSIFIER - ADDITIONAL FIGURES FOR CASES A & B

Figures B.11 and B.12 show the results for the training data-set A. Figures B.11 and B.12 in the appendix show the results for the test data-set B.

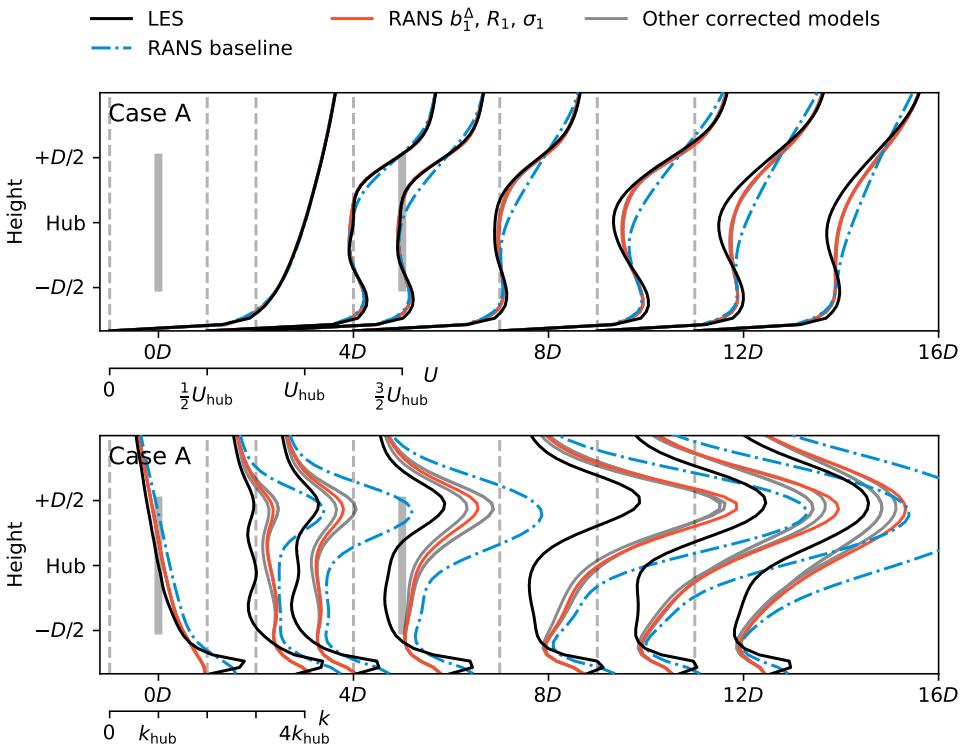


Figure B.11: LES, RANS baseline, and corrected RANS model predictions for Case A.

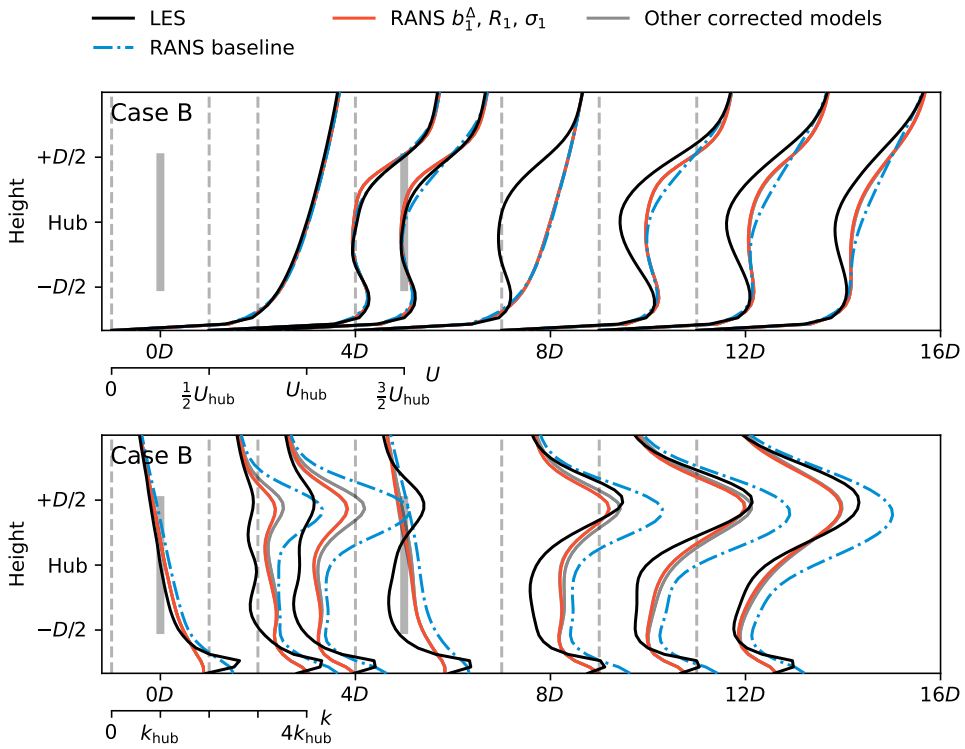


Figure B.12: LES, RANS baseline, and corrected RANS model predictions for Case B.

List of Publications

J. Steiner, A. Viré, F. Benetti, W.A. Timmer, R.P. Dwight, *Parametric slat design study for thick-base airfoils at high Reynolds numbers*, In: *Wind Energy Science*, 5 (3), pp. 1075-1095(2020), DOI: 10.5194/wes-5-1075-2020.

J. Steiner, R.P. Dwight, A. Viré, *Data-driven turbulence modeling for wind turbine wakes under neutral conditions*, In: *Journal of Physics: Conference Series, 1618 (6), Science of Making Torque from Wind 2020, TORQUE 2020, 2020-09-28 to 2020-10-02, Online, Virtual, Online, Netherlands (2020)*, (2020), DOI:10.1088/1742-6596/1618/6/062051.

A. Viré, B.P. LeBlanc, **J. Steiner**, W.A. Timmer, *Experimental study of the effect of a slat on the aerodynamic performance of a thick base airfoil*, In: *Wind Energy Science*, 7 (2), pp. 573-584 (2022), DOI: doi.org/10.5194/wes-7-573-2022.

J. Steiner, R.P. Dwight, A. Viré, *Data-driven RANS closures for wind turbine wakes under neutral conditions*, In: *Computers & Fluids*, 233 (2022), DOI: 10.1016/j.compfluid.2021.105213.

J. Steiner, R.P. Dwight, A. Viré, *Classifying Regions of High Model Error Within a Data-Driven RANS Closure: Application to Wind Turbine Wakes*, In: *Flow, Turbulence and Combustion*, 109 (3), pp. 545-570 (2022), DOI: 10.1007/s10494-022-00346-6.