Limited Area Forecasting and Statistical Modelling for Wind Energy Scheduling

Rosgaard, Martin Haubjerg; Hahmann, Andrea N.

Publication date:
2015

TECHNICAL UNIVERSITY OF DENMARK

Limited Area Forecasting and Statistical Modelling for Wind Energy Scheduling

by

Martin Haubjerg Rosgaard

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the
Department of Wind Energy
Meteorology Section

April 2015

With the increasingly intrusive insight gained since the dawn of the Atomic Age the 'cowboy' economy rumbling across the apparent illimitable plains of the past may today and in the future similarly be called the 'spaceman' economy, in which Earth has become a single spaceship, without unlimited reservoirs of anything, either for extraction or for pollution, and in which, therefore, man must find his place in a cyclical ecological system.

Freely adapted from the essay
The Economics of the Coming Spaceship Earth
K. E. Boulding
1966

TECHNICAL UNIVERSITY OF DENMARK

Abstract

Department of Wind Energy
Meteorology Section

PhD thesis

by Martin Haubjerg Rosgaard

The worldwide deployment of wind energy has continually accelerated during the last few decades, and the implications of day ahead predictability for this highly fluctuating renewable energy source for feasible wind power integration in electrical grids are multifaceted. This thesis concerns forecast accuracy for operational wind power scheduling.

Numerical weather prediction history and scales of atmospheric motion are summarised, followed by a literature review of limited area wind speed forecasting. Hereafter, the original contribution to research on the topic is outlined. The quality control of wind farm data used as forecast reference is described in detail, and a preliminary limited area forecasting study illustrates the aggravation of issues related to numerical orography representation and accurate reference coordinates at fine weather model resolutions.

For the offshore and coastal sites studied limited area forecasting is found to deteriorate wind speed prediction accuracy, while inland results exhibit a steady forecast performance increase with weather model resolution. Temporal smoothing of wind speed forecasts is shown to improve wind power forecast performance by up to almost 1 %, and the explanatory value for wind power forecasting of six different prognostic and diagnostic weather model variables modelled semi-parametrically is found to differ depending on the local terrain. In terms of wind speed ramp predictability, the study finds consistent improvement for better resolved forecasts, and indications of wind speed fluctuation phase-drift with weather model integration time are countenanced, which in part explains the faster decline in limited area forecast performance with leadtime, relative to global model forecasts. The limited area forecasting study is rounded off with a demonstration of the feasibility of forecasted wind speed variability for predicting wind power uncertainty. Finally, a statistical postprocessing framework for numerical wind speed forecasts is developed and evaluated, and the proposed methodology made possible the discovery of the lifted index weather model diagnostic as containing systematic corrective potential for wind speed forecasts generated by the weather model studied.

DANMARKS TEKNISKE UNIVERSITET

Dansk Resumé

Institut for Vindenergi
Meteorologi sektionen

PhD afhandling

af Martin Haubjerg Rosgaard

Vindenergikapaciteten på verdensplan øges i stadig højere tempo og følgelig er forudsigelighedsgraden af denne bæredygtige energiform af væsentlig betydning for effektiv integration i elnettet. Denne afhandling drejer sig om vejrprognosenøjagtighed med henblik på operationel drift af vindparker og effektiv håndtering af den producerede el.

Numerisk vejrprediktion gennemgås historisk og forskellige skalaer for atmosfæredynamik opridses, efterfulgt af et litteraturstudie af vindhastighedsprediktion med regionale vejrmodeller. Kvalitetskontrol af vindparkdatasæt der anvendes som vejrprognosereferencer beskrives detaljeret, og den negative indflydelse af relativt små afvigelser i det anvendte terrænelevationsdatasæt, samt i vindpark referencekoordinatet, illustreres via et forstudie for en af vindparkerne.

For de studerede vindparker offshore og på kysten forværrer de genererede regionale vejrprognoser nøjagtigheden mht. vindhastighed, hvorimod nøjagtigheden af højopløste vejrprognoser forbedres i komplekst terræn. Tidslig udjævning af vindhastighedsprognoser giver en forbedring af vindkraftprognosenøjagtighed på knap 1 %, og forskellig præstation afhængig af lokationen findes ved semiparametrisk modellering af seks prognostiske og diagnostiske vejrmodelvariable. Det vises at vindhastighedsramper forudsiges bedre for højopløste regionale vejrprognoser, og ydermere demonstreres det at modelprognoserne gradvis forsinkes ift. referencemålinger hvilket delvis forklarer det hurtigere tab af nøjagtighed for regionale- vs. globale vejrprognoser. Vha. ændringer i den regionale models kildekode vises det at vindhastighedsvariabilitet er en lovende prediktor i fremtidige statistiske modeller for vindkraftusikkerhed. Endelig udvikles og demonstreres – for førnævnte globale vejrmodel – en statistisk efterbehandlingsmetode til at opnå forbedrede vindhastighedsprognoser, og vha. af denne fremgangsmåde findes det at den diagnostiske *lifted index* variabel indeholder systematisk korrigerende værdi.

Acknowledgements

First and foremost I thank Andrea Hahmann, my principal supervisor, for introducing me to meteorology and numerical weather prediction. The patient guidance during the early stages of the project supported my steady progress, and numerous discussions and feedback provided later on have been invaluable for my gradual transition to professional scientist. I would like to thank my colleagues at ENFOR A/S; Sven Thomsen, Stig Mortensen, Henrik Nielsen, and Torben Nielsen, for generously sharing their extensive data analysis and statistical modelling knowhow, and for the active participation in shaping the investigations described in the thesis. I benefitted greatly from discussions on meteorology-related research and associated technical aspects with Neil Davis and Claire Vincent of DTU Wind Energy, and by the same token I appreciate comments from Mark Kelly and Jake Badger on results near the end of the project. I am thankful to Gregor Giebel, Henrik Madsen and Poul Sørensen for their suggestions during project meetings. I would also like to thank my fellow PhD students for the good atmosphere at the departments of Wind Energy and Applied Mathematics and Computer Science at the Technical University of Denmark.

The Danish Meteorological Institute kindly provided access to courses on Earth-system modelling and predictability undertaken in March and May 2013, respectively, at the European Centre for Medium-Range Weather Forecasts. Being a newcomer to the fields of meteorology and numerical weather prediction the intensive, high-quality courses gave me an overview of frontline research challenges, disseminated by employees at the world leader of medium-range weather forecasts.

During the summer of 2013 I went on external stays at two institutions in the United States of America. Joshua Hacker hosted my visits to the Naval Postgraduate School in Monterey, California, and the National Center for Atmospheric Research in Boulder, Colorado, and introduced me to the ensemble Kalman filter approach to data assimilation. Thus digressing a bit from the project, I modified a tool implementing this approach to allow wind power data to be assimilated, aided by Nancy Collins from the NCAR Mesa Lab. My stay abroad brought me a quantum leap forward in understanding numerical weather prediction in practice. Besides the friendly guidance from Josh and Nancy, I greatly value professional discussions as well as recreational activities with my office mate at NPS, Jared Lee, along with Andrew Penny, Kurt Nielsen and Walter Kolczynski, likewise from NPS. The same goes for Caroline Draxl, who also hosted my 1-day visit to the National Renewable Energy Laboratory in Boulder.

Finally, I am grateful for the most important dynamics under the sky; my wife and children, who occasionally drag me away from the computer and out into actual weather.

Contents

Abstract iii

Dansk Resumé iv

Acknowledgements v

List of Figures ix

List of Tables xi

Abbreviations xii

Nomenclature xiv

Journal Publication Drafts xv

1 Introduction 1
- 1.1 Historical rundown of numerical weather prediction . . . 2
- 1.2 Spatio-temporal scales of atmospheric dynamics . . . 4
- 1.3 Wind speed forecast dependence on model resolution . . . 5
- 1.4 Research contribution . . . 9
- 1.5 Software tools . . . 12

2 Forecast reference data 13
- 2.1 Wind farm data vs. measurement tower data . . . 14
- 2.2 Algorithm for processing of raw wind farm data . . . 15
- 2.3 Detection of anemometer/power data inconsistency . . . 18
- 2.4 A reference for the limited area forecasting studies . . . 23

3 Limited area forecasting 25
- 3.1 Initialisation of limited area weather model forecasts . . . 27
- 3.2 Parameterisations for sub-grid scale physics . . . 28
- 3.3 Terrain representation on high-resolution model grids . . . 31
  - 3.3.1 Forecast performance assessment strategy . . . 31
  - 3.3.2 Forecast performance as a function of spatial smoothing . . . 33

3.3.3 Remarks on the findings . . . 38
3.4 Introduction to appendices A and B . . . 39
3.5 Multidimensional nearest neighbour smoothing . . . 41
3.6 Supplementary discussion on wind speed ramps . . . 42
3.7 Error sources pertaining to limited area forecasting . . . 43
3.7.1 Starting out right — forecast initialisation . . . 43
3.7.2 Spurious perturbations — lateral boundary condition schemes . . . 45

4 Weather model deficiency **48**
4.1 Statistical forecasting by general linear modelling . . . 51
4.2 Introduction to Appendix C . . . 52
4.3 Multicollinearity . . . 54
4.4 Autocorrelation . . . 55
4.5 Error-free predictors . . . 57

5 Conclusion & further research **59**

A Journal publication I — *draft* **63**

B Journal publication II — *draft* **80**

C Journal publication III — *draft* **99**

Bibliography **113**

List of Figures

2.1 Anemometer readings vs. power production scatter density plots. The power curve region is heuristically framed by the dashed and solid lines. UPPER LEFT: Horns Rev. UPPER RIGHT: Rejsby Hede. BOTTOM: Stor-Rotliden. 19

2.2 Photograph of a Stor-Rotliden wind turbine blade during an icing event [108]. 20

2.3 Cumulative count of data points falling outside the visually inferred power curve region framed by dashed and solid lines in Figure 2.1. UPPER LEFT: Horns Rev. UPPER RIGHT: Rejsby Hede. BOTTOM: Stor-Rotliden. 21

2.4 **N.B.: Full-year GFS bias and scaling error has been removed cf. *Appendix A Section 2.1*.** LEFT: Monthly mean of farm-averaged wind turbine nacelle anemometer data for Stor-Rotliden, along with forecast leadtime hour 0 ↔ 12 twice-daily GFS wind speed data, as well as the corresponding forecast bias. RIGHT: Cumulative squared- and absolute forecast error for the 10 minute dataset. 22

3.1 Sketch of how a finer-resolved Arakawa C-grid [117] computational domain is nested within the parent domain. θ denotes scalar quantities such as pressure and temperature, while U and V represent the horizontal components of atmospheric flow. The figure is borrowed from [119] with permission. 26

3.2 Sketch of atmospheric dynamics that cannot be resolved explicitly and therefore needs to be parameterised. Reproduced from [126] with permission. 29

3.3 Sketch of terrain tiles representing the lower boundary of atmospheric dynamics simulations. The figure is borrowed from [146] with permission. 31

3.4 Wind speed forecast performance as a function of spatial smoothing by Eq. (3.1) for the two high-resolution WRF model grids considered in the studies introduced in Section 3.4, based on the *official* wind farm coordinate [151]. Forecast leadtime dependency is resolved according to the strategy laid out in Section 3.3.1. The number of grid points within r is noted above (below) 1.1 km (3.3 km) PCC values for the 36 ↔ 48 forecast leadtime hour group. 34

3.5 Representation of the Stor-Rotliden site in the WRF model. TOP: Layout of the wind farm turbines, reproduced here with permission [152]. LOWER LEFT: Orography map [153] with near-turbine 1.1 km domain grid points marked in red, numbers denote the ordinally ranked proximity to the wind turbine average coordinate **X**. The official wind farm coordinate [151] for which the four WRF model computational domains of the study in Section 3.4 have a grid point specified is marked with **O**. LOWER RIGHT: Representation of terrain elevation above mean sea level in the WRF model; generated using NCL [154]. 35
3.6 Wind speed forecast performance for the WRF modelled 1.1 km grid points sketched in Figure 3.5. 37
3.7 Cf. Figure 3.4, though the smoothing reference wind farm coordinate is replaced by the average of the 40 individual wind turbine coordinates constituting the Stor-Rotliden wind farm. 37

4.1 Illustration of the complexly coupled physical processes that need to be accurately simulated in weather and climate models in order to reproduce surface measurements of atmospheric variables. Black arrows denote surface and boundary layer processes, red arrows radiation, while green arrows are land surface processes. Solid (dashed) lines represent positive (negative) feedback. Circled numbers 1 to 8 display the origin and effect of the sketched physical processes, respectively. The figure is adapted and reproduced with permission [243]. 53
4.2 The black, vertical bars denote autocorrelation for the 12 $\leftrightarrow$ 24, 24 $\leftrightarrow$ 36, and 36 $\leftrightarrow$ 48 forecast leadtime groups considered in Appendix C (year 1). The blue circles denote autocorrelation predictions from a first-order autoregressive model, AR(1), for the residuals of the full MOS model considered in Appendix C. Similarly, the red bullets mark AR(2) model predictions. The grey, horizontal lines delimit an approximate white noise 95 % confidence interval cf. [249]. TOP: The offshore Horns Rev wind farm. MIDDLE: Coastal Rejsby Hede. BOTTOM: Stor-Rotliden, situated in central Sweden. 56

List of Tables

1.1 Hydrostatic and nonhydrostatic NWP model formulation traits. 3
1.2 Characteristic atmosperic dynamics scales as proposed by Orlanski [35]. . 4
1.3 Characteristic atmosperic dynamics scales as proposed by Lackmann [36]. 5

2.1 Wind farm data for the forecast reference sites considered in this work. The wind farms are operated by the Swedish energy company *Vattenfall*. . 14
2.2 Raw **wind speed** data for **Horns Rev** processed according to the algorithm described in Section 2.2. 16
2.3 Raw **wind speed** data for **Rejsby Hede** processed according to the algorithm described in Section 2.2. 16
2.4 Raw **wind speed** data for **Stor-Rotliden** processed according to the algorithm described in Section 2.2. 16
2.5 Raw **wind power** data for **Horns Rev** processed according to the algorithm described in Section 2.2. 17
2.6 Raw **wind power** data for **Rejsby Hede** processed according to the algorithm described in Section 2.2. 17
2.7 Raw **wind power** data for **Stor-Rotliden** processed according to the algorithm described in Section 2.2. 18
2.8 Missing data percentage out of the yearlong, processed wind farm datasets applied as forecast reference in Chapter 3 and appendices A, B. 23

3.1 Observed [103] vs. WRF modelled terrain height above mean sea level for 1.1 km grid points closest and $6th$ closest to the average of the 40 coordinates of the wind turbines constituting the Stor-Rotliden wind farm. 36
3.2 Resolution and output frequency employed in WRF model simulations. . 40

Abbreviations

3DVAR	*Three-dimensional variational data assimilation*
4DVAR	*Four-dimensional variational data assimilation*
AGL	*Above ground level*
AR(x)	*Autoregressive model of order x*
BIC	*Bayesian information criterion*
CET	*Central European Time*
CSI	*Critical success index*
DART	*The Data Assimilation Research Testbed* (NCAR)
ECMWF	*European Centre for Medium-Range Weather Forceasting* (UK)
EnKF	*Ensemble Kalman filter (data assimilation)*
ETS	*Equitable threat score*
FDDA	*Four-dimensional data assimilation*
GCM	*General circulation model*
GFS	*Global Forecast System* (NCEP)
IC	*Initial condition*
IFS	*Integrated Forecast System* (ECMWF)
LAM	*Limited area (weather) model*
LBC	*Lateral boundary condition*
LLS	*Linear least square*
MAE	*Mean absolute error*
MetUM	*UK Meteorological Office Unified Model*
MM5	*The 5th generation PSU-NCAR mesoscale model*
MOL	*Monin-Obukhov length*
MOS	*Model output statistics*

NCAR *National Center for Atmospheric Research* (USA)

NCEP *National Centers for Environmental Predictions* (USA)

NWP *Numerical weather prediction*

PBL *Planetary boundary layer*

PCC *Pearson correlation coefficient*

PDE *Partial differential equation*

PSU *Pennsylvania State University* (USA)

RAMS *Regional Atmospheric Modeling System*

RMSE *Root mean square error*

TKE *Turbulent kinetic energy*

WPPT *Wind Power Prediction Tool*

WRF *Weather Research and Forecasting (model)* (NCAR)

UTC *Coordinated Universal Time*

Nomenclature

Wind speed	w	forecasted by GCM or LAM
Wind speed	w	wind turbine averaged anemometer readings
Predictor	$x/z/w$	an independent variable in a statistical model
(Model) response	W	dependent variable in a statistical model
Statistical model coefficient	θ_i	for the ith predictor (except in Figure 3.1)
Statistical model error	ε	a stochastic quantity
Statistical model residual	ϵ_i	for the ith time series data point
LLS wind speed	w_{LSS}	linear corrected forecasted wind speed
LLS error	$w_{LLS} - \mathsf{w}$	forecast error post bias/scaling error-removal

Journal Publication Drafts

Title to be submitted	to journal
Limited area forecasting for wind energy scheduling. Part I: Performance assessment and enhancement studies	**Wind Energy**
Limited area forecasting for wind energy scheduling. Part II: Predictability of wind speed variability and uncertainty	**Wind Energy**
Correction of NWP model deficiencies by statistical postprocessing	**Quarterly Journal of the Royal Meteorological Society**

Chapter 1

Introduction

Unexpected wind power fluctuations impede the transmission system operator's ability to balance supply and demand and may entail substantial costs for wind energy companies [1], which therefore rely on numerical weather prediction (NWP) of wind speed for operational wind energy scheduling to forestall events potentially harmfull to business. Wind speed is considered one of the most challenging meteorological parameters to predict [2], yet the accuracy of wind speed forecasts has direct implications for efficient integration of wind energy [3]. A concise overview of wind power forecasting methods anno 2010 is given in Foley *et al.* [4], updated in their 2012 paper [5]. Foley *et al.* [4, 5] classify methods as either statistical, physical or hybrid. A range of statistical modelling techniques based on measurements are employed, though without weather model input all of these are only feasible for prediction a few hours ahead, as pointed out e.g. in Giebel *et al.* [6, 7]. More comprehensive in scope than can be achieved in a journal publication format, as in Foley *et al.* [4, 5], the work of Giebel *et al.* [8] is a technical report from 2011 on the state of the art in wind power prediction, thus providing an extensive review of the latest research results on the topic. Physical methods encompass the field of NWP as well as downscaling from NWP model output to power forecasts, calibrated to wind farm sites of interest based on physical principles and the local topography. An example closer to the former is the adaptive and recursive semi-parametric wind power prediction tool (WPPT) [9, 10], whereas the Previento model [11, 12] to a greater extent is based on physical principles. Common for all wind power forecasting methods feasible in practice for day ahead wind energy scheduling is the dependence on input from NWP model data [6, 7].

Global NWP models, or general circulation models (GCMs), are often unable to adequately resolve local topography forcing on surface layer atmospheric dynamics in the vicinity of a site of interest in complex terrain, as well as meso- and microscale meteorological features of importance to weather forecast end-users. Limited area models (LAMs) describe atmospheric dynamics for a specified geographical region only, taking initial conditions (ICs) and lateral boundary conditions (LBCs) from GCM data. Therefore higher spatio-temporal resolution can be accommodated for reasonably attainable computational resources, hence making LAM simulations feasible for end-users to carry out themselves, or outsource to third parties, if accurate forecasts covering a certain geographical region are needed. End-users may be interested in customising the spatio-temporal resolution, sub-grid scale physical process parameterisations, output variables for decision model input, or apply high-resolution topography data, in order to optimise the LAM configuration for the application at hand.

This work aims to quantify the extent to which limited area forecasting can increase wind power forecasting accuracy, and proposes an original statistical postprocessing framework for systematically correcting for NWP model deficiencies. A brief historical review of the scientific discipline of weather simulation is given next, followed by a discussion on different scales of atmospheric motion, leading to a literature review of limited area forecasting studies, and finally a summary of the research contribution presented in the thesis.

1.1 Historical rundown of numerical weather prediction

In 1904 Bjerknes [13, 14] framed the challenge of forecasting the weather by the following two statements:

1. One has to know with sufficient accuracy the state of the atmosphere at a given time.
2. One has to know with sufficient accuracy the laws according to which one state of the atmosphere develops from another.

He developed the idea of describing atmospheric dynamics solely on the basis of physical and mechanical laws [15] discovered since Euler's derivation of the general equations of

PROGNOSTIC MODEL FORMULATIONS FOR ATMOSPHERIC DYNAMICS SIMULATION

Hydrostatic	Nonhydrostatic
The primitive equations: Simplified equation for vertical motion	Relaxation of hydrostasis assumption: Unaltered equation for vertical motion
Less computation intensive	More computation intensive
Unreliable simulations for high model resolutions and for very complex terrain	Suitable for atmospheric flow modelling over complex terrain

TABLE 1.1: Hydrostatic and nonhydrostatic NWP model formulation traits.

motion from Newton's Second Law of Motion [16, 17]. Richardson formulated the atmospheric prediction equations into a form similar to that used in current NWP [18], and explains in his book from 1922 [19] how he performed weather prediction calculations based on these equations. A few years later the pioneering work of Courant, Friedrichs and Lewy on the discretisation of partial differential equations (PDEs) [20, 21] laid a sound theoretical foundation for solving PDE systems in practice, an endeavour facilitated with the advent of the Electronic Numerical Integrator and Computer (ENIAC) in 1946, on which the first machine-aided NWP experiments were carried out in the following years. In the late 1940s Charney [22] and Eady [23] derived models for baroclinic instability that largely explain observations of midlatitude synoptic fluctuations not directly related to diurnal solar heating. Charney reasoned that largescale atmospheric dynamics can only be analysed in practice if approximations to separate high-frequency gravity- and sound waves from largescale atmospheric motion are made [22, 24]. In 1950, Charney, Fjörtoft and von Neumann [25] first publicised application of the electronic computer for weather prediction, integrating in time the barotropic vorticity equation numerically. This achievement was the first demonstration of NWP execution on a machine based on spatially discretised forms of mathematical relations originating from hydro- and thermodynamics theory, integrated as a stepwise process in time. During the 1950s attention turned first to NWP models for simple baroclinic flow [26] before the primitive equations were considered [27, 28].

NWP models can roughly be categorised in the four combinations spanned by hydrostatic/nonhydrostatic and shallow/deep-atmosphere approximations [29]. See Table 1.1 and e.g. Xu *et al.* [30], Janjic *et al.* [31], Saito *et al.* [32], Staniforth and Wood [33] for an overview of mathematical and numerical aspects of nonhydrostatic NWP, and namely Saito [34] provides a meticulous overview of various formulations of the governing equations for atmospheric dynamics and particular nonhydrostatic NWP models in

use worldwide operationally and for research purposes — including the LAM considered in this work.

1.2 Spatio-temporal scales of atmospheric dynamics

An atmospheric dynamics scale reference commonly referred to in the literature is proposed in Orlanski [35], reproduced in Table 1.2. A more recent definition can be found in Lackmann [36], reproduced in Table 1.3.

The weather experienced by an observer on the ground is a superposition of dynamics across scales, from polar front jet streams sustained by planetary rotation and solar heating to boundary layer turbulence due to vertical wind speed shear and thermal stratification. The former is a westerly flow that reaches wind speeds of ∼150 km per hour along strong upper troposphere pressure gradients and thus does not directly affect surface weather conditions important for wind energy applications, contrary to weaker low level jet-like weather phenomena that occur under certain conditions, e.g. barrier jets along mountain ranges, as well as other types of offshore and onshore persistent horizontal flow at wind turbine height [37, 38]. At spatial scales less than a few hundred meters the average kinetic energy spectrum of lower atmosphere horizontal motion is understood to be generated by isotropic three-dimensional turbulence, cf. Kolmogorov [39], while at spatial scales greater than a few hundred kilometers, Charney [40] predicts a dependence on the wavenumber associated with horizontal velocity to the minus third with his theory on geostrophic turbulence, later verified by Gage and Nastrom [41] among others. On the intermediate scale, the mesoscale (as defined in Table 1.3), which

Scale	Range	Examples of atmospheric dynamics
Microscale-γ	< 20 m	Turbulence, plumes, roughness
Microscale-β	20-200 m	Dust devils, thermals, wakes
Microscale-α	200-2000 m	Tornadoes, short gravity waves
Mesoscale-γ	2-20 km	Thunderstorm convection, complex terrain flows, urban effects
Mesoscale-β	20-200 km	Nocturnal low-level jets, cloud clusters, sea breezes
Mesoscale-α	200-2000 km	Fronts, low-pressure systems, hurricanes
Macroscale-β	2000-20000 km	Baroclinic waves
Macroscale-α	2000-20000 km	Tidal waves

Table 1.2: Characteristic atmosperic dynamics scales as proposed by Orlanski [35].

Scale	Length [km]	Time	Examples of atmospheric dynamics
Microscale	<1	< 1 hour	Turbulence, planetary boundary layer
Mesoscale	1-100	1 hour to 1 day	Thunderstorm, land-sea breeze
Synoptic	1000-6000	1 day to 1 week	Upper-level troughs, ridges
Planetary	>6000	> 1 week	Polar front jet stream, trade winds

TABLE 1.3: Characteristic atmosperic dynamics scales as proposed by Lackmann [36].

is relevant to weather forecasting for wind energy scheduling, there are two prevalent theories. On one hand, VanZandt [42] and Dewan [43] suggest that observed kinetic energy spectra are created by gravity waves, while Gage [44] and Lilly [45] highlight evidence for quasi-two-dimensional turbulence. Cho *et al.* [46] investigate whether horizontal wind fluctuations on the mesoscale, as defined in Table 1.3, are mainly due to the former or the latter theory, and find stronger evidence for the quasi-two-dimensional turbulence theory in their study.

Lorenz [47] derives, in his 1969 paper, theoretically a pessimistic limit on atmospheric predictability of ∼1 hour for a horizontal resolution of about 40 km, based on a two-dimensional form of the vorticity equation. In practice, however, the error growth were later shown to be much slower for mesoscale prediction than anticipated by Lorenz [47], see e.g. [48–52], making day ahead NWP feasible for wind energy scheduling. Numerical weather modelling approaches feasible on the mesoscale are well-established, while sub-kilometer scale turbulence poses a severe challenge for weather simulation on the microscale [53]. Skamerock [54] defines effective resolution of a NWP model as the wavenumber at which the kinetic energy spectrum of the model starts to be damped relative to corresponding observations, cf. [41, 55–57]. Based on Weather Research and Forecasting (WRF) model [58, 59] simulations with computational grids of 22, 10 and 4 km horizontal resolutions, Skamarock [54] estimates the effective spatial resolution of the WRF model to be 7 times as coarse as the specified model resolution.

The study presented in Chapter 3 is based on WRF model version 3 [60, 61] simulations.

1.3 Wind speed forecast dependence on model resolution

Worldwide GCM forecasts are generated daily, e.g. by spectral NWP models such as the European Integrated Forecast System (IFS) [62] and the Global Forecast System (GFS)

[63] in the USA, or based on finite differences as the UK Meteorological Office Unified Model (MetUM) [64], though restrictions on computational resources limit model resolutions attainable. The accuracy of local predictions near the Earth's surface depends on the ability of a NWP model to resolve atmospheric processes on a range of spatio-temporal scales, and the local topography must be represented on a scale consistent with the spatial resolution of the NWP model. In order to further increase the resolution of weather simulations operationally a LAM must be employed, initialised based on assimilated realtime atmospheric measurements and/or taking as IC spatial meteorological fields provided by a GCM used to force the outermost computational grid laterally for the duration of the LAM forecast. The feasibility of such configurations depends on the resolution required to resolve the features of interest with sufficient accuracy, computational resources available, and the forecast leadtime horizon of interest. However, before reviewing studies pertaining to operational forecasting WRF model performance in the context of climatology simulation in considered, for completeness.

Jiménez *et al.* [65] study the WRF model's [60] ability to reproduce the wind field of six typical wind patterns for the northeastern part of the Iberian Peninsula. They extend the study by evaluating the accuracy of the wind field simulations conditioned on synoptic pressure patterns, as identified in previous work by Jiménez *et al.* [66]. They demonstrate that regional details unavailable in the IFS-based reanalysis forcing data (ERA-40 and ERA-Interim) are simulated by WRF with acceptable accuracy. Wind speed is underestimated at mountain tops and hills, while overestimated at plains and in valleys. Wind direction scores higher at sites at hills and plains than in valleys. Apart from the dependence on the accuracy of the synoptic scale forcing data, Jiménez *et al.* [65] attribute WRF model errors to IC and LBC issues, as well as topography missrepresentation in the WRF model. Horvath *et al.* [67] find that the WRF model [60] generally overestimates the mean wind speed, especially during the night, whereas the model's predecessor; the fifth-generation Pennsylvania State University – National Center for Atmospheric Research Mesoscale Model (MM5), underestimates winds in general. The experiments span horizontal computational grid points 27 km down to $\frac{1}{3}$ km apart and show that numerical scales down to at least near-kilometer grids are necessary for wind speed climate downscaling over the complex terrain studied. Hahmann *et al.* [68] study the wind climate over the North- and Baltic seas using the WRF model [60] and meteorological tower- and lidar measurements for validation. They demonstrate

clear inferiorness of ERA-Interim data relative to WRF model simulations initialised and laterally forced with said reanalyses. The annual wind speed bias of the WRF model data is small, though the wind speed variability is lower than that for observations. This is a well-known NWP model issue [54]. The results indicate that the WRF model spin up time over land is larger than 12 hours for the model configuration employed. The spin up time is the time atmospheric dynamics represented on the LAM grid needs in order to "recover" from the discretisation-shock induced due to the interpolation from coarserly gridded forcing data during initialisation, as well as differences in governing equations approximations, discretisation schemes, and sub-grid scale parameterisations [52]. LAM initialistion is discussed further in Section 3.1. Laprise *et al.* [69] carry out two LAM experiments, based on the Canadian Regional Climate Model (CRCM) [70], with forcing from 6 hourly analyses generated by the Canadian Meteorological Centre [71] and synthetic "GCM" data; i.e. CRCM runs for a large region and subject to filtering of short scales (shorter than 500 km). The experiment suggests that a LAM cannot predict finescale structures deterministically if the information is not explicitly present in both IC and LBCs.

Deppe *et al.* [38] briefly touch upon the subject of LAM resolution comparison for day ahead wind speed forecasts, yet decide to drop their WRF model [60] configuration's computational domain of 4 km horizontal resolution early on in the study, as data extracted from the parent 10 km grid exhibits lower wind speed mean absolute error (MAE), based on 80 m above ground level (AGL) mast measurements. Another study more thoroughly investigating the effect of LAM resolution-increase on wind speed and power forecasts is that of Lazić *et al.* [72]. They examine day ahead forecast performance of a 22 km horizontally resolved computational domain with 45 vertical layers providing ICs and LBCs for a nested domain of 4 km horizontal resolution and 60 vertical layers, using the Eta model [73–76] with integration time steps 72 s and 12 s, respectively. ICs for the coarse Eta model domain are taken from the US National Centers for Environmental Prediction (NCEP) global reanalysis which is not available in realtime – i.e. the experiment does not resemble an operational forecast situation – and the LBCs are obtained from the GFS. Lazić *et al.* [72] thoroughly evaluate the performance of the high-resolution Eta model predictions of wind speed and power at different heights AGL, yet do not compare the performance metrics to corresponding values for the coarser Eta domain, nor global model data. Deppe *et al.* [38] and Lazić *et al.* [72] are fairly recent

studies from 2013 and 2010, respectively.

Going back 10–15 years, Rife *et al.* [77] compare wind speed forecasts from four different NWP models, namely the Rapid Update-Cycle (RUC) and Eta models, both at 40 km horizontal resolution, as well as 1° data from the GFS and MM5, the latter with 1.33 km horizontal resolution. They find no significant benefit from the high-resolution MM5 data in terms of wind speed. However, anomaly correlation results illustrate unambiguous improvement of the fine MM5 grid compared to forecasts from the other models. In a study published in the following year, Rife and Davis [78] return to the question of whether high-resolution NWP skill increases relative to corresponding coarser resolved wind forecasts. They hypothesise that traditional forecast verification metrics are not suitable estimators of "value" added from increased NWP model resolution as these sharply penalise spatio-temporal deviations from observations. Using MM5, the predecessor of the WRF model [60] used in this work, wind speed forecast performance of two model configurations with 30 km and 3.3 km horizontal resolutions are compared in terms of root mean square error (RMSE), anomaly correlation, ramp objects, and equitable threat score (ETS) for dichotomous temporal variance data. Overall, Rife and Davis [78] find incremental improvements from the high-resolution MM5 forecasts, but the skill relative to the coarser MM5 runs increases when temporal dilation is allowed; less so for the slight tolerance introduced in the wind speed ramp object verification and more so for the full-forecast-span dichotomous variance ETS metric. It should be mentioned that Rife and Davis [78] do not nest the high-resolution domain within the coarse resolution domain; the two runs are from separate MM5 configurations both predicting the atmosphere over the reference sites, though with computational domain borders outlining different geographical regions. Colle and Mass [79] find improvement in the simulation of downslope winds in a mountaneous region when increasing MM5 horizontal resolution from 27 km to 1 km. Expanding the range of considered meteorological variables beyond wind, Mass *et al.* [80] address the question "Does increasing horizontal resolution produce more skillful forecasts?" anno 2002. First, they review the latest results around the time when research intensified for spatial scales beyond hydrostatic-approximation validity, that is; scales where nonhydrostatic formulations [30–33] of the governing equations are necessary and the adverse effects of convection parameterisation begin to outweigh the benefits [81–83]. Mass *et al.* [80] then present results from a multiyear verification of mesoscale forecasts in the Pacific Northwest, including a case

study in which the model resolution is varied. Based on ∼2 years of twice-daily 48 hour MM5 forecasts for 36 km, 12 km and 4 km horizontally resolved computational domains, they find consistent, albeit small, wind speed MAE reduction when going from 36 km to 12 km, though not when going from 12 km to 4 km resolution, and the 4 km wind speed results are by far the most biased. For wind direction, both MAE and bias consistently improve with resolution, although very little from 12 km to 4 km. For their case study, a frontal passage is studied and the spatial structure is clearly approaching that of radar observations as model resolution increases. The front does, however, make landfall approximately 1 hour early, which severely affects traditional scores, as also discussed in Rife *et al.* [78]. Overall, many studies do provide evidence that increasing NWP model resolution do indeed better describe the fundamental physics of the atmosphere. McQueen *et al.* [84] find improvement to the realism of meteorological structures over land when increasing Regional Atmospheric Modeling System (RAMS) forecasts from 10 km to 2.5 km horizontal resolution, due to the better topography representation of the high-resolution domain, as do Davis *et al.* [85] when increasing MM5 horizontal resolution from 10 km to 1.1 km. By increasing the horizontal resolution of University of New South Wales High Resolution Limited-Area Atmospheric Prediction (HIRES) model forecasts from 50 km to 10 km, Buckley and Leslie [86] demonstrate greatly improved accuracy of wind and pressure predictions for a storm case.

1.4 Research contribution

As stated above, the work presented in the thesis concerns limited area forecasting and statistical modelling for wind energy scheduling.

Cf. Section 1.3, a number of studies have looked into how increased LAM resolution affects forecast accuracy for wind speed and other meteorological variables. The work presented here differs from part of the literature reviewed on the following points.

- The experiment is carried out subject to realistic operational conditions

LAM forecasts are generated solely using information that would be available operationally and thus resemble a realistic forecasting situation, contrary e.g. to the case in the Lazić *et al.* [72] study where reanalysis data is used for LAM initialisation.

- Type of reference measurements used for NWP model performance assessment

In this work, the wind speed measurements forming the basis of forecast assessment are farm-averaged anemometer readings from wind turbine nacelles. In comparison, data from measurement masts in the vicinity of- or inside a wind farm is less robust to topography and wind farm wake effects for arbitrary wind directions. Another argument pro the use of farm-averaged observations for comparison to NWP model data is the fact that a computational grid point represents a spatio-temporal mean of local atmospheric dynamics, depending on the choice of grid constants and integration time step. As such, comparison of NWP wind speed data to a farm-average of wind speed measurements from each wind turbine, as would represent the spatial extent of the wind farm, is arguably more reasonable than comparison to e.g. mast measurements, which only represent a single point within or outside the spatial extent of the wind farm. Finally, many previous studies use 10 m AGL wind observations which are plausibly corrupted more by turbulence due to surface roughness than data measured at wind turbine hub-height — where the wind power is harvested.

- Forecast leadtime horizon differentiation

Contrary to all reviewed literature on the topic, traditional forecast performance metrics are here consistently evaluated separately for three forecast leadtime horizon groups, together spanning the day ahead time span relevant for operational wind energy scheduling. Thus, not only the effect of LAM resolution increase is examined, but also the relative forecast performance decrease as a function of the leadtime is resolved, and each reference data point is represented once within each forecast leadtime group.

- A non-traditional scalar accuracy metric for forecast performance assessment

Based on the review of Section 1.3, as well as other literature related to wind power prediction [8, 87], traditional forecast scores include bias, MAE and RMSE. Bias for a NWP modelled meteorological variable of interest is, naturally, very important to quantify if one needs accurate predictions of said variable for a given application. Knowledge of the systematic deviation from corresponding atmospheric observations allow for correction prior to further transformation of the variable, as e.g. when NWP model forecasts of

wind speed are transformed to power for day ahead quantification of the amount of energy that can be delivered by a given wind farm. Provided wind speed observations are continually available, realtime bias correction based on e.g. a few months data leading up to the forecast initialisation time can be integrated in wind power prediction models [88]. Metrics like MAE and RMSE penalise bias and are therefore misleading forecast performance evaluation quantities for NWP data providing input for bias-correcting wind power prediction models. A metric such as Pearson's correlation coefficient (PCC) is robust to systematic over- or underprediction, and is thus a more pure measure of NWP model ability to time wind speed fluctuations correctly. Therefore, the PCC is the primary forecast performance metric in Chapter 3 and Appendix A, rather than the traditional MAE and RMSE metrics.

- LAM resolution comparison is based on nested computational domains

In contrast to e.g. the Rife *et al.* [77] and Rife and Davis [78] studies, LAM resolution comparison is here based on the same model configuration, consequetively nested babushka dollwise. I.e. the first nested LAM domain is forced laterally from the coarsest LAM domain, the second nest is forced laterally from the first nest, etc. thereby enabling a somewhat more fair comparison than if e.g. domain borders of one out of two or more different configurations traverse more challenging terrain than the other model configuration(s), thus being more prone to induce differing adverse effects due to e.g. spurious gravity waves amplified by the shortcomings of the ill posed relaxation LBC formulation [89] typically employed in LAMs [52].

- Several other atmospheric variables than wind components are considered in statistical modelling for weather model prediction postprocessing

Few studies reviewed by the author go beyond including a handfull of NWP modelled variables as predictors in statistical postprocessing frameworks to improve wind speed forecasts. Lange and Focken [90, 91] relate wind speed forecast RMSE to meteorological situations, using principal component- and cluster analysis, and Vincent and Hahmann [92] use Kohonen's Self-Organising Map approach [93] to arrive at 36 large scale weather pattern categories of which a few are identified to be associated with severe wind variability at the offshore Horns Rev wind farm, which is also studied in this work. Both

studies restrict attention to mean sea level pressure in the classification procedures. Ranaboldo *et al.* [94] consider temperature, pressure, planetary boundary layer height, heat flux, Monin-Obukhov length, and Richardson number in their model output statistics (MOS) [95] framework. Instead of the forward stepwise selection [96] of predictors employed in Ranaboldo *et al.* [94], the original MOS postprocessing framework proposed in Appendix C is based on backward elimination.

1.5 Software tools

As mentioned last in Section 1.2, the LAM study of Chapter 3 and appendices A, B is based on WRF model [60] simulations. The R language [97] is used extensively throughout the thesis for data analysis, statistical modelling, and visualisation of results. Computational grids considered in WRF model simulations cf. Appendix A and the wind farm sites map in Appendix C are sketched using the NCL language [98], as is the orography plot of Figure 3.3.

Zhu and Genton [88] review the challenges related to translating weather modelled wind speed forecast time series to wind power predictions, and several methods are found in the literature. In this work, the conversion of NWP model data to wind power is carried out using the commercial WPPT [9, 10] mentioned above. As described e.g. in the concise overview provided in Cutler *et al.* [99], the WPPT estimates, based on a few recent months' wind power measurements and forecast data, parameters in a wind power curve model with forecasted wind speed and direction as predictors — with separate model parameter fits for each forecast leadtime step to account for deteriorating NWP model performance with forecast integration time. The prediction model is a realtime data-driven weighting between the current and previous wind power forecasts, the power curve model response, diurnally varying terms, and a bias term.

Chapter 2

Forecast reference data

Oreskes *et al.* [100] discuss the concepts of verification and validation in the context of Earth system modelling, and emphasise that the primary value of models is heuristic; useful tools perhaps, but not susceptible to proof. Natural systems are never closed. The challenge outlined in Chapter 1, however, is not concerned with underlying "truth" in nature, but instead has a rather practical scope: *To what extent can limited area forecasting and statistical modelling facilitate wind energy integration in the power grid?*

Wind turbine anemometer readings averaged farm-wide constitute forecast reference data throughout this work, along with corresponding power production time series. The three sites considered are subject to offshore, coastal and inland weather conditions and thus cover a wide range of terrain types. Compared to measurement tower observations, farm-averaged wind turbine anemometer readings

1. spatially approximates the *spatio-temporal computational grid cell mean* wind speed that NWP simulations of atmospheric flow represent, and
2. are for arbitrary wind directions more robust to adverse effects due to wind farm wakes and topography complexity and/or misrepresentation in the NWP model.

E.g. Cutler *et al.* [101] find that power curve models based on farm-averaged nacelle measurements of wind speed results in more accurate power predictions than if mast measurements are used, even when calibration corrections are made to the mast measurements. N.b. that *measurement tower* and *meteorological mast* is used interchangably in this work to reflect the widespread use of both terms in the literature.

	Horns Rev	Rejsby Hede	Stor-Rotliden
Local terrain type	Offshore	Coastal	Inland
Hub-height [m]	70	45	95
Coordinates	55.5297° north, 7.9061° east [102]	55.0373° north, 8.6699° east [103]	64.2166° north, 18.3829° east [104]
Number of turbines	80	39	40
Turbine make & capacity [MW]	Vestas V 80 [105] 2 (all turbines) [105]	Micon [106] 0.6 (all) [106]	Vestas V 90 [107] 2 (29), 1.8 (11) [104]
Farm capacity [MW]	160	23.4	77.8
Commissioned [year]	2002	1995	2011

Table 2.1: Wind farm data for the forecast reference sites considered in this work. The wind farms are operated by the Swedish energy company *Vattenfall*.

Wind farm data is listed in Table 2.1, and the structure of files describing wind farm measurements is reviewed and associated issues are discussed and resolved. Data processing steps are argued for and applied sequentially, accompanied by tables and graphical assessment of the effect of a given processing action on the data. The result is a measurement dataset that constitute a suitable reference for forecast validation and statistical modelling in the studies presented in the following chapters and appendices.

The shortest dataset for sites listed in Table 2.1 is that for Stor-Rotliden, commencing in April 2012. As the GCM used to initialise LAM forecasts cf. Chapter 3 was upgraded 22 May 2012, and henceforth includes e.g. 80 m and 100 m AGL wind speed variables, the study period chosen w.r.t. the studies carried out in Chapter 3 and appendices A, B spans 23 May 2012 through 23 May 2013, thus covering all seasons.

N.b. that Chapter 2 concerns the time period relevant to Chapter 3 and appendices A, B. The raw wind farm measurements underlying reference data applied in Chapter 4 and Appendix C, spanning the time period from 1 October 2012 to 30 September 2014, is processed according to the procedure outlined in the this chapter.

2.1 Wind farm data vs. measurement tower data

The findings of Cutler *et al.* [101] are elaborated on and discussed in the following. As mentioned they find, when used in conjunction with NWP model data, that time series of farm-averaged wind turbine nacelle measurements of wind speed consistently outperform measurements from nearby meteorological masts for translating atmospheric flow to wind power production. Although flow disruption by wind turbine blades upwind partially distorts anemometer readings at the wind turbine nacelle, the study [101] indicates

that the limited spatial representativeness of local conditions offered by meteorological mast readings is a greater liability. A forecast user may be more interested in weather conditions in the vicinity of particular coordinates, as reflected by measurement tower data, yet highly accurate point forecasts are not reasonable to require from a global NWP model representing atmospheric dynamics on a discrete grid on which the spatio-temporal continuum of actual atmospheric dynamics is approximated. Smoothing the time series spatially, e.g. by averaging individual wind turbine anemometer readings, facilitates more consistent comparison to the four-dimensional grid cell average NWP modelled quantities represent, in that the study becomes more robust to sensitivity to topography and other sources of local wind field heterogeneity that mast measurements would be more vulnerable to.

2.2 Algorithm for processing of raw wind farm data

The raw data files for the three wind farms share the same technical quirks[1], and hence a single processing function is developed and applied to the data files for each site, respectively.

- Instantaneous farm-averaged wind speed measurements and wind power production are available as 5 minute time series, respectively, and these time series are averaged to 10 minute values to match the LAM output frequency employed in Chapter 3 and appendices A, B.

- To ensure fairly homogeneous inter-turbine shadowing and spatial representation of wind turbines throughout the dataset an arbitrary limit of one third of farm capacity is set, below which data is discarded. N.b. that wind turbine wakes for wind farm capacities above this heuristic lower limit constitute an error source in the wind *speed* forecast reference datasets used in the following chapters and appendices, though in terms of wind *power* forecasts by the WPPT any directional bias is addressed in the power curve model, cf. Section 1.5.

[1] Time stamp/file name issues at transition between CET winter and summer time.

Missing data time period	Hours duration	Percentage
2012-09-26 15:50:00 to 2012-09-27 00:00:00	8	
2012-10-08 21:25:00 to 2012-10-09 00:00:00	3	
2012-10-19 12:30:00 to 2012-10-20 00:00:00	12	
Average duration of outage	8	
Median duration of outage	8	
Missing data due to flat period removal		0.5
Total missing data		0.8

TABLE 2.2: Raw **wind speed** data for **Horns Rev** processed according to the algorithm described in Section 2.2.

Missing data time period	Hours duration	Percentage
2012-06-28 00:00:00 to 2012-06-29 00:00:00	24	
2012-09-26 15:50:00 to 2012-09-27 00:00:00	8	
2012-10-08 21:25:00 to 2012-10-09 00:00:00	3	
Average duration of outage	12	
Median duration of outage	8	
Missing data due to flat period removal		2.7
Total missing data		3.1

TABLE 2.3: Raw **wind speed** data for **Rejsby Hede** processed according to the algorithm described in Section 2.2.

Missing data time period	Hours duration	Percentage
2012-06-28 00:00:00 to 2012-06-29 00:00:00	24	
2012-09-26 15:50:00 to 2012-09-27 00:00:00	8	
2012-10-08 21:25:00 to 2012-10-09 00:00:00	3	
Average duration of outage	12	
Median duration of outage	8	
Missing data due to flat period removal		4.7
Total missing data		5.1

TABLE 2.4: Raw **wind speed** data for **Stor-Rotliden** processed according to the algorithm described in Section 2.2.

- Records are kept on available wind farm capacity for any given time stamp, i.e. the fraction of wind turbines actively producing power times farm capacity, is used to scale produced power in order to compensate power data for periods with inactive wind turbines (temporarily reduced farm capacity). There are instances where the wind turbine availability scaling results in power values exceeding farm capacity. This is assumed to occur as a consequence of too pessimistic online capacity (i.e. more wind turbines than indicated are in fact producing power) and such overshot power values are reduced to farm capacity.

- From inspection of the raw data, wind speed and power measurements are sometimes observed to be exactly constant in time for one or more consecutive time

Missing data time period	Hours duration	Percentage
2012-06-20 15:35:00 to 2012-06-21 00:00:00	8	
2012-09-26 16:00:00 to 2012-09-27 00:00:00	8	
2012-10-08 21:25:00 to 2012-10-09 00:00:00	3	
2012-10-19 12:30:00 to 2012-10-20 00:00:00	12	
2012-12-24 21:40:00 to 2012-12-25 00:10:00	2	
2012-12-29 05:10:00 to 2012-12-29 05:50:00	1	
2013-02-25 15:40:00 to 2013-02-25 15:45:00	0	
Average duration of outage	5	
Median duration of outage	3	
Missing data due to flat period removal		0.7
Total missing data		1.1

TABLE 2.5: Raw **wind power** data for **Horns Rev** processed according to the algorithm described in Section 2.2.

Missing data time period	Hours duration	Percentage
2012-06-20 15:35:00 to 2012-06-21 00:00:00	8	
2012-08-13 11:50:00 to 2012-08-15 09:15:00	45	
2012-09-26 16:00:00 to 2012-09-27 00:00:00	8	
2012-10-08 21:25:00 to 2012-10-09 00:00:00	3	
2012-12-24 21:35:00 to 2012-12-25 07:10:00	10	
2012-12-26 01:20:00 to 2012-12-26 06:15:00	5	
2012-12-30 23:45:00 to 2012-12-31 03:15:00	4	
2012-12-31 03:20:00 to 2012-12-31 06:15:00	3	
2013-01-30 23:05:00 to 2013-01-31 03:25:00	4	
2013-03-09 12:55:00 to 2013-03-09 15:45:00	3	
2013-04-28 13:20:00 to 2013-04-28 14:50:00	2	
Average duration of outage	9	
Median duration of outage	4	
Missing data due to flat period removal		0.2
Total missing data		1.3

TABLE 2.6: Raw **wind power** data for **Rejsby Hede** processed according to the algorithm described in Section 2.2.

stamps. Such unnatural behaviour is likely due to anemometer and/or wind turbine control software malfunction, and is removed from the dataset if wind speeds greater than 0.1 m/s or power values exceeding 1 % of farm capacity are observed for more than four consecutive time stamps.

Tables 2.2 through 2.7 show particular time periods for which data is missing/removed prior to filtering events of four consequetive constant values, along with the additional data point percentage removed due to more than four consequetive constant values. Finally, each table lists the total percentage of data points not available in forecast reference datasets post processing by the algorithm described in Section 2.2.

Missing data time period	Hours duration	Percentage
2012-06-20 15:35:00 to 2012-06-21 00:00:00	8	
2012-06-22 16:00:00 to 2012-06-25 09:00:00	65	
2012-07-01 15:20:00 to 2012-07-01 15:25:00	0	
2012-07-16 17:50:00 to 2012-07-16 22:45:00	5	
2012-07-17 16:15:00 to 2012-07-17 23:30:00	7	
2012-07-24 05:20:00 to 2012-07-24 17:15:00	12	
2012-07-25 05:40:00 to 2012-07-25 07:10:00	2	
2012-07-25 11:20:00 to 2012-07-31 13:15:00	146	
2012-08-08 06:35:00 to 2012-08-08 09:40:00	3	
2012-09-12 04:25:00 to 2012-09-12 13:55:00	10	
2012-09-26 16:00:00 to 2012-09-27 00:00:00	8	
2012-10-03 11:10:00 to 2012-10-03 18:50:00	8	
2012-10-08 21:25:00 to 2012-10-09 00:00:00	3	
2012-10-22 09:45:00 to 2012-10-22 14:00:00	4	
2012-10-23 07:30:00 to 2012-10-23 13:55:00	6	
2012-11-02 13:05:00 to 2012-11-02 14:40:00	2	
2012-11-26 10:30:00 to 2012-11-26 14:20:00	4	
2012-12-04 06:30:00 to 2012-12-04 19:00:00	12	
2012-12-19 02:40:00 to 2012-12-19 13:40:00	11	
2012-12-19 14:30:00 to 2012-12-19 14:35:00	0	
2013-01-01 04:55:00 to 2013-01-01 10:15:00	5	
2013-01-06 17:30:00 to 2013-01-06 22:30:00	5	
2013-02-05 17:30:00 to 2013-02-06 00:45:00	7	
2013-02-27 07:00:00 to 2013-02-27 10:00:00	3	
2013-03-17 22:35:00 to 2013-03-20 21:10:00	71	
2013-04-04 11:05:00 to 2013-04-04 11:35:00	0	
2013-05-06 16:00:00 to 2013-05-07 08:55:00	17	
2013-05-17 17:45:00 to 2013-05-17 20:00:00	2	
2013-05-22 01:50:00 to 2013-05-22 07:35:00	6	
Average duration of outage	15	
Median duration of outage	6	
Missing data due to flat period removal		0.7
Total missing data		5.6

TABLE 2.7: Raw **wind power** data for **Stor-Rotliden** processed according to the algorithm described in Section 2.2.

2.3 Detection of anemometer/power data inconsistency

Scatter density plots of anemometer readings vs. power production data are shown in Figure 2.1. Ideally, a slim power curve tracing out the wind farm transformation of kinetic energy in the atmospheric flow to electric energy is expected. However, the following digressions are readily observed:

- Several "tongues" of data points protrude from the Stor-Rotliden power curve to the high wind speed side. Given the inland location and high latitude of the site these are likely wind turbine blade icing events during which power production is reduced, cf. e.g. Davis *et al.* [109]. Figure 2.2 is a photo taken in cold weather

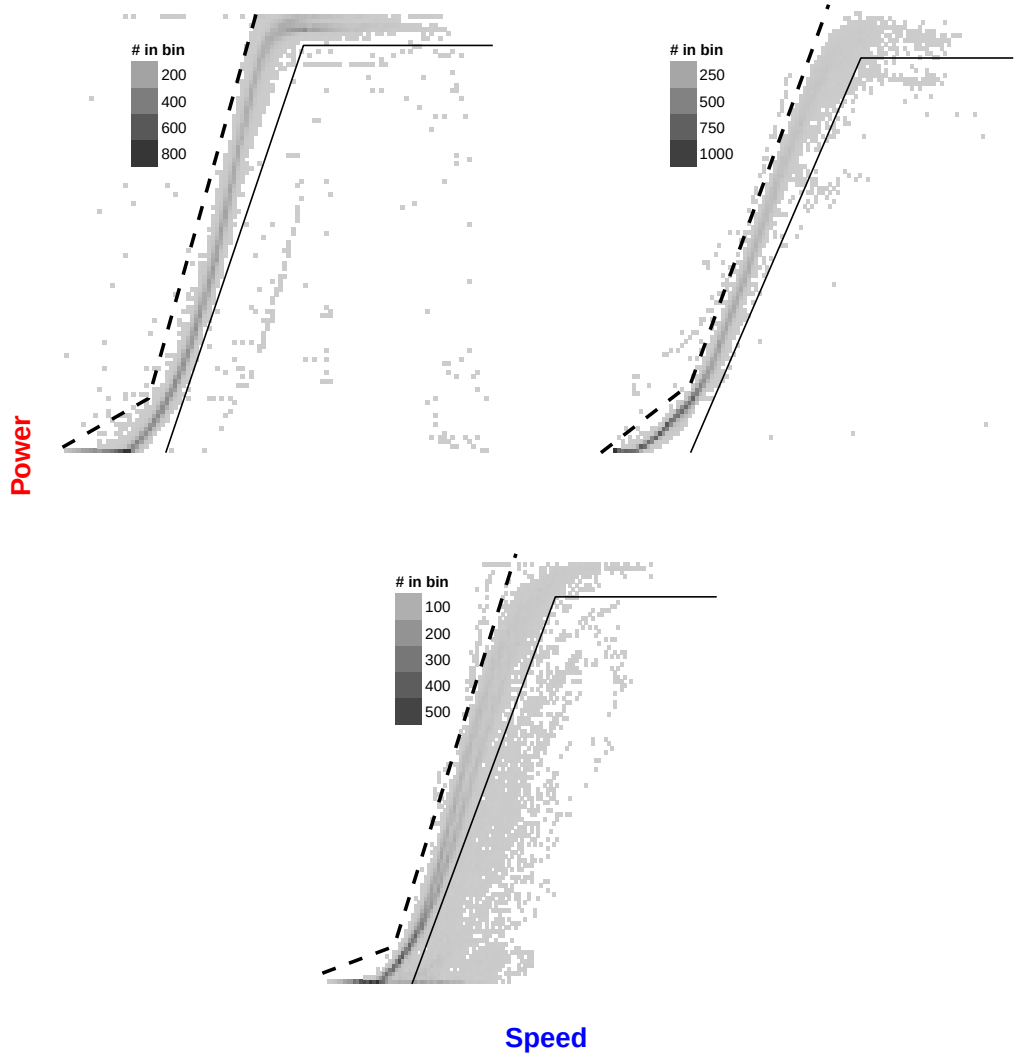


FIGURE 2.1: Anemometer readings vs. power production scatter density plots. The power curve region is heuristically framed by the dashed and solid lines. UPPER LEFT: Horns Rev. UPPER RIGHT: Rejsby Hede. BOTTOM: Stor-Rotliden.

and suggests that wind turbine blade icing events are plausible. *Suggested action*: Keep wind speed data, power data below the solid line is removed from the dataset.

- On the low wind speed side of power curves for Stor-Rotliden and Rejsby Hede a few data points trace out a curve roughly parallel to the power curves, as also observed on the high wind speed side of the Horns Rev power curve. This would occur if recordings of online wind turbine availability indicate less (more) capacity than that actually available at Stor-Rotliden and Rejsby Hede (Horns Rev). *Suggested action*: Keep wind speed data, power could be heuristically rescaled though

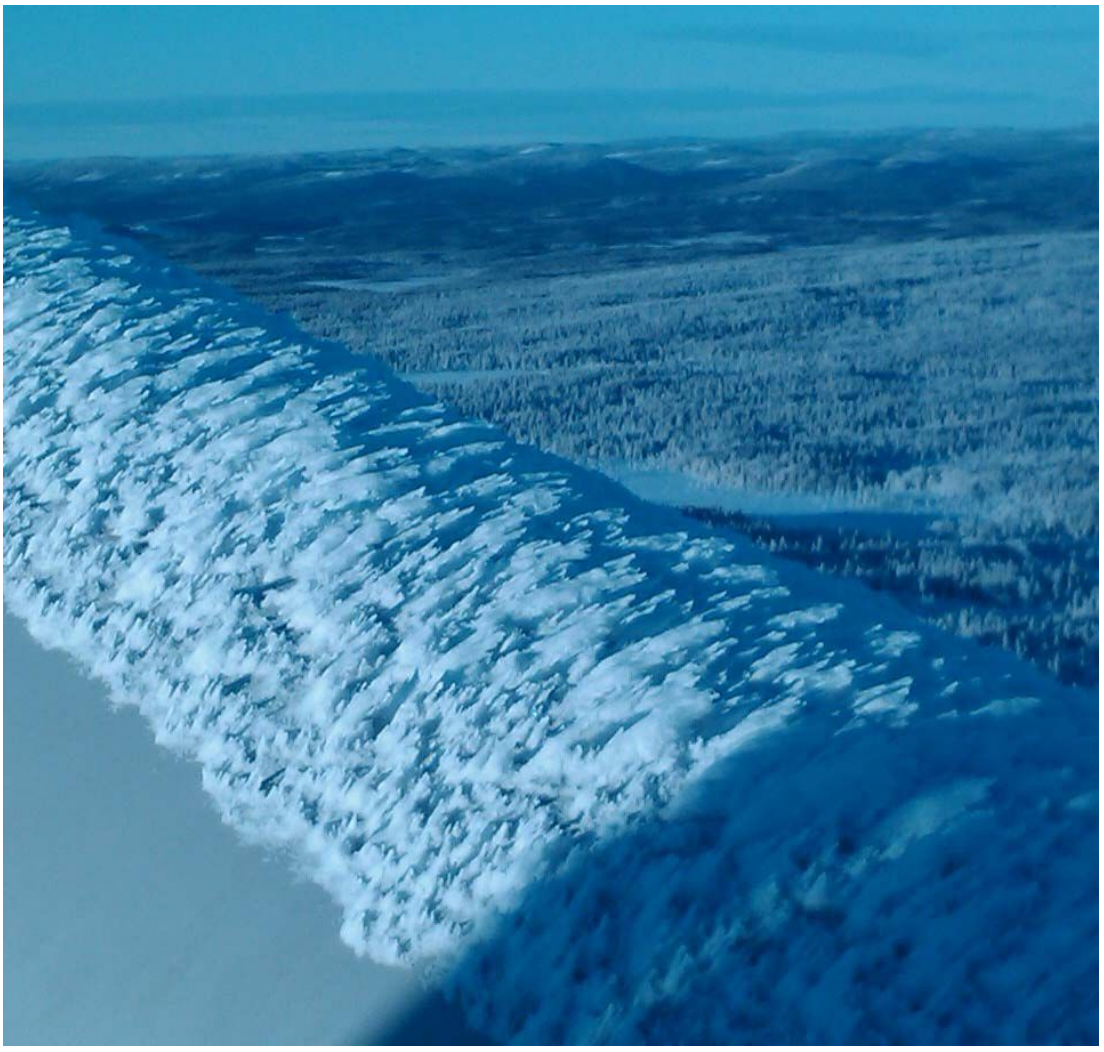


FIGURE 2.2: Photograph of a Stor-Rotliden wind turbine blade during an icing event [108].

due to the relatively few data points falling into this category such an endeavour is not worth the effort and the power data above the dashed line is instead removed.

- The data points on the low wind speed side of the Horns Rev power curve exhibit no reasonable temporal dependency nor scatter plot pattern and therefore the inconsistent scatter plot relation cannot be attributed to errors in neither wind speed nor power data. Likewise, there is an unexplainable lump of Rejsby Hede data points in the vicinity of the cut-in wind speed, where power data should be close to zero. *Suggested action*: Discard wind speed and power data above the Horns Rev and Rejsby Hede dashed lines.

- Upon closer inspection, the remaining data points digressing from the three power

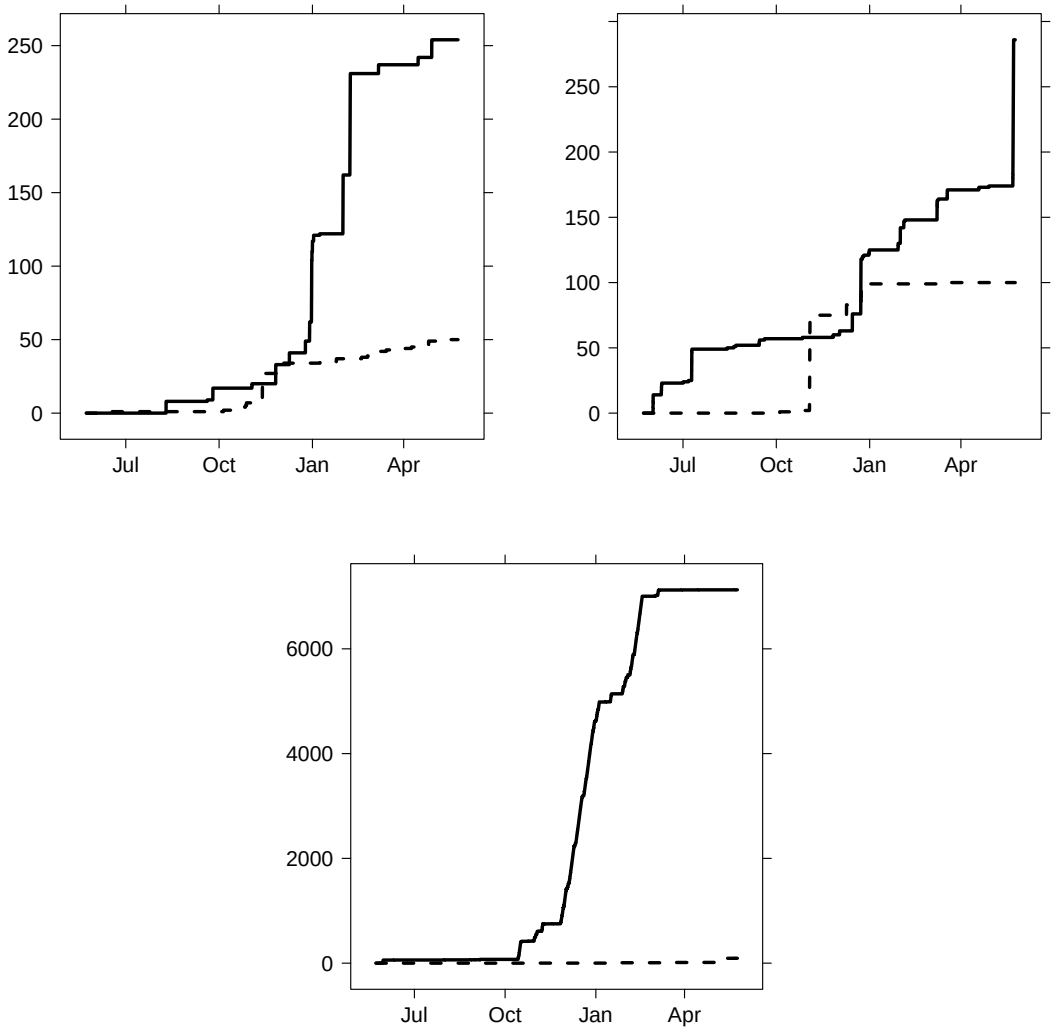


FIGURE 2.3: Cumulative count of data points falling outside the visually inferred power curve region framed by dashed and solid lines in Figure 2.1. UPPER LEFT: Horns Rev. UPPER RIGHT: Rejsby Hede. BOTTOM: Stor-Rotliden.

curves do suggest temporal and scatter plot relation. As such, erroneous availability flags and/or wind farm curtailment is assumed responsible by which wind speed measurements are unaffected. *Suggested action*: Keep wind speed data, discard power data.

Based on the heuristic arguments put forth above the following data processing is executed: As only few wind speed data points are sacrificed by removing all wind speed data above dashed lines, this is done in order to process wind farm data similarly for the three sites. All power data above (below) dashed (solid) lines is removed.

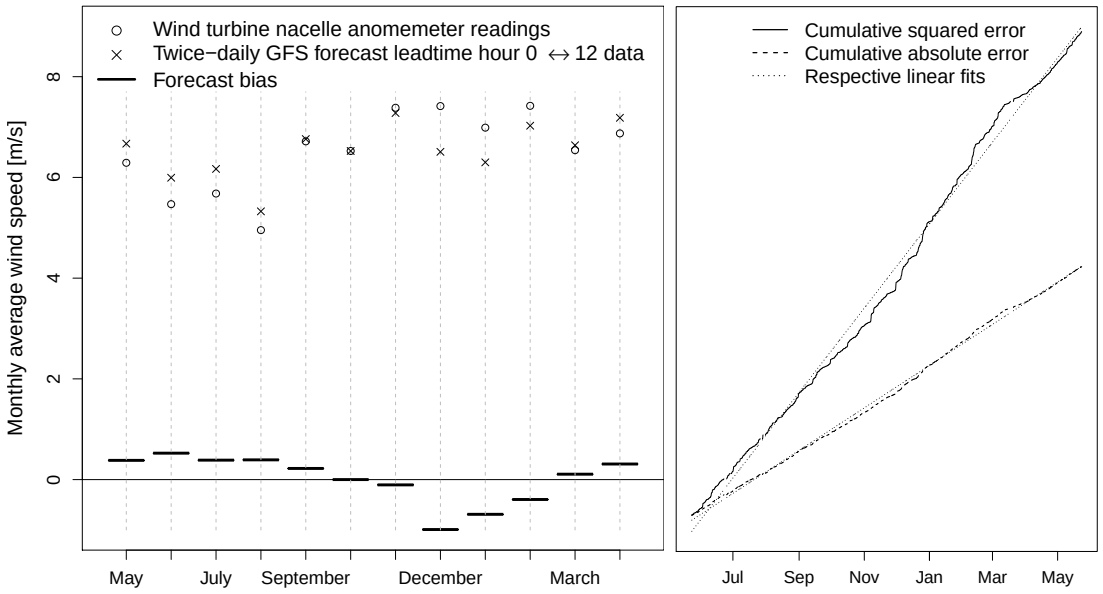


FIGURE 2.4: **N.B.: Full-year GFS bias and scaling error has been removed cf. *Appendix A Section 2.1*.** LEFT: Monthly mean of farm-averaged wind turbine nacelle anemometer data for Stor-Rotliden, along with forecast leadtime hour 0 ↔ 12 twice-daily GFS wind speed data, as well as the corresponding forecast bias. RIGHT: Cumulative squared- and absolute forecast error for the 10 minute dataset.

Figure 2.3 shows the cumulative count of 10 minute data points outside the power curve regions framed by dashed and solid lines in Figure 2.1. Near-vertical slopes in Figure 2.3 indicate that several temporally-consecutive data points digress from the power curve region, while steady count-increase may well suggest diurnally conditioned digression from the power curve region. The Stor-Rotliden winter season is characterised by finite slopes which may be interpreted as icing events, oscillating diurnally between melting and freezing weather conditions.

As an attempt to infer whether Stor-Rotliden anemometer readings are corrupted during what appears to be icing events, monthly biases are shown in Figure 2.4 based on NWP data from GFS operational forecasts. Figure 2.4 displays severe forecast underprediction for December and January, in particular. However, had the discrepancy been due to nacelle anemometer freezing one would expect *lower* measured wind speed relative to forecasts, i.e. the inverse of the case in Figure 2.4. Hence, the winter bias cannot be explained in terms of iced anemometers as the addition of a layer of frozen water can hardly be imagined to accelerate anemometer rotation. Also, monthly scatter density plots cf. Figure 2.1 for October, November and February (not shown) indicate icing periods (data points below the solid line) as well, though no corresponding inconsistency

Wind farm	Wind speed	Wind power
Stor-Rotliden	5.3	19.1
Rejsby Hede	3.3	1.8
Horns Rev	0.9	1.6

TABLE 2.8: Missing data percentage out of the yearlong, processed wind farm datasets applied as forecast reference in Chapter 3 and appendices A, B.

stands out in Figure 2.4 for these months. Slight cumulative squared- and absolute wind speed error slope-increases are observed, respectively, during the winter months, and may suggest inferior NWP performance during winter where large-scale weather system driven storms are more frequent than during summer.

In summary, all wind power data points above dashed lines and below solid lines are removed from the dataset. In addition, Rejsby Hede and Horns Rev wind speed data falling above the dashed lines in Figure 2.1 is removed.

2.4 A reference for the limited area forecasting studies

Table 2.8 displays missing data percentages for the three wind farms datasets post processing for application as forecast referenc data in the limited area forecasting studies. As noted first in this chapter, forecast reference data applied in Chapter 4 and Appendix C are processed by the criteria established in sections 2.2 and 2.3.

In summary, cf. point 3) in Warner's [110] *modelling practices* checklist, Chapter 2 describes in detail the processing of raw 5 minute nacelle anemometer and power production data from an inland, a coastal and an offshore wind farm into a dataset suitable for comparison to 10 minute resolved forecasts issued twice-daily, each integrated 48 hours beyond initialisation time. As discussed in the following chapter, performance assessment for operational forecasts from a GCM, namely the GFS briefly considered in Section 2.3, and respective forecasts deriving from four nested LAM domains of increasing resolution is carried out. Cf. Section 2.2, a heuristically set minimum of a third of the wind turbines of a given wind farm are required to be online. If not, power data is discarded due to concerns that, below this minimum, wind turbine wake effects will have a strong impact on power readings, depending on the wind direction, and there is no way to account for different offline patterns as information on the status of individual

wind turbines is not available. However, wind direction bias is addressed in the WPPT when translating wind speed forecasts to wind power predictions. A heuristically set maximum of *four* temporally-consecutive time series data points are allowed for wind speeds over 0.1 m/s and power data exceeding 1 % of full wind farm capacity. Such occurences are assumed to be due to malfunction of anemometer and/or wind turbine control software and are therefore discarded from the dataset. Based on scatter density plots of nacelle-measured wind speed vs. power production data, a simplistic piece-wise linear framing of power curve data points is inferred from visual inspection, and wind power data outside the power curve region is removed due to the inconsistence with anemometer readings. In addition to removal of power data digressing from power curves, wind speed data above dashed lines in Figure 2.1 is also discarded for the three sites since erroneous online-capacity flags (giving rise to inflated/deflated power scaling) alone cannot explain the scatter patterns observed above Figure 2.1 dashed lines.

Chapter 3

Limited area forecasting

Development of the Weather Research and Forecasting (WRF) model [59] commenced around the turn of the millenium and is ongoing, with numerous yearly contributions to the open source code from the worldwide WRF community. Two dynamical cores are available, of which the Advanced Research WRF (ARW) system [60, 61] is employed in this work and henceforth, as in Chapter 1, referred to simply as *the WRF model*.

The temporal evolution of atmospheric dynamics is characterised by the three equations for momentum conservation, the continuity equation expressing conservation of mass, and the thermodynamic energy equation theorising energy conservation for a fluid element, cf. e.g. Holton and Hakim [111]. In practice, the continuum of atmospheric dynamics underlying observed weather is represented either by spectral [112] or grid NWP models [113], of which the WRF model adhere to the latter category, as do most *limited area* models for climate and weather. The governing equations of the WRF model [114] include nonhydrostatic (cf. Table 1.1) momentum equations for each spatial dimension, along with four additional prognostic equations for potential temperature, simulation layer thickness in terms of atmospheric pressure, geopotential height, and atmospheric moisture, respectively. The four first-mentioned PDEs involve forcing terms representing parameterised sub-grid scale physics, spherical projections, and Earth's rotation. Furthermore, the diagnostic equation of state along with an equation for the inverse density of dry air are included in the dynamical core formulation. The time integration is based on Runge-Kutta $2nd$ and $3rd$ order schemes, cf. Wicker and Skamarock [115], and $2nd$ through $6th$ order accurate flux divergence discretisations are available for advection

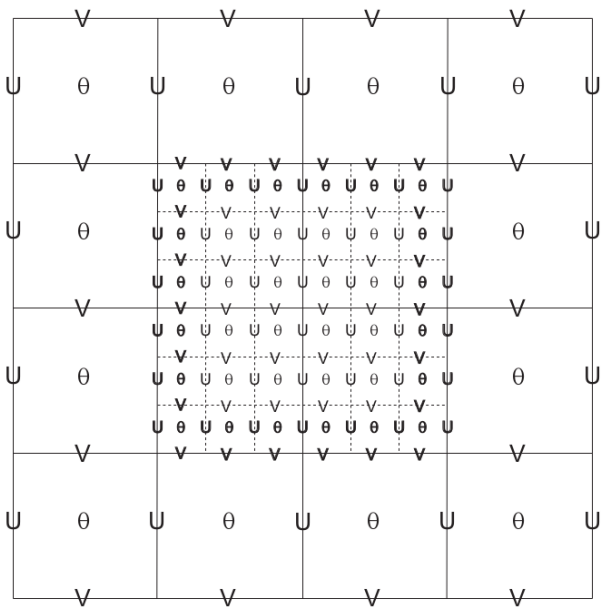


FIGURE 3.1: Sketch of how a finer-resolved Arakawa C-grid [117] computational domain is nested within the parent domain. θ denotes scalar quantities such as pressure and temperature, while U and V represent the horizontal components of atmospheric flow. The figure is borrowed from [119] with permission.

modelling [116]. The horizontal plane on which the spatial scale of advective processes is resolved numerically is spanned by rectangular grid points with Arakawa C-grid staggering [117], and computational domains can be consequetively ingested with increasing *horizontal* (not vertical) resolution for each nesting level, cf. Figure 3.1. The governing equations for atmospheric dynamics are formulated on sigma levels cf. Laprise [118], i.e. horizontal grids are discretised on a terrain-following hydrostatic-pressure vertical coordinate.

In Section 3.1 LAM initialisation strategies are discussed, Section 3.2 concerns atmospheric dynamics description on scales unresolved in NWP models, and Section 3.3 opens the LAM study of this work with a discussion on orography representation in the WRF model. The article drafts in appendices A and B are introduced in Section 3.4, and Chapter 3 is wrapped up with a discussion on error sources in the LAM studies of appendices A and B — and limited area weather and climate modelling studies in general.

3.1 Initialisation of limited area weather model forecasts

There are two alternatives for LAM forecast initialisation; namely *cold* and *warm* ICs, respectively [120, 121].

- **Cold start:** Direct interpolation of the GCM fields to LAM computational grids.
- **Warm start:** At initialisation time LAM variable fields are inferred from the previous LAM variable field forecast, in combination with assimilation of atmospheric soundings, surface observations, satellite data, etc.

The latter accommodates continually developing atmospheric dynamics. E.g. vertical motion and heat fluxes associated with mesoscale convective systems are well-established in the simulation at forecast initialisation, and hence likely providing more accurate spatio-temporal prediction of precipitation, convective storms, and squall lines, that is; elongated thunderstorms that usually form along weather fronts. Cold start LAM simulations need to gradually "spin up" the various atmopsheric dynamics aspects forced by the GCM at initialisation and via LBCs for the duration of the forecast, and hence need to recover a physical solution to the governing equations for atmospheric dynamics on the LAM computational grid(s). The initial-shock numerical artefacts causing spurious fluctuations in meteorological fields simulated decrease with LAM integration during the first few forecast leadtime hours, until a credible representation of atmospheric dynamics is recovered. Though in the midlatitudes, including Scandinavia and the North Sea region, flow above ∼1km altitude can often be assumed in geostrophic balance [122] and largescale pressure systems typically dominate local weather at the surface — for horizontal advection, in particular. As the application at hand concerns atmospheric flow, and spin up of sub-grid scale physics parameterisations of namely convective dynamics is of less consequence than in tropical regions, a few advantages of the cold start strategy are highlighted.

Warm IC weather forecasts drag along to subsequent forecasts errors induced by sparse or erroneous observations during data assimilation, whereas cold start simulations start anew with every forecast initialisation and thus do not accumulate errors cf. the warm IC case. This warm start issue could be circumvented e.g. by initialising LAM forecasts using only recently predicted cold start variable fields and/or nudging the fields with

observations for a shorter period leading up to initialisation time, though this would still allow LBC errors [28, 123, 124] longer time to corrupt meteorological features within the interior LAM domains close to the target sites, cf. e.g. [48, 51, 52]. Atmospheric dynamics simulated in cold start WRF model forecasts are assumed to have largely recovered from adverse numerical artefacts due to the initial shock 12 hours into the forecast [54], although e.g. the study by Hahmann *et al.* [68] indicate that proper spin up may take longer. Based on the discussion above, and the fact that the forecast leadtimes of interest here lie beyond said integration time in the simulation of the governing equations for atmospheric dynamics (cf. Section 3.3), the cold start strategy is employed in this study.

3.2 Parameterisations for sub-grid scale physics

Holton and Hakim [125] narrow in the most important classes of sub-grid scale physical processes as pertaining to:

1. Radiation.
2. Clouds and precipitation.
3. Turbulent mixing and exchange.

In a wind energy scheduling context, namely the third class arguably directly impacts simulation of near-surface atmospheric flow. However, the modelling accuracy of incoming solar radiation and resulting exchange of heat from the ground – controlled by the land surface model applied – feeds into 3. in terms of thermal boyuancy driving vertical turbulent flow thus nudging simulated weather towards unstable conditions, and hence e.g. affecting the mean flow with tangible implications for wind power generation. 2. in turn feeds into 1. in that correct representation of the cloud fraction in part influences incoming solar radiation, and hence ultimately also atmospheric stability cf. the discussion before. A sketch of the processes parameterised in NWP models is shown in Figure 3.2. The key message is that testing individual parameterisations for sub-grid processes in an NWP formulation is a challenging task, as all aspects interact via the coupled nonlinear PDEs constituting the weather model's dynamical core.

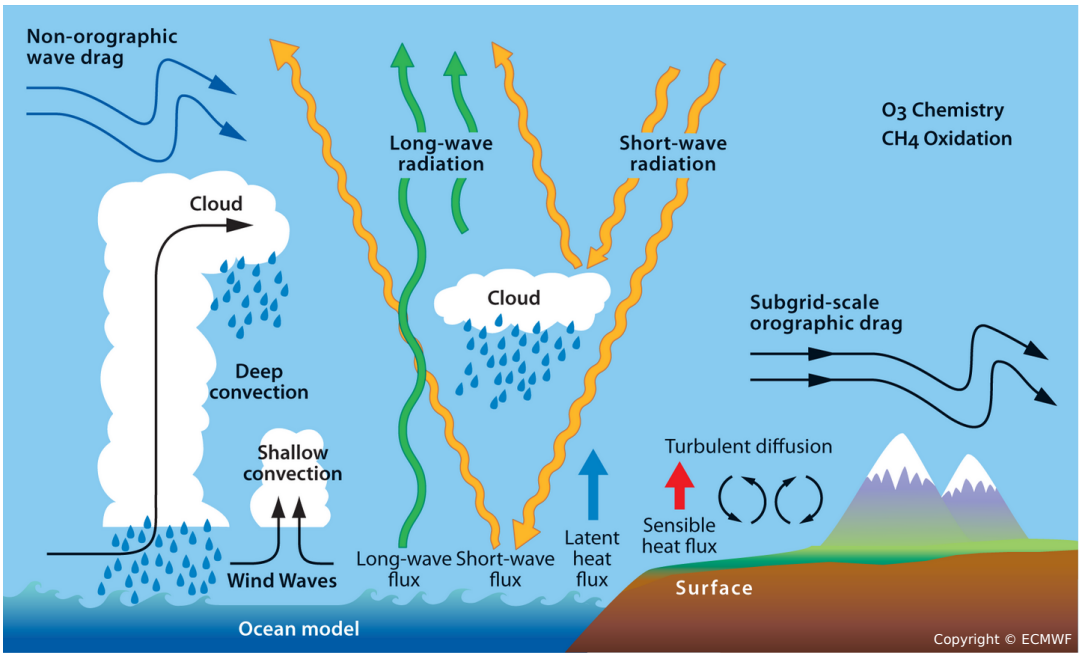


FIGURE 3.2: Sketch of atmospheric dynamics that cannot be resolved explicitly and therefore needs to be parameterised. Reproduced from [126] with permission.

A thorough review of the technical aspects and details underlying each class of NWP model parameterisation for unresolved processes lies beyond the scope of this work, and the interested reader is instead referred to e.g. Warner [127], and [128] for a description of land surface modelling aspects, or Pielke [129] for more in-depth details on the topic.

However, a few characteristics of *planetary boundary layer* (PBL) schemes is discussed in some detail, as the relative merits of alternative PBL parameterisations are arguably central to the feasibility of NWP model forecasts for wind energy scheduling, cf. e.g. Lee *et al.* [130] who find strong influence of the particular PBL representation approach on WRF modelled 10 m wind speed forecasts. PBL schemes represent turbulent vertical fluxes of momentum, heat, moisture, etc. and a closure scheme must be selected in order to evaluate turbulent fluxes based on mean values of atmospheric quantities [131] available from prognostic and diagnostic NWP variables. See e.g. Alapaty *et al.* [132] for a concise overview and application of different closures, and e.g. Stull [133] for an in-depth discussion on the topic. Turbulent kinetic energy (TKE) closure schemes introduce a TKE prognostic equation to the NWP model formulation in order to describe more of the physics of the atmospheric planetary boundary layer than simpler first-order closure schemes, which model atmospheric turbulent transfer alike molecular transport [131, 134]. Such *local closure* schemes rely on NWP variables and associated gradients

at the local computational grid point only. In contrast, *nonlocal* PBL schemes accommodate interaction – though only vertically, due to computational resource constraints – with NWP model variables at vertical layer grid points beyond the vicinity of the grid point considered. Many approaches have been applied in the literature, either based on explicit vertical coupling, cf. e.g. [135], or parameterised interaction, cf. e.g. [136]. The asymmetric convective model, version 2, (ACM2) is available in the WRF model, and offers hybrid local/nonlocal PBL closure, including nonlocal transport only during unstable weather conditions when convection dominates local atmospheric dynamics, cf. Pleim [137, 138].

A number of comparative PBL parameterisation studies for WRF model options have been carried out, see e.g. Hu *et al.* [139] getting smaller midlatitude (Texas, USA) biases for the nonlocal Yonsei University (YSU) [140] and hybrid ACM2 schemes relative to the local Mellor-Yamada-Janjic (MYJ) TKE-approach [75] for surface observations. Extending said PBL parameterisation ensemble by the Boulac scheme [141], Xie *et al.* [142] find superior accuracy of the ACM2 for 10 m wind speeds in a subtropical region (Hong Kong, China). Further north, the WRF modelled wind speed results in the study by Draxl *et al.* [143] hint local closure (MYJ) performance superiority subject to stable weather conditions, while the ACM2 hybrid performs best under neutral conditions, and the nonlocal YSU scheme outperforms the others for unstable conditions (northern Europe). Finally, the six-PBL scheme study by Deppe *et al.* [38] find (Iowa, USA) that the Mellor-Yamada-Nakanishi-Niino (MYNN) level 2.5 local closure scheme [144, 145] better captures rapid changes in the atmospheric flow — so-called *ramp events*. Accurate forecasts of wind speed ramps are important in a wind energy scheduling context [3], and hence this topic is considered in Appendix B. Due to the results of Deppe *et al.* [38], as well as the typically largescale pressure system dominated weather conditions cf. Section 3.1 in the regions considered, local closure schemes may be slightly more feasible for the application at hand – cf. Draxl *et al.* [143] – on average during a yearlong study period for three Scandinavian locations assumed mainly characterised by stable to near stable weather conditions. Hence, the TKE-based MYNN 2.5 PBL parameterisation is applied in WRF model simulations carried out in this work. See Section 3.4 and Appendix A for an overview of the WRF model configuration.

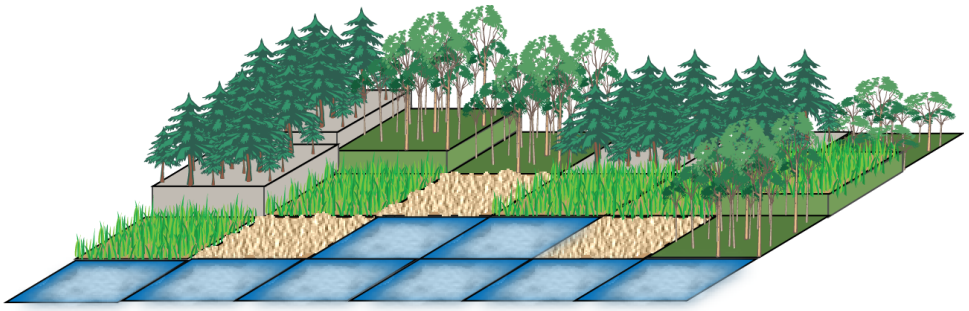


FIGURE 3.3: Sketch of terrain tiles representing the lower boundary of atmospheric dynamics simulations. The figure is borrowed from [146] with permission.

3.3 Terrain representation on high-resolution model grids

As e.g. McQueen *et al.* [84] find in their study based on the RAMS, also mentioned in Section 1.3, high-resolution NWP modelling may be required to accurately simulate atmospheric flow. Coarsely-resolved GCM data is not able to represent complex terrain in high detail, and hence added wind energy scheduling value from LAM forecasting is expected for the inland Stor-Rotliden wind farm, in particular. The computational domains used to generate the WRF model forecasts are sketched in *Figure 1 of Appendix A*. The terrain tiles of horizontal grids are vaguely visible and each of these are assigned a specific land use category with associated topographic attributes. For example, Figure 3.3 illustrates discretised representation of orography, needle- and leaf forest, grassland, beach, and ocean/lake surfaces as "viewed" by the discretised PDEs of the WRF model. As the McQueen *et al.* [84] study indicate for RAMS simulations, e.g. Santos-Alamillos *et al.* [147] find that terrain representation has a strong impact on WRF modelled wind speed. Hence, a preliminary study on the sensitivity to terrain representation in the WRF model for the aforementioned complex terrain site in central Sweden is carried out prior to the forecast performance studies for wind farms in different terrain types, introduced in Section 3.4.

3.3.1 Forecast performance assessment strategy

On the Elspot day ahead market for energy trading in northern Europe, the auction closes at noon CET [148] which in turn implies that only forecast leadtimes beyond

12 hours ahead of this deadline are relevant for wind energy scheduling. When operational GCM forecasts are initialised, weather observations collected globally each day are assimilated into GCM-predicted spatially gridded fields for meteorologial variables, which are derived from the previous forecast based on a time-window traversing initialisation time for the current forecast. Hence, GCM computations commence with a few hours delay from initialisation time and take an additional hour or so to compute, depending on forecast length and model resolution. The WRF model forecast runtime further adds to the delay from initialisation time, depending on the number of grid points and integration time step specified for each computational domain, and hence a total delay of 6–12 hours from initialisation time to forecast completion is realistic in an operational setting for wind energy scheduling. The GCM applied for WRF model initialisation and lateral forcing is the GFS [63], generating daily global weather forecasts initialised 0000, 0600, 1200, and 1800 UTC, respectively. Hence, daily GFS midnight UTC initialisations, or 0600 UTC depending on WRF model runtime, and forecast leadtime hour span 12–48 are relevant for the WRF model application at hand.

An aspect of interest in this work is to gauge forecast accuracy as a function of forecast leadtime, which can be illustrated in a number of ways. The dataset could be grouped according to each forecast hour such that a performance metric is evaluated for wind speed forecasts one hour ahead against corresponding observations, with similar evaluations for the other forecast leadtime hours. This would provide performance leadtime dependency information of high temporal resolution. However, since the NWP model studied in this work do not provide forecasts initialised each hour operationally, different leadtime hours are tied to specific hours of the day. This intrinsic diurnal influence thwarts inference on the extent to which performance depends on forecast leadtime. On the other hand, if a single performance metric evaluation is carried out for the full dataset wind speed measurements would be recurrent which would constitute an indirect violation of the assumption of independent and identically distributed (i.i.d.) statistical model residuals underlying the statistical model development process outlined in *Appendix C Section 3.4*, and recurring data points arguably decrease the *effective* number of independent forecast reference measurements even further than the effect from chronological data point dependency, cf. Section 4.4. Therefore, the dataset is binned according to forecast leadtime spans of width equal to the time period between NWP model forecast initialisations. Data within each forecast leadtime hour group is then

augmented chronologically such that all wind speed measurement data points occur exactly once in each dataset. In this way all hours of the day are represented in each forecast performance metric evaluation and inference on forecast performance leadtime dependency is thus readily assessed from forecast leadtime group results.

In order to resolve forecast leadtime dependency, both 0000 and 1200 UTC GFS initialisations are included in the study, and hence three forecast leadtime horizon groups are available to resolve forecast performance dependency on the leadtime; namely the three forecast leadtime groups 12 $\leftrightarrow$ 24, 24 $\leftrightarrow$ 36 and 36 $\leftrightarrow$ 48 hours ahead of forecast initialisation. N.b. that this is only done for wind *speed*, while wind *power* forecast performance assessment is condensed to a single forecast leadtime group to summarise results, cf. appendices A and C.

The magnitude of the Pearson correlation coefficient (PCC) is invariant under linear transformations, and therefore not sensitive to NWP model bias when used for forecast performance assessment. Shevlyakov and Smirnov [149] summarise a range of alternative estimators of the correlation coefficient robust to outliers, that is; focusing evaluation on "good" data, and thus potentially capturing a true relation otherwise masked in noisy data. However, in this study a neutral assessment evaluating data subject to severe forecast errors on par with better forecasts is preferred to sustain realistic conditions in the study. Fosdick and Raftery [150] note that the PCC is a Maximum Likelihood estimator provided the constituent variable datasets are drawn from a bivariate normal distribution, which can typically be assumed for continuous variable time series data. Hence, subject to this assumption, the PCC is a minimum variance statistically unbiased estimator for wind speed forecast performance, and this scalar accuracy metric is applied for forecast performance assessment in Section 3.3.2, as in all three appendices.

3.3.2 Forecast performance as a function of spatial smoothing

The highest resolved computational grids are assigned 3.3 km and 1.1 km horizontal resolution, respectively, and all WRF model domains have one grid point specified at the official Stor-Rotliden wind farm coordinate at 64.2166° latitude, 18.3829° longitude, cf. Table 2.1. The 1.1 km resolution is so fine that several grid points cover the horizontal extent of the wind farm, and hence a proper weighted grid-point wind speed mean may provide a better overall forecast in comparison to farm-averaged wind turbine nacelle

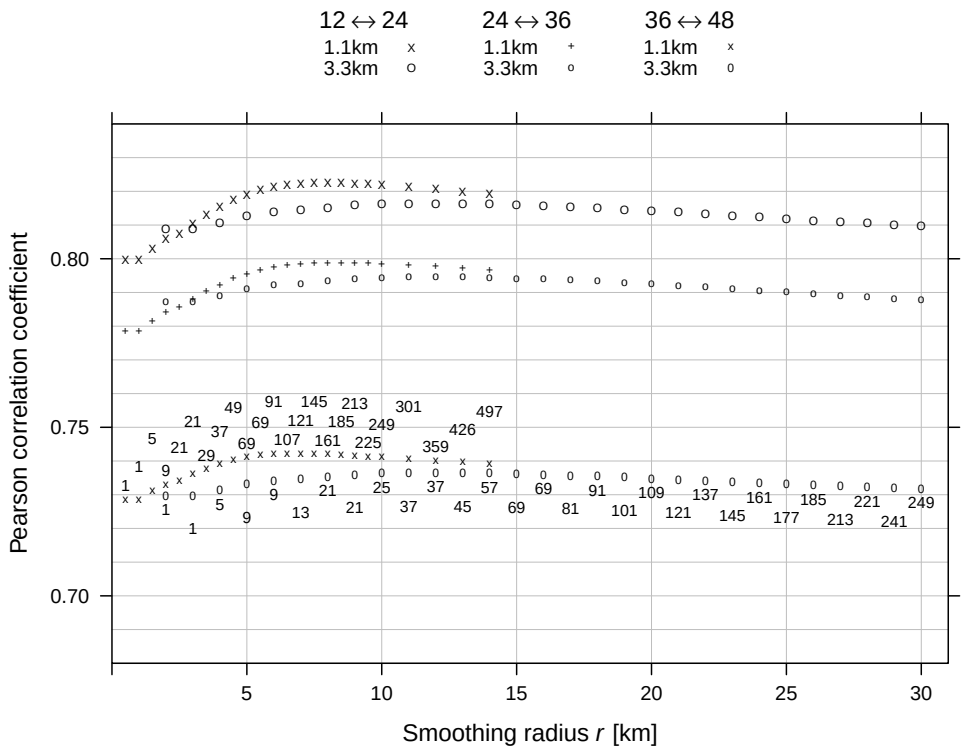


FIGURE 3.4: Wind speed forecast performance as a function of spatial smoothing by Eq. (3.1) for the two high-resolution WRF model grids considered in the studies introduced in Section 3.4, based on the *official* wind farm coordinate [151]. Forecast leadtime dependency is resolved according to the strategy laid out in Section 3.3.1. The number of grid points within r is noted above (below) 1.1 km (3.3 km) PCC values for the 36 ↔ 48 forecast leadtime hour group.

anemometer readings. Therefore, the yearlong forecast dataset is assessed as a function of the degree of spatial smoothing by the tricube weight function,

$$w_t(\boldsymbol{x}) = \begin{cases} \dfrac{\left(r - \|\boldsymbol{x} - \boldsymbol{y}\|\right)^3}{\sum\limits_{\|\boldsymbol{x}-\boldsymbol{y}\|<r} \left(r - \|\boldsymbol{x} - \boldsymbol{y}\|\right)^3} & \text{for } \|\boldsymbol{x} - \boldsymbol{y}\| < r \\ \\ 0 & \text{for } \|\boldsymbol{x} - \boldsymbol{y}\| \geq r \end{cases} \tag{3.1}$$

The tricube weight $w_t(\boldsymbol{x})$ is only nonzero when the smoothing radius r is greater than the Euclidean distance $\|\boldsymbol{x} - \boldsymbol{y}\|$ between coordinates $\boldsymbol{x}$ for the grid point to be assigned a weight and the wind farm reference coordinates $\boldsymbol{y}$. The results are based on hourly averages for the 10 minute dataset and shown in Figure 3.4. Note the inferior forecast performance for the 1.1 km domain relative to the coarser 3.3 km grid for $r < 3$ km, transcending to superior 1.1 km wind speed forecast performance for $r > 3$ km. This

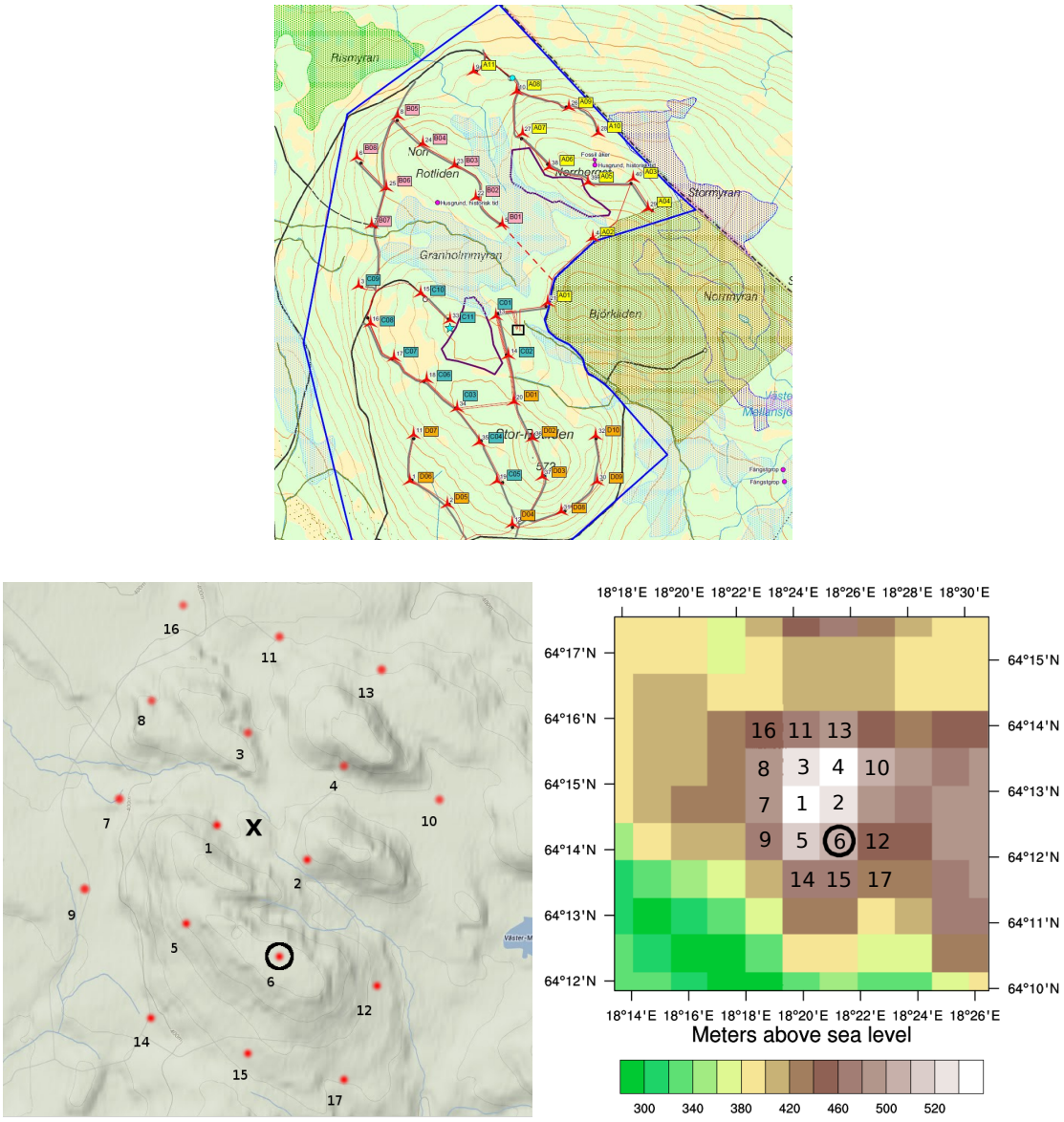


FIGURE 3.5: Representation of the Stor-Rotliden site in the WRF model. TOP: Layout of the wind farm turbines, reproduced here with permission [152]. LOWER LEFT: Orography map [153] with near-turbine 1.1 km domain grid points marked in red, numbers denote the ordinally ranked proximity to the wind turbine average coordinate **X**. The official wind farm coordinate [151] for which the four WRF model computational domains of the study in Section 3.4 have a grid point specified is marked with **O**. LOWER RIGHT: Representation of terrain elevation above mean sea level in the WRF model; generated using NCL [154].

finding prompts a closer inspection of the 1.1 km grid points representing the weather in the vicinity of the Stor-Rotliden wind turbines.

Coordinates for the 40 Stor-Rotliden wind turbines were obtained after carrying out the WRF model runs. See Figure 3.5 for a sketch of the locations, 1.1 km domain grid points

	Height above mean sea level [m]	
Ranked proximity of 1.1km grid points to the average of the turbine coordinates; **X**	Observed	Modelled
1	495	537
6	569	505

TABLE 3.1: Observed [103] vs. WRF modelled terrain height above mean sea level for 1.1 km grid points closest and 6*th* closest to the average of the 40 coordinates of the wind turbines constituting the Stor-Rotliden wind farm.

near wind turbines, and orography as modelled in the WRF model, cf. the illustration in Figure 3.3. Two issues are apparent:

1. The mean wind turbine coordinates are ∼1 km from the official coordinates for the wind farm [151], i.e. the 30 km, 10 km, 3.3 km, and 1.1 km computational grids of the studies introduced in Section 3.4 do not each have a grid point at the horizontal centre of the wind farm. That is, the WRF model grid points at **O** represent atmospheric dynamics of the *south-eastern* corner of the wind farm.

2. Orography in the vicinity of the wind farm is modelled erroneously. As Table 3.1 demonstrates, the terrain-tile heights above mean sea level at **X** and **O** appear to be swapped relative to the observed terrain. The hill at the southern end of the wind farm is named *Stor-Rotliden* – the location that the official wind farm coordinates [151] in fact denotes – is erroneously modelled as being *next to* a hill.

Wind speed forecast performance for each 1.1 km grid point sketched in Figure 3.5 is shown in Figure 3.6. The three forecast leadtime group PCCs for the grid point at **O**, the 6*th* closest to **X**, are ∼0.02–0.03 lower than for the grid point closest to **X**. Overall, there is a steady drop in performance as a function of distance from **X**, though with a few performance-humps for the 7*th*, 8*th*, 10*th*, 11*th* and 16*th* closest grid points to **X**. Note from Figure 3.5 that these are all outermost grid points outlining the northern part of the wind farm area, furthest away from the wrongly represented Stor-Rotliden hill orography.

Upon changing the smoothing reference $\boldsymbol{y}$, cf. Eq. (3.1), from **O** to **X** cf. Figure 3.5, the wind speed forecast performance results shown in Figure 3.7 are obtained. Note that the highest resolved WRF model domain now consistently outperforms the coarser parent domain.

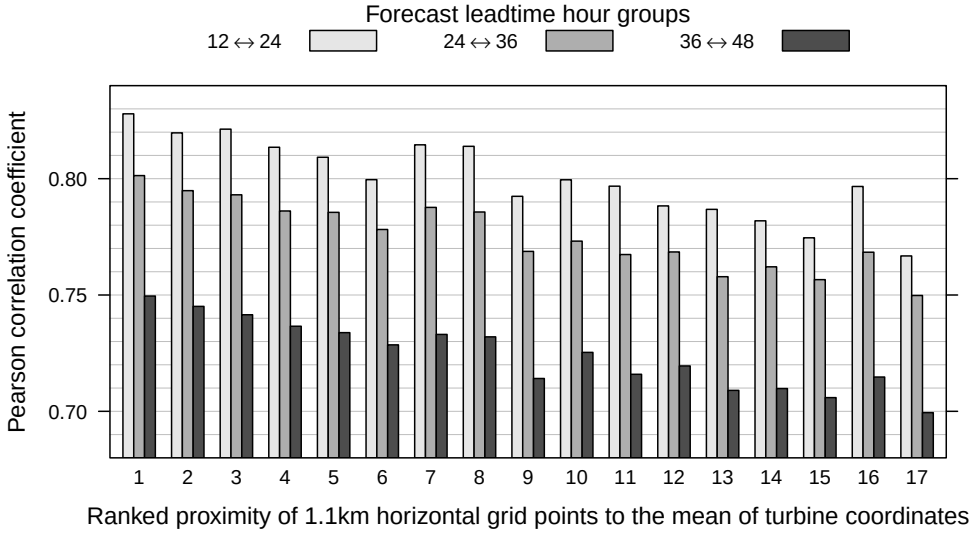


FIGURE 3.6: Wind speed forecast performance for the WRF modelled 1.1 km grid points sketched in Figure 3.5.

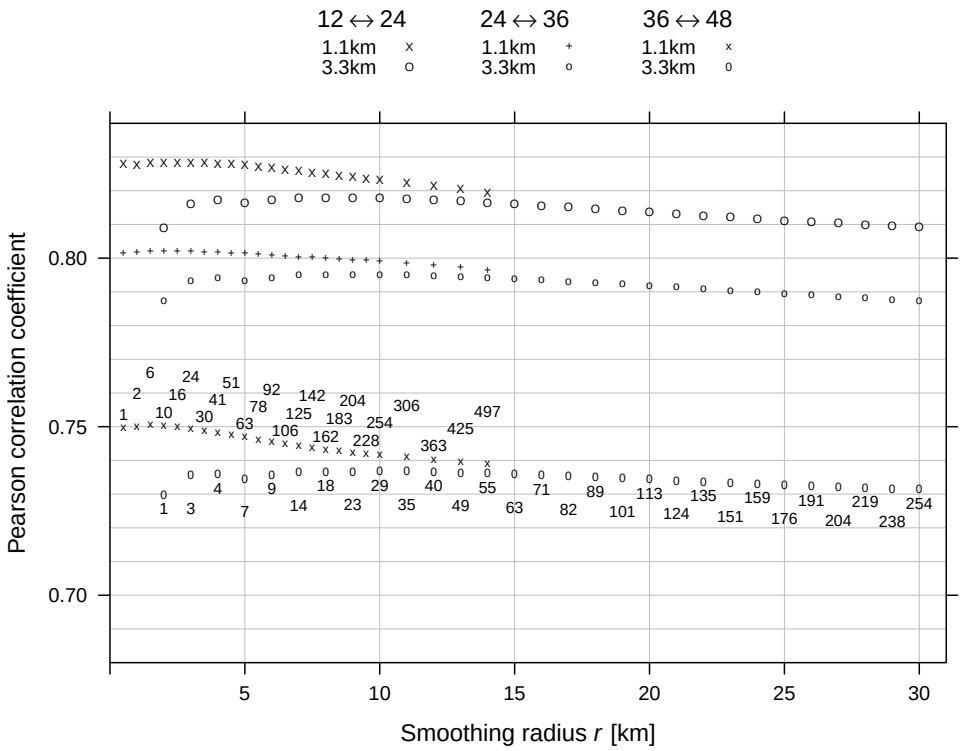


FIGURE 3.7: Cf. Figure 3.4, though the smoothing reference wind farm coordinate is replaced by the average of the 40 individual wind turbine coordinates constituting the Stor-Rotliden wind farm.

3.3.3 Remarks on the findings

The WRF model study by Jiménez *et al.* [65] partially attribute wind field simulation errors to terrain misrepresentation, as is also demonstrated here and in the already mentioned study by Santos-Alamillos *et al.* [147]. The results of Section 3.3.2 emphasise the importance of careful inspection and verification of prerequisite information, as small spatial deviations in reference coordinates are shown to cause high-resolution forecasts to perform inferior to the coarser laterally forcing data. Retrospectively, the WRF modelled inner-most computational grid should have been based on higher-resolved terrain dataset than the applied Moderate-resolution Imaging Spectroradiometer (MODIS) [155] satellite data for landuse and United States Geological Survey Global 30 Arc-Second Elevation (USGS GTOPO30) [156] orography dataset, both resolved at the same spatial scale as the WRF model 1.1 km domain.

Upon correcting the smoothing reference consistent improvement with WRF model resolution is found. From the study of the 17 individual 1.1 km grid points overlapping wind turbine coordinates, the three points closest to the average of the 40 turbine coordinates – arguably the *correct* wind farm centre, in contrast to that proposed in [151] – are the three best performing grid points. This supports the notion that the WRF model does indeed reproduce observed atmospheric dynamics consistently in space and time. The PCC drops by 0.05–0.06 (6–7 %) between the best and worst performing grid points, which are closest and furthest from the wind farm centre, respectively. The performance of the grid point at the Stor-Rotliden hill, the $6th$ closest to the wind farm centre, has 0.02–0.03 smaller PCC than the best performing grid point closest to the wind farm centre, corresponding to a decrease in performance of 3–4 %.

The peak performance radius for the 1.1 km computational grid is found to decrease slightly as a function of the forecast leadtime group. This can be explained by increased phase errors which cause computational grid points in the vicinity of the wind farm to be less likely to capture in space the timing of wind speed fluctuations as observed at wind turbine hub-height. Thus, including these grid points penalises the PCC accuracy metric. Computational grid points close to the wind farm centre has greater phase error tolerance, in that, no matter in which meteorological direction the wind speed fluctuation timing is spatially off between grid point simulation and reality, there is a higher probability that the wind farm data – which, cf. Chapter 2 and Figure 3.5, is

based on measurements from 40 turbines covering several 1.1 km grid points – captures behaviour seen at the close-to-central grid point, than for a WRF modelled 1.1 km grid point in the vicinity of the wind farm (which is more likely to see information not captured by simultaneous measurements, given finite phase errors).

Although there is a slight advantage from including more than just the 1.1 km grid point closest to the wind farm centre, this never exceeds ∼0.002 improvement in PCC, which corresponds to ∼0.3 %. This, combined with the fact that the optimal smoothing radius is ∼1.5 km (6 computational grid points with nonzero tricube weights) for the 36 ↔ 48 forecast leadtime hour group, leads to the definition of an optimal smoothing radius of no more than 1.5 km for the 1.1 km grid.

For the 3.3 km computational grid there is a ∼0.007 improvement independent of forecast leadtime group when going from the grid point closest to the wind farm centre ($r = 2$ km) to including the three closest grid points ($r = 3$ km). As the sketch of high-resolution grid points shows, the 3.3 km grid point closest to the wind farm centre is on the top of the Stor-Rotliden hill, the location of which is marked by the publicly available reference coordinates for the Stor-Rotliden *wind farm*. As the sketch shows, this is *more than 1km off* the average of individual wind turbine coordinates, and moreover is at the peak of a hill which is not representative for all 40 wind turbines, since most are situated on hill slopes and moderate depressions. As the potential benefit of moving beyond 4 km smoothing radius for the 3.3 km grid is vanishingly small (∼0.1–0.2 % better for ∼10 km), the optimal smoothing radius is defined to be 4 km, implying nonzero Eq. (3.1) weights for the four grid points closest to the wind farm centre.

Perhaps not surprisingly, no performance gain can be documented for spatial smoothing of the 10 km and 30 km WRF model grid data also considered in the studies introduced in Section 3.4, and these grids are hence not considered in Section 3.3.2.

3.4 Introduction to appendices A and B

A key objective of this work is to assess the value of limited area forecasting for wind energy scheduling, which in a northern European context implies energy trading on the Elspot market [148] based on NWP modelled wind speed for leadtimes spanning the 12th

	Domain 1	Domain 2	Domain 3	Domain 4
Horizontal distance between computational grid points [km]	30	10	$\sim$3.3 $\left(\frac{10}{3}\right)$	$\sim$1.1 $\left(\frac{10}{9}\right)$
Unstaggered vertical levels ($\sim$ four below 100 m AGL)	40	40	40	40
Integration time step [s]	120	40	$\sim$13.3 $\left(\frac{40}{3}\right)$	$\sim$4.4 $\left(\frac{40}{9}\right)$
Output frequency [minutes]	10	10	10	10

TABLE 3.2: Resolution and output frequency employed in WRF model simulations.

to 48*th* hour beyond forecast initialisation. The WRF model configuration employed is shown in Table 3.2, cf. Appendix A for further details.

Although important for accurate simulation of sub-grid scale processes involving e.g. turbulent momentum and thermal fluxes [157], a recent study [158] indicates that the vertical resolution has little effect on the simulation accuracy for the mean flow across a NWP model grid cell (though recall also McQueen *et al.* [84], arriving at the opposite conclusion in their RAMS study). The relatively sparse 40 vertical layers (41 staggered levels) – cf. e.g. the 64 layers in the GFS, albeit covering a greater vertical extent – for each of the computational grids outlined in Table 3.2 are assessed adequate and also serves to conserve computation walltime for the twice-daily 48 hour WRF model forecast generation covering a year.

Integration time steps are kept sufficiently small to avoid violation of the Courant-Friedrichs-Lewy (CFL) condition [20, 21]; roughly put atmospheric flow must not traverse more than one horizontal grid point in a single time step in order for the numerical integration process to stay stable. A common practice for WRF modelling is to set the time step (in seconds) equal to six times the horizontal resolution (in kilometers) [146], and hence this study is on the conservative side to substantiate numerical stability and avoid simulation crashes due to violation of the CFL stability criterion.

In order to reduce generation of spurious inertia-gravity waves and other numerical artefacts, cf. Warner *et al.* [52], the horizontal resolution of the outermost computational grid (30 km) is kept close to that underlying NWP by the GFS ($\sim$ 27 km; though $\sim$ 13 km as of mid-January 2015 [159]) that forces WRF model runs laterally and for initialisation, although GFS output is only available at coarser 0.5° ($\sim$55 km) horizontal grid points.

For high-resolution computational grids, a number of studies report realistic nonhydrostatic model simulation of a range of mesoscale convective systems, see e.g. [160–163]. Weisman *et al.* [163] mention horizontal resolutions from 4 km and finer, while Bernadet *et al.* [164] find that 2 km resolution is necessary to capture convection explicitly in their RAMS runs. Finally, Sato *et al.* [81] find that simulations with the cloud-resolving Nonhydrostatic Icosahedral Atmospheric Model (NICAM) succesfully reproduce precipitation observations at 3.5 km horizontal resolution, while 7 km is off in both magnitude and timing. Based on this cumulus parameterisation literature review, the Kain-Fritsch scheme [165] employed here, cf. *Appendix A Table I*, is switched off for the 3.3 km and 1.1 km WRF model nests in this work. The microphysics, describing the physical processes underlying precipitation from water and ice particle growth within a cloud, are parameterised according to the Thompson parameterisation [166], the choice of which is substantiated e.g. by the study of Rögnvaldsson *et al.* [167], and the study of a 2008 thunder storm in India by Rajeevan *et al.* [168] find better updraft/downdraft description from this scheme compared to three alternatives, albeit similar simulation of surface wind variation.

3.5 Multidimensional nearest neighbour smoothing

The conditional parametric modelling study presented in Appendix A could be extended with coefficient multipredictor-dependence; $\theta(z_1, z_2, \ldots, z_n)$. However, for z candidates uniformly distributed within a p-dimensional hypercube, a fraction r of the unit volume (corresponding to the fraction of data points covered to estimate the coefficient) translates to an expected hypercube side-length, that is; individual z predictor (i.e. individual-z-dimension) range of $r^{1/p}$, cf. [169]. Hence, for modest smoothing degrees r – corresponding to very local smoothing – a large fraction of each z candidate range is necessarily included and the procedure becomes *less local* for each z predictor involved in the nonparametric $\theta(z_1, z_2, \ldots, z_n)$ estimate. Consider e.g. inclusion of three z candidates, implying that a smoothing degree of 10 % of the available data; $r = 0.1$, corresponds to inclusion of 46 % of each of the z candidate data points! This issue is known as *the curse of dimensionality* [170], and imposes a practical limit on the feasible of conditional parametric modelling by the method of Nielsen [171] of no more than two

simultaneously included z predictors [172]. Hence only single-z conditional parametric modelling is explored in Appendix A.

3.6 Supplementary discussion on wind speed ramps

The WRF model output is at 10 minute resolution, the underlying integration time steps ranging from 2 minutes to ∼4 s, cf. Table 3.2. As the GFS data is available at 18 times coarser (3 hourly) temporal resolution and resolves topography in the vicinity of the wind farm sites at a ∼27 km horizontal grid constant, much coarser than can be accommodated by the WRF 1.1 km computational domains, GFS wind speed data is arguably less sensitive to local topography and hence likely closer to mean wind speed than corresponding high-resolution WRF model data on average. To illustrate why this may lead to better GFS forecast performance relative to corresponding WRF model performance in terms of scalar accuracy metric assessement, cf. Appendix A, substitute a sinus curve as the forecast reference time series. Subject to a π phase error an otherwise perfectly matching high-resolution sinus curve would nowhere in the dataset be closer to the forecast reference than the constant zero-mean that coarse forecast data is taken to represent in this thought experiment. Hence, high-resolution forecasts of inferior scalar accuracy metric performance due to more or less severe phase errors may in fact yet contain superior information in terms of the temporal evolution of meteorological structures of interest. The results presented in Appendix B support this notion. In the following, adverse implications of wind speed ramps for wind power integration in the electrical grid is discussed, thus supplementing the literature review in Appendix B.

An up-ramp event overestimated in magnitude or forecasted too soon may induce a down-regulation penalty for the wind power producer if the electrical grid is already balanced, depending on the spot (day ahead) and regulation (realtime adjustment) prices [173]. Severe oversupply of wind power can e.g. occur when storms sweep across northern Europe, when consumption is low, and neither German, Danish nor Swedish wind farm production can be exported to neighbouring countries. Although rare, such situations may potentially inflict severe damage on power cable bottlenecks lest wind farms are shut down immediately. On the other hand, a misforecasted down-ramp may induce an up-regulation penalty when power sold on the Elspot market cannot be delivered fully [174], unless the absence of the promised amount of power helps the market. This

situation is potentially very expensive for the energy producer and challenging to handle for the transmission system operator, scaling with the amount of wind power in question. Curiously, the WRF model study by Deppe *et al.* [38] find that less than a fifth of observed ramp events occur during frontal passages or severe storms in the ramp-analysis part of their study spanning one year, focussing on a region in Iowa, USA.

3.7 Error sources pertaining to limited area forecasting

Deferring discussion of general weather model deficiency to Chapter 4, this section concluding Chapter 3 concerns forecast error sources within the context of LAMs, although Section 3.7.1 is indeed relevant also for GCMs.

3.7.1 Starting out right — forecast initialisation

A realistic outset is vital for a computer simulation to attempt to reproduce reality. Hence, two obvious weather modelling error sources concern measurement uncertainty and shortcomings of data assimilation schemes employed to transfer information from atmospheric measurements into weather models, let aside adverse effects from simplifying assumptions necessary for practical feasibility of NWP model formulations and numerical artefacts pertaining to different discretisation schemes. The World Meteorological Organisation facilitates access to atmospheric observations worldwide several times daily, and ongoing improvement of measurement instruments and in quality control algorithms run at weather centres ensure continual reduction in NWP corruption by poor-quality observations. Hólm [175] provide an accessible, yet concise overview of NWP data assimilation techniques, see also Bouttier and Courtier [176] for a more comprehensive review. Kalnay [177] offers an overview of the field of NWP, culminating with a thorough review of data assimilation theory. A few of the techniques in widespread use today are briefly outlined in the following paragraph.

Using the previous forecast to initialise the current forecast is one of the two main types of four dimensional data assimilation (FDDA). The other is known as Newtonian relaxation and involves "nudging" NWP data towards observations by adding forcing functions to the NWP model equations, see e.g. the pioneering works of [178, 179], and [180] for an overview of the technique development over the first few decades. In

practice though, three data assimilation schemes which, in contrast to FDDA, applies NWP model variable covariances for estimating a suitable balance between observed and modelled fields are in use for operational forecasting. Two of these are variational techniques involving minimisation of a cost function that estimates the difference between forecast and observations, though differing in the way the prognostic variable (background) error covariance matrix is estimated. This matrix controls the effect of observations on the NWP model variables and is a key dependency of the cost function. In the simplest of the two variational schemes; three-dimensional variational data assimilation (3DVAR), the covariance is constant and based on historical data, while the more advanced four-dimensional variational data assimilation (4DVAR) technique includes realtime information within a given time window. The ensemble Kalman filter (EnKF) approach also utilises realtime information, though the background error covariance is here inferred from an ensemble of forecasts, each forecast member with a slightly different initial perturbation, cf. e.g. Evensen [181]. While it is well-established that the 3DVAR approach generally performs inferior to the 4DVAR and EnKF approaches – namely with surface observations and in complex terrain [182] – it is less clear-cut which of the latter two performs better [183]. However, Zhang *et al.* [184] find the EnKF approach to be superior in their forecast experiment using the WRF model in combination with the Data Assimilation Research Testbed (DART) [185–187], particularly for forecast leadtime horizons of 2–3 days ahead. As upper air measurements are typically already assimilated in GCM forecasts providing ICs and LBCs for LAM forecasts, any potential for improvement of the initialisation of LAM simulations should derive from surface observations not available, or not ingestible, for GCM initialisation. Ha and Snyder [188] recently demonstrated improved precipitation forecasts up to the $12th$ forecast leadtime hour from EnKF data assimilation of temperature and moisture measurements 2 m AGL and 10 m AGL winds at initialisation time, based on the WRF-DART system also used in the Zhang *et al.* [184] study mentioned above, and attribute forecast improvements to better representation of the depth and structure of the PBL and the reduction of surface bias errors.

The presented literature review concerns application of data assimilation for NWP model initialisation. This section concludes with an alternative application of data assimilation, also with implications to forecast accuracy, namely to infer the nature of NWP model errors. Hacker and Angevine [189] employ the EnKF approach for assimilation

of temperature, water mixing ratio and wind components in order to characterise NWP model error. Results of their single-column WRF-DART experiment could indicate that observed errors arise from unmet assumptions underlying the Monin-Obukhov similarity theory applied in the surface layer parameterisation to provide lower boundary condition turbulent fluxes for the PBL parameterisation.

3.7.2 Spurious perturbations — lateral boundary condition schemes

Deficiencies of LBC formulations are known to cause deteriorating LAM forecast performance as integration time increases [52], cf. Section 3.1. The related issue of lower boundary conditions handled via land surface modelling can also be viewed as a sub-aspect of physical process parameterisation cf. Section 3.2 — a discussion of this complex topic is not considered in detail in this work, although derived adverse effects are indirectly studied in appendices A and C via statistical modelling of the wind speed based on NWP model variables. Upper LAM (and GCM) boundary conditions are arguably less important than LBCs for 2 day LAM simulated atmospheric dynamics near the Earth's surface, as is relevant for the weather modelling application at hand. See e.g. [190–192] for theoretical discussions on upper atmosphere boundary conditions.

Sundström [193] derives two different LBC formulations for which unique solutions to the barotropic vorticity equation exist within a limited area domain, though Oliger and Sundström [194] show that the more complex initial boundary value problems for the primitive equations used in modern NWP are ill posed for open boundaries fixed in space, as is the case for all LAMs. However, the work of Temam and Tribbia [195] carried out a few decades later demonstrates that the same problems are well posed if a mild vertical viscosity is added. Termonia [196] illustrates how limited area modelling of a rapidly moving meteorological feature crossing a LAM domain is dominated by errors coming from the inflow boundary after a short time, with consequently poor performance relative to that of the coarser GCM forcing data. A follow-up study from 2009 [197] suggests that forecasts should be restarted as soon as the severe event has entered the domain, and it is further argued that a scale-selective digital filter be employed for the re-initialisation. Ren and Shepherd [124] provide a theoretical treatment of the influence of LBCs on slow advective timescale atmospheric dynamics and Charney [28] demonstrates the adverse effects of overspecification of the LBCs for the shallow-water equations; a simplified

form of the primitive equations that describe atmospheric flow. Davies [123] extend this work with a theoretical outline of the dependence of the uniqueness of solutions to the barotropic and baroclinic primitive equations on the LBCs. Paegle *et al.* [51] argue that the main source of LAM error growth with model integration time is due to small uncertainties on larger spatial scales rather than small uncertainties on the smaller scales, and furthermore associate the adverse largescale impact on LAM predictions to LBCs rather than ICs.

Virtually all LAMs used in operational NWP use the relaxation LBCs proposed by Davies in 1976 [89], one exception being the Eta model [73–76] which uses Mesinger's forward-backward scheme, first presented in 1977 [198]. A recent study by Mesinger and Veljovic [199] raises doubts as to the justification of the widespread use of Davies' relaxation LBCs and points out that the scheme makes no effort to respect the basic mathematics of the problem. They compare the Davies [89] and Mesinger [198] LBC schemes on the basis of six 32-day forecasts, assessing for one of the forecasts also the effect of replacing the linear change of relaxation in the Davies approach by Lehmann's optimised choice [200] that minimises the unwanted partial reflection of waves leaving the LAM domain at the boundaries under idealised conditions. Gustafsson *et al.* [201] find that LBC errors are significant for one of two studied cases of poor 12 hour forecasts, while the IC errors are large in both cases.

However, other authors downplay the impact of LBC errors on predictability in limited area weather modelling. Hong and Juang [202] challenge the focus on LBC schemes by suggesting that their orography blending technique applied across the lateral boundary is likely more efficient than elaborating the numerics of a LBC scheme to eliminate systematic error. Davies [109] demonstrates that LBC errors constitute only a small part of the overall error in simulations made with the MetUM, see e.g. [203] for an overview of the history of the development of this finite difference NWP model. Torn *et al.* [204] find that interior-domain WRF model ensemble simulations are not particularly sensitive to whether GCM ensemble forcing or perturbations around a deterministic forecast are applied at the lateral boundary. That is; the study suggests that LAM ensemble forecasts are not very sensitive to LBC errors far away from the lateral boundaries.

Be that as it may, the science of computational mathematics has evolved since the 1970s and 80s when LAM LBCs were intensively studied, though the widespread application

of Davies approach of 1976 [89] leads scientists to believe that the lateral boundary value problem is inherently ill posed, cf. e.g. Hong and Kanamitsu [205]. However, the work by e.g. Funaro and Gottlieb [206], Carpenter *et al.* [207], Nordström and Svärd [208], Hieber and Monniaux [209], and Cheng *et al.* [210] suggest that accurate LBC formulations can indeed be applied for limited area NWP, ensuring *well posedness* of the initial boundary value problem. That is,

- a solution exist,
- the solution is unique,
- the solution's behaviour changes continuously with the initial conditions.

Hesthaven [211] and Svärd *et al.* [212] show how to contruct a semi-discrete, energy-stable and well posed boundary condition scheme for the compressible Navier-Stokes equations, and Kreiss and Wu [213] discuss the relation between stability of semi-discrete and fully discrete schemes for initial boundary value problems.

Davies [214] and McDonald [215] provide overviews of various LBC schemes employed in limited area NWP. The method first proposed by Engquist and Majda in 1977 [216] is revisited by McDonald [217–219]. In his 2005 study, he derives transparent LBCs for two linear systems; a two-layer model of two superposed immiscible fluids of different densities and a multilevel model of the hydrostatic primitive equations in two dimensions (horizontal and vertical), and demonstrate that the boundaries are transparent to outgoing waves while externally imposed incoming waves enter without distortion. Two follow-up studies ensued in 2006 and 2009 where McDonald adressed potential vorticity waves [220] and vertical shear [221], respectively. Termonia and Voitus [222] build on the work of McDonald in trying to disentangle the transparent LBC formulation from the dynamical core of semi-implicit semi-Lagrangian NWP models in order to ease future maintance and scientific development. The following year, Voitus *et al.* [223] introduced a well posed LBC formulation for a one-dimensional model employing Fourier spectral horizontal discretisation, and the study finds improvement in terms of stability properties.

Chapter 4

Weather model deficiency

The limitations on weather model accuracy pertain to the following seven error sources.

1. A fundamental limit on atmospheric dynamics predictability

In his famous paper from 1963 [224], Lorenz presents his discovery of chaotic dynamics, demonstrated by integrating numerically three coupled nonlinear equations describing thermal convection, as proposed the previous year by Saltzmann [225]. The initial value problem considered is a relatively simple coupled set of equations, two of which contain nonlinear terms;

$$\frac{\mathrm{d}x}{\mathrm{d}t} = c_1 x + c_1 y, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = xz + c_2 x - y, \quad \frac{\mathrm{d}z}{\mathrm{d}t} = xy + c_2 x - c_3 z, \tag{4.1}$$

and choosing $0 < c_2 < 1$ results in the stable, steady-state solution $(x, y, z \to 0)$ independent of c_1 and c_3, while $c_2 > 1$ leads to two solutions that are stable – i.e. attractors in the system described by Eqs. (4.1) – for

$$1 < c_2 < \frac{c_1(c_1 + c_3 + 3)}{c_1 - c_3 - 1}.$$

When viewing the temporal development in the phase space spanned by x, y and z, the solutions based on Lorenz' choice of c_2 exceeding the upper stability limit spirals outward from one of the two $c_2 > 1$ stable solutions a number of times, before suddenly transitioning to spiral outward from the other stable solution. The phase space trajectory never repeats and the number of cycles taken around one attractor before transition

to the other cannot be predicted. Moreover, Lorenz shows [224] that an infinitesimal perturbation away from one trajectory may follow the original trajectory closely for a while, but they will eventually diverge completely and become independent. More complex nonlinear systems have been shown to exhibit such chaotic behaviour [226], and as NWP models are based on nonlinear PDEs Lorenz' results for a simpler system hints a fundamental limitation on the predictability of coupled systems of nonlinear PDEs, e.g. Baines [227] mentions 10–14 days ahead.

2. Garbage in → garbage out

Successful application of the theoretical foundation underlying weather and climate modelling is subject to proper initialisation of meteorological fields on the spatial grids on which the governing equations for atmospheric dynamics are integrated, cf. Section 3.7.1. Hence, the feasibility of the endeavour depends on the extent to which the grid-initialising snap-shot data representing the current state of the atmosphere is accurately measured, as well as on the accuracy of previous weather or climate model predicted fields used in combination with observations to propagate the simulation forward.

3. Representing fractal nature on a finite number of pixels

Judd *et al.* [228] view the inability of the discretised governing equations for atmospheric dynamics to properly represent observations across a continuum of spatio-temporal scales as the computer simulation occupying a phase space of much of lower dimension than actual dynamics.

4. Physical processes that do not derive from first principles

Leith [229] distinguishes the former error source from inadequate representation of atmospheric dynamics on numerically unresolved scales, of which Holton and Hakim [125] note the three most important physical process classes to be radiative, cloud and precipitation, and turbulent mixing and exchange, respectively, recalling Section 3.2. Associated challenges are explored e.g. in the study by Brown *et al.* [230], considering boundary layer wind profiles. They address issues related to representation of sub-grid scale

convection in two operational NWP models with spectral and finite difference formulations, respectively, and Milton and Wilson [231] illustrate the sensitivity of atmospheric simulations to the complex balance between parameterisations together representing mechanical dissipation.

5. Approximations applied to avoid numerical artefacts

As pointed out by Charney [22, 24, 232] from the outset of NWP, approximations to the governing equations for atmospheric dynamics are necessary to avoid numerical noise from poorly resolved smallscale dynamics, while preserving phenomena of meteorological interest. More or less severe inaccuracies accompany the choice of approximations applied to remove unwanted, though physically correct, solutions such as acoustic waves. Keller [233] shows that the hydrostatic approximation, commonly employed in operational GCMs, severely impacts the realism of simulated gravity wave propagation. Filtering of smallscale atmospheric dynamics may gradually become a forecast accuracy liability as NWP model resolution increases.

6. From analytical ideal to numerical pragmatics

As discussed e.g. in Linz [234], artefacts of more or less severe impact on results derive from spatial and temporal discretisation schemes applied to the governing equations for atmospheric dynamics in order to implement these numerically. Derivatives estimated by finite difference approximations are subject to unavoidable truncation errors, and e.g. Bénard [235] studies temporal discretisation for a range of schemes commonly used for NWP and finds associated stability properties to be highly dependent on fundamental choices like the vertical coordinate. Novel NWP model formulations have appeared in the literature recently, two finite element based examples being the Icosahedral Nonhydrostatic (ICON-AIP) model by Gassmann [236], designed to conserve mass and energy on a hexagonal grid [237], and the Model for Prediction Across Scales (MPAS) by Skamarock *et al.* [238] featuring smooth model resolution increase across the vicinity of areas of interest for high-resolution NWP modelling. Whether such modern NWP formulations are less prone to suffer from adverse effects sensitive to the choice of discretisation schemes remains an open question.

—— **7.** Bugs uncaught during model code verification ——

Usually left out of discussions on things that can go wrong, the physicist/programmer may be in error at any point in the process of casting and implementing the model equations on discretised form. Elusive issues due to implementation error may indirectly be picked up by weather model output postprocessing frameworks, as discussed in sections 4.1 and 4.2 introducing Appendix C.

4.1 Statistical forecasting by general linear modelling

With regard to points **4.** through **7.** above there are ways to rectify weather model predictions to some extent, and one such postprocessing class is univariate general linear modelling. A mathematical relation for an observed quantity – wind speed in the context of the study introduced in Section 4.2 – denoted W is assumed related to a series of additive terms, each of which taken to be proportional to an unknown model coefficient θ_i,

$$\mathsf{W} = \theta_0 + \sum_{i=1}^{n} x_i \theta_i + \varepsilon, \tag{4.2}$$

and the relation is modelled as subject to stochastic error ε, for which a distribution is usually assumed for θ_0 and θ_i inference purposes. For normal theory models the method of ordinary least squares (OLS), based on squared model residual minimisation, corresponds to Maximum Likelihood estimation [239], and this approach is employed in Appendix C to be introduced in the next section. The key challenge is to identify x_i predictors able to systematically account for measurements corresponding to the model response W.

Classical statistical forecasting does not involve weather models in that the modelling task is to construct regression equations relating atmospheric measurements for one or more predictors to a response observed a specified forecast leadtime ahead. An application of the technique for prediction of hurricane trajectories is outlined in Miller *et al.* [240], using observed geopotential heights and geostrophic winds, among other atmospheric observations, as predictors in a set of regression equations developed separately

for four leadtime horizons each spanning 12 hours. Today, such purely statistical models are only competitive in nowcasting applications that concern leadtimes up to a few hours ahead, at least in a wind power forecasting context [6, 7]. Klein *et al.* [241] first combine statistical modelling of an observed response variable, related mathematically to simultaneously observed predictor variables, with evaluation based on corresponding NWP model variables as predictors. In contrast to the purely statistical approach this method, known as *perfect prog*, does not involve a time lag between predictor(s) and response, in that the idea is to develop the statistical model solely using observations, and then apply it using corresponding NWP derived predictors. Hence, the perfect prog method do not account for NWP model error. MOS is a third approach, pioneered by Glahn and Lowry [95], where both statistical model development and evaluation is based on NWP datasets with matching forecast leadtimes and the MOS technique can therefore be used to quantify systematic NWP model errors. In a subsequent study comparing the latter two techniques for probability of precipitation forecasts, Klein and Glahn [242] find better improvement from the MOS postprocessing of NWP data.

The study introduced in Section 4.2 is based on MOS modelling.

4.2 Introduction to Appendix C

Representing weather and climate with numerical models based on well-established theory [111] entails multifarious simplifications and approximations. With each choice of discretisation scheme and parameterisation for sub-grid scale processes discrepancy between the practical implementation of the governing equations for atmospheric dynamics and observations is potentially introduced. Figure 4.1 illustrates physical process interactions that need to be represented numerically, and it is not hard cf. Section 3.2 to imagine e.g. error in the solar radiation parameterisation to propagate into other physical processes as a function of model integration time, with adverse implications for the simulation accuracy of moisture flux, atmospheric stability, etc. which ultimately affects prediction of atmospheric flow, the weather feature of interest in this work.

The study presented in Appendix C concerns MOS postprocessing of GCM data, based on a wide range of predictor candidates, and hence indirectly addresses weather model deficiency. That is, the succesful MOS modelling predictor candidates provide guidance

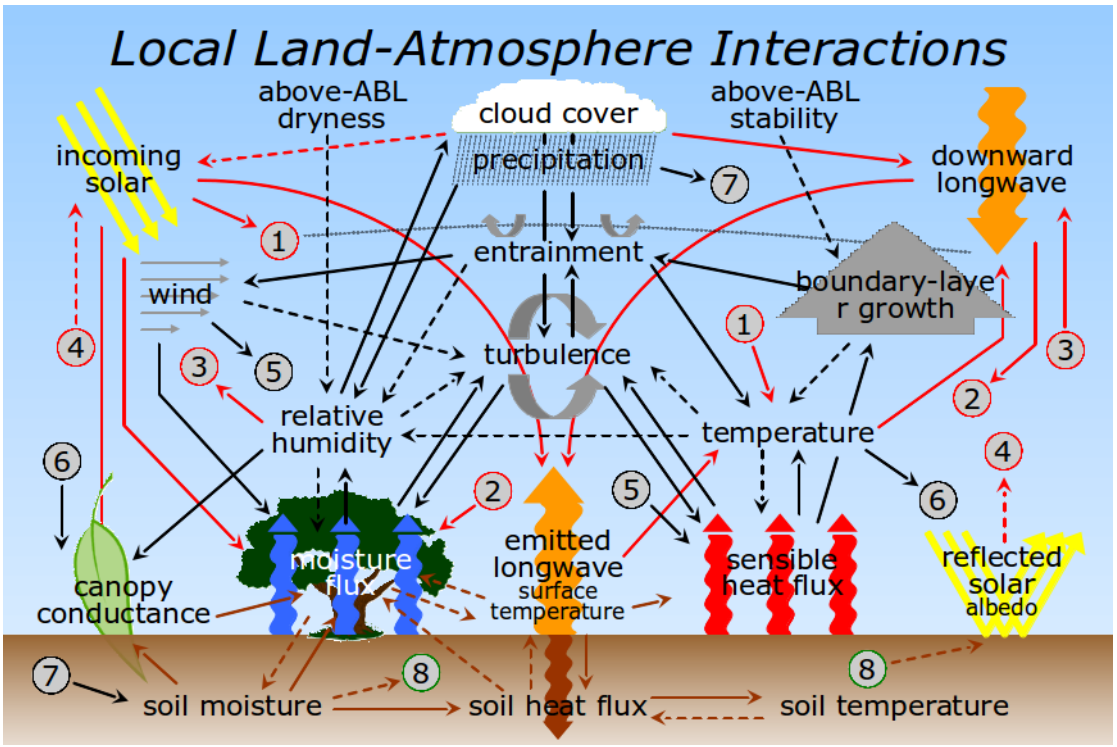


FIGURE 4.1: Illustration of the complexly coupled physical processes that need to be accurately simulated in weather and climate models in order to reproduce surface measurements of atmospheric variables. Black arrows denote surface and boundary layer processes, red arrows radiation, while green arrows are land surface processes. Solid (dashed) lines represent positive (negative) feedback. Circled numbers 1 to 8 display the origin and effect of the sketched physical processes, respectively. The figure is adapted and reproduced with permission [243].

as to which GCM formulation aspect should be in focus with respect to future upgrades of the weather model. Results presented in *Appendix C Section 4* reflect forecast improvement, achievable in practice operationally, from statistical postprocessing by the method of *Appendix C Section 3* applied to forecasts derived from a state of the art NWP model providing freely available global forecasts; the GFS.

In contrast to the LAM study of Chapter 3 and appendices A, B, the study presented in Appendix C is based on a two-year dataset, and the forecast reference data is processed according to the approach specified in Chapter 2. As remarked e.g. in Jacks *et al.* [244], having several years of NWP data available for MOS model parameter estimation increases the robustness of statistical forecasts. However, major and minor upgrades of operational GCMs are regularly carried out which may not impact regression model performance severely for minor NWP model changes, as observed e.g. by Erickson *et al.* [245], though statistical model development should in general be based on data from a

static NWP model implementation to avoid the risk of making false conclusions on the explanatory value of one or more predictor candidates. Müller [246] proposes adaptive short-term MOS, with daily model parameter updates based on data for the previous 30, 60 and 90 days, respectively, allowing the MOS equations to "recover" consistency faster than is the case for longer training datasets traversing NWP model upgrades. The inconsistent-dataset issue is circumvented in the study of Appendix C by selecting data from a time period during which no changes were made to the GFS. Hence, the data period commences a month after a land surface model bug fix was implemented and the GCM remains unaltered until the dataset ends 30 September 2014 (+ 2 days).

4.3 Multicollinearity

Provided two or more Eq. (4.2) predictors are highly correlated, the model design matrix $\boldsymbol{M}$, containing predictors along the column dimension and observations along the row dimension, multiplied with it's transpose, is near-degenerate since $\boldsymbol{M}$ itself is near-degenerate,

$$\det\left(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}\right) = \det\left(\boldsymbol{M}^{\mathrm{T}}\right) \cdot \det(\boldsymbol{M}).$$

Recasting Eq. (4.2) in matrix notation,

$$\mathsf{W} = \begin{bmatrix} 1 & \boldsymbol{x}^{\mathrm{T}} \end{bmatrix} \boldsymbol{\theta} + \varepsilon,$$

with $\boldsymbol{x}$ and $\boldsymbol{\theta}$ respectively denoting column vectors with x_i and θ_0 & θ_i elements, the OLS estimator for $\boldsymbol{\theta}$ subject to the assumption of independent and identically distributed (i.i.d) ε is

$$\hat{\boldsymbol{\theta}} = \left(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}\right)^{-1} \boldsymbol{M}^{\mathrm{T}}\mathbf{W},$$

with covariance

$$\mathrm{Cov}\left[\hat{\boldsymbol{\theta}}\right] = \sigma_{\varepsilon}^{2}\left(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}\right)^{-1}$$

cf. e.g. Pawitan [247]. The model coefficient estimates become uncertain, cf. e.g. Valentine [248], which may lead to relatively large model coefficients – of opposite sign

– being attributed to the two (or more) correlated predictor terms. Thus, assuming two severely collinear predictors, each of these may appear to explain a large portion of the model variation, while the combined effect of the two in fact may almost completely cancel. In terms of the variation histogram approach for predictor relative-value screening, cf. *Appendix C Section 3.5*, severe collinearity would result in wide 95 % quantile ranges for the two said predictors. In order to detect such false predictor-value signals, $x\hat{\theta}$ variation histograms can be supplemented by ordering according to forecast performance reduction due to removal of a given predictor. That is, the predictor term causing the most severe performance reduction when removed from the model is at the top of the hierarchy, and so forth as the adverse performance impact from predictor removal decreases. If a predictor with a relatively wide 95 % quantile range assumes a low ranking this indicates strong interpredictor coupling that may be confirmed from interpredictor correlation matrix and scatter plots, cf. *Appendix C Section 3.3*. Valentine [248] proposes to substitute the individual terms with a single linear combination of the collinear predictors, though also emphasises the entailing challenge of choosing appropriate weights for each of these. Hence, his other suggestion of simply leaving out one of the collinear predictors is more straight-forward. To screen all predictors for true explanatory value these are left out by turn, the model re-assessed, and severety-ranking of adverse impact on performance in terms of accuracy metrics can be applied to root out exaggerated-value (due to collinearity) predictors. Thus, inconsistency between $x\hat{\theta}$-variation contribution vs. the same term's removal severity-ranking leads to exclusion of said predictor.

4.4 Autocorrelation

Another issue of the statistical modelling approach employed in Appendix C is the assumption of independent observations. If the wind speed was 1 m/s five minutes ago, the current wind speed is not likely to be far from this value, say 25 m/s. As measured and forecasted wind speed time series are both subject to finite autocorrelation, MOS model residuals ϵ_i will exhibit some autocorrelation — or put differently; there will be nonzero off-diagonal (covariance) elements in the variance matrix for ε cf. Eq. (4.2). Continuous data typically exhibits finite autocorrelation and hence an assumption underlying the predictor-reduction process involving the Bayesian Information Criterion (BIC) is not

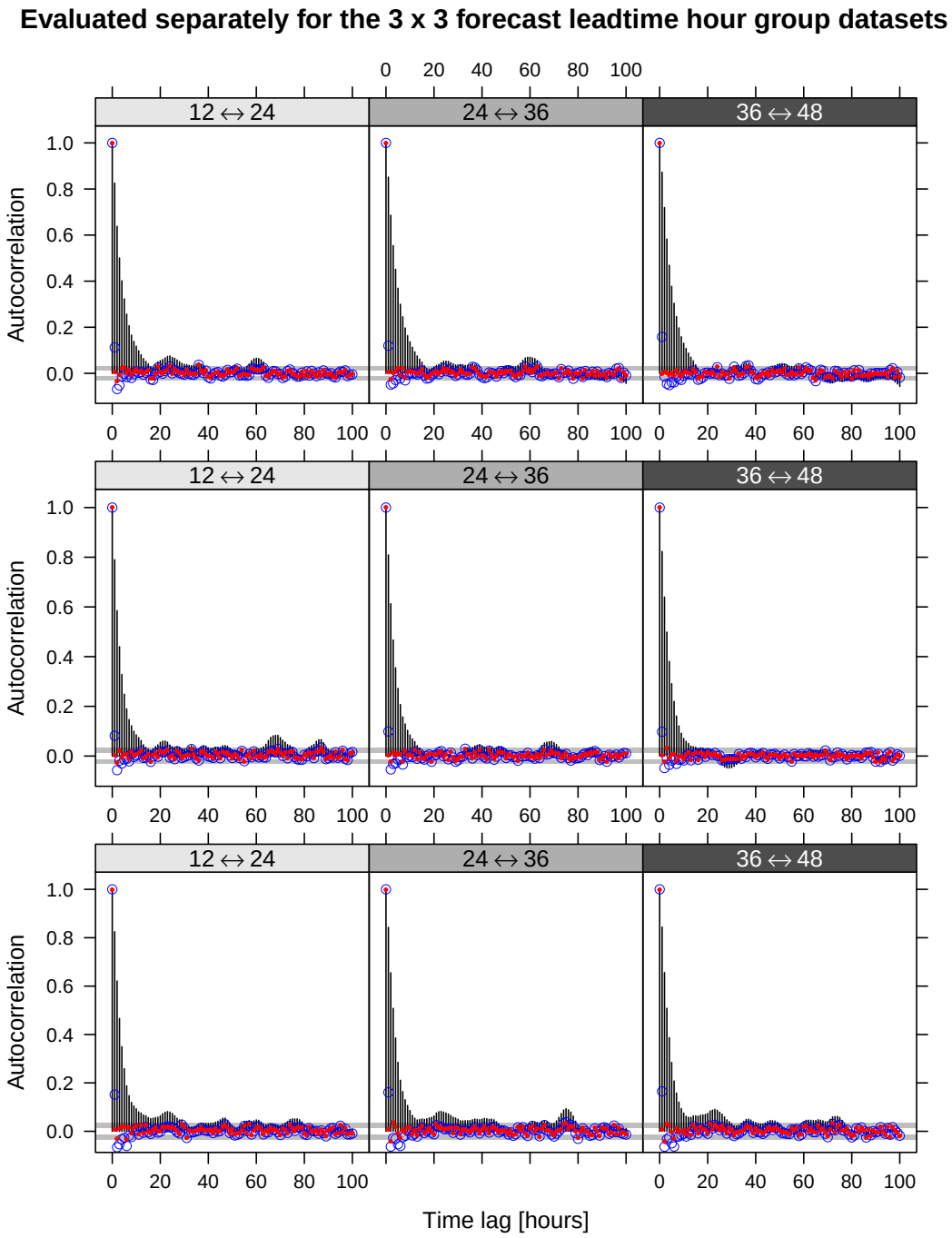


FIGURE 4.2: The black, vertical bars denote autocorrelation for the 12 ↔ 24, 24 ↔ 36, and 36 ↔ 48 forecast leadtime groups considered in Appendix C (year 1). The blue circles denote autocorrelation predictions from a first-order autoregressive model, AR(1), for the residuals of the full MOS model considered in Appendix C. Similarly, the red bullets mark AR(2) model predictions. The grey, horizontal lines delimit an approximate white noise 95 % confidence interval cf. [249]. TOP: The offshore Horns Rev wind farm. MIDDLE: Coastal Rejsby Hede. BOTTOM: Stor-Rotliden, situated in central Sweden.

satisfied, and redundant predictors likely survive the process cf. the discussion in *Appendix C Section 3.4*. Hence, the final development step based on variation histograms is introduced.

Alternatively the MOS model considered in Appendix C can be brought to account for autocorrelation by inferring the number of autoregressive (AR) process parameters needed to model residual autocorrelation, and then integrate the chronological ϵ_i-dependency into the residual matrix structure, cf. e.g. Gałecki and Burzykowski [250]. Cf. e.g. Madsen [251], the p-order autoregressive process for Appendix C MOS model residuals ϵ_i – AR(p) – is written,

$$\epsilon_i + \epsilon_{i-1}\theta_1 + \epsilon_{i-2}\theta_2 + \cdots + \epsilon_{i-p}\theta_p = \varepsilon_i, \tag{4.3}$$

where ε_i here denotes a white noise process cf. [252]. Figure 4.2 indicates that suitable inclusion of AR(2) process parameters θ_1 and θ_2 in the residual variance-covariance matrix (i.e. nonzero off-diagonal elements are introduced) for the MOS model studied in Appendix C can effectively circumvent the autocorrelation issue by modelling the chronological dependence as part of the Appendix C MOS model coefficient estimation, introducing θ_1 and θ_2 coefficients of Eq. (4.3) as additional parameters to be estimated. However, the computational expense was found to increase exponentially with the number of data points included, and statistical model fitting turned out to be infeasible for the purpose at hand; iterative BIC-reduction for nine datasets each based on yearlong hour-resolved time series.

4.5 Error-free predictors

A final remark pertaining to the approaches outlined in appendices A and C concerns the implicit assumption of error-free predictors. All predictors considered in Appendix C derive from GFS forecasts, initialised in part with 1) surface and upper-air observations of the current state of the atmosphere, and 2) meteorological fields derived from the previous GFS forecast. Each are subject to errors related to the data assimilation process constituting forecast initialisation, due to

- the spatio-temporal grid of measurements is finitely and unevenly resolved,

- mathematical/technical quirks of the data assimilation scheme employed,
- sensitivity of predicted trajectories to initial condition accuracy cf. [47, 224, 227],
- errors associated with measurement devices — satellites, measurement towers, etc.

Jonsson [253] considers the influence of predictor errors in a simple linear regression model, and his study suggests that ordinary OLS estimation is quite feasible subject to heteroscedastic predictor uncertainty. Effort toward accommodating above-discussed error sources in the modelling framework laid out in Appendix C is not pursued in this work.

Chapter 5

Conclusion & further research

The studies described in part aim to quantify the prospects of limited area forecasting for use in operational wind energy scheduling, and also propose and assess an original methodology for inferring deficiencies in the discretised formulations for atmospheric dynamics under the hood of the weather models; the predictions of which are essential to electrical grid stability subject to substantial wind power penetration.

For the particular WRF model configuration selected for the limited area forecasting study, offshore and coastal LAM performance is inferior to corresponding GCM data when assessed in terms of the scalar accuracy metrics of interest for wind energy scheduling. From visual inspection of individual wind speed forecasts it is found that meteorological features on short time scales are indeed largely captured in the high-resolution forecasts, yet subject to some phase error which is severely penalised by the forecast performance metrics. Hence, a spatio-temporally coarse scale GFS forecast oblivious to fast atmospheric flow fluctuations may achieve better scores in terms of the relevant assessment criteria, and for the full-year dataset considered the adverse effect of phase errors in finely resolved forecasts outweight the insensitivity to local terrain and temporal volatility of a GCM. Inland, a dip in forecast performance is observed upon numerically translating GCM data to the outermost LAM grid of comparable horizontal resolution, though wind speed forecast performance steadily increases beyond that of the GFS at high resolution. However, this LAM superiority for wind speed forecasts does not transfer to the wind power domain.

Temporal smoothing of wind speed forecasts is shown to consistently improve performance, further supporting the notion that slight phase errors account for part of the inferiorness of raw LAM data. Applying this simple NWP model forecast postprocessing renders inland high-resolution LAM forecasts outperforming corresponding GCM predictions, albeit by less than 1 % of wind farm capacity. From the conditional wind speed modelling on prognostic and diagnostic weather model simulated quantities, a pronounced dependence on the geographical region is established, and the method is less feasible in terms of wind power at the coastal wind farm — except when considering the surface pressure variable. The five other quantities are all strongly influenced by the local topography and land surface model response, while atmospheric pressure largely reflects synoptic scale dynamics. Hence, physical processes in the vicinity of the coastal site may not be properly represented in the WRF model simulations. Inland, in complex terrain, temperature, moisture, turbulent kinetic energy and Monin-Obukhov length are all found relatively valuable; the two former to a greater extent for in-sample assessment in terms of wind speed, while the two latter exhibit more potential when assessment is instead based on adaptive and recursive (out-of-sample) wind power forecast evaluation. These four quantities are all strongly coupled to the land surface model forcing, and offshore – where surface feedback is more smooth – all six modelled quantities display slight to moderate potential in a wind power forecast modelling context.

As also inferred from individual wind speed forecast data, power spectral density estimates show that the WRF modelled flow dynamics approaches corresponding spectra based on measurements as a function of grid resolution. However, leaving spectral space, a closer inspection of LAM wind speed forecast performance as a function of integration time reveals that the wind speed predictions of the WRF model configuration gradually lags behind simultaneously measured wind speed time series, and a tendency towards unstable weather conditions with forecast leadtime is revealed. This could e.g. be caused by excessive heat exchange from the ground, perhaps due to exaggerated parameterised amounts of downward short wave radiation, which in turn may point towards the microphysics Thompson scheme representing cloud fractions. To further infer the presence of temporal lag between predictions and observations, the performance of a wind speed ramp detection algorithm not penalised for slight phase errors – as are scalar accuracy metrics – is evaluated, and consistent performance increase with LAM resolution is found in terms of this assessment based on meteorological objects with implications for wind

energy scheduling. A linear relation between LAM forecasted wind speed variability and wind speed uncertainty is established for offshore and coastal sites, while the relation is less evident inland. This link is re-addressed in the discussion on further research below.

The statistical postprocessing framework proposed for correction of NWP model deficiencies results in seven competing MOS models, and the two fundamentally different strategies employed for performance assessment in terms of wind speed and wind power, respectively, points toward the simulated processes underlying the lifted index diagnostic as containing systematic coupling to the wind speed forecast error. To further improve the methodology, the ignored – albeit very present – autocorrelation in wind speed can be accounted as explained, yet as also mentioned the associated computational costs were found too severe for the application at hand. If these practical limitations can be circumvented, it is also possible to address heteroscedasticity via model residual covariance modelling.

Concerning the LAM configuration employed in the limited area forecasting study, the uneven LBC resolution around the outer LAM domain – due to the degree-resolved GFS data – may potentially induce horizontally-skewed forcing of meteorological fields on LAM grids. Equidistantly-resolved LAM forcing data may result in more spatially balanced numerical simulations of the atmospheric fluid and hence yield better forecasts. Also, the adverse effects associated with warm LAM initialisation could turn out to be less severe than the cold start shock that the presented LAM study is subject to. Of more general concern is that each combination of the numerous NWP modelling aspects together entail a multitude of alternative solutions — though only one configuration is investigated. Exploring different LAM configurations accounting for all seasons, and with enough data points to support statistically candid statements, is not feasible in practice. To put things into perspective, the LAM study carried out here ran on the university's high-performance computing cluster during a time period of three months, and afterwards broken simulations had to be identified and patched up. Several years worth of human hours and computational resources are necessary to assess every physics parameterisation setup, model resolution sensitivity, spatial extent of computational domains, LBC formulations, data assimilation schemes and forecast initialisation approaches, etc., and new NWP model schemes are continually introduced. Hence, in order to efficiently probe the feasibility of several weather forecasting tools and/or configurations of these, future studies along the lines of isolating *distinct synoptic weather*

situations with *distinct impacts* on wind power forecasting is proposed. One such approach could e.g. be based on SOMs as in Vincent and Hahmann [92], identifying days with severe wind speed variability as discussed last in the introduction. The method proposed by Lange [90] relates synoptic weather situations to wind speed forecast error based on principal components and clustering analysis, and hence addresses the task at hand; see also Lange and Focken [91]. Extending aforementioned endeavours further into applicability for efficient objective identification of approximately optimal LAM configurations for a given geographical site or region will in turn promote the vision of substantial wind power penetration viably supplemented by other energy sources. In the mean time, statistical postprocessing such as the MOS method suggested here finds practical use for wind energy scheduling, and another recent development is the analogue ensembles approach, cf. e.g. [254, 255]. In the context of the wind power uncertainty analysis performed in this work, the SOMs technique could be envisioned as a largescale component in a realtime wind power uncertainty warning system, based primarily on forecasted wind speed variability to reflect more local and short term scales.

In a broader NWP modelling context, Wedi recently pointed out [256] that more sophisticated strategies than mere blunt increase of the horizontal resolution are needed in the near future in order to accommodate deeper understanding of multiscale interactions in the atmosphere and the corresponding numerical implementation on future massively parallel computer architectures. In a few years time NWP formulations may well incorporate stochasticity, cf. e.g. Annunziato [257, 258] and Palmer [259], and as remarked previously and encouraged in Wedi [256] the sophistication of NWP formulations and the associated practical implementability is an active research topic, and novel approaches are frequently added to the growing literature on the subject of numerical weather forecasting [260, 261].

Appendix A

Journal publication I — *draft*

WIND ENERGY
Wind Energ. 2015; **00**:1–16
DOI: 10.1002/we

RESEARCH ARTICLE

Limited area forecasting for wind energy scheduling. Part I: Performance assessment and enhancement studies

Martin H. Rosgaard[1,2], Andrea N. Hahmann[1], Torben S. Nielsen[2], Gregor Giebel[1], Henrik Aa. Nielsen[2], Henrik Madsen[3]

[1]DTU Wind Energy, Technical University of Denmark, Risø Campus, DK-4000 Roskilde, Denmark. [2]ENFOR A/S, DK-2970 Hørsholm, Denmark. [3]DTU Compute, Technical University of Denmark, Lyngby Campus, DK-2800 Kgs. Lyngby, Denmark

ABSTRACT

The implications of global- and limited area weather forecast input for wind energy scheduling is assessed and compared, based on twice-daily 48 hour operational Global Forecast System (GFS) data and Weather Research and Forecasting (WRF) model simulations for limited area computional grids. The forecast reference dataset spans wind farm sites offshore, onshore near the coast, and inland in complex terrain. For the offshore and coastal sites, both in the North Sea region, high-resolution forecasting is found to deteriorate performance in terms of scalar accuracy metrics, while consistent wind speed improvement with resolution, beyond the global forcing data, is found for the site in complex topography inland. Temporal smoothing of individual wind speed forecasts is shown to improve wind speed and power predictions, and peak performance improvement is achieved at tricube-weighted inclusion of past and future information up 7 hours away from the time stamp of the smoothed forecast data point. Systematic explanatory value of WRF model simulations for boundary layer height, surface pressure, temperature, moisture, turbulent kinetic energy, and Monin-Obukhov length is established within a conditional parametric modelling framework. The particular implications of each of these variables for wind energy scheduling is found to depend on the region considered; offshore the prognostic variables temperature and moisture contain most explanatory value, on the coast only pressure has a positive impact in terms of power forecast performance, and inland turbulent kinetic energy and Monin-Obukhov length systematically enhance predictability of the atmospheric flow more than do the other mentioned variables. Copyright © 2015 John Wiley & Sons, Ltd.

KEYWORDS

numerical weather prediction; wrf mesoscale wind speed modelling; wind power forecasting; conditional parametric modelling; wind farm observations; wind turbine anemometer measurements

Correspondence

Martin Haubjerg Rosgaard, Technical University of Denmark, Department of Wind Energy. Email: mhros@dtu.dk

Contract/grant sponsor

Danish transmission system operator *Energinet.dk* public service obligation

Contract/grant number

PSO 10464

Received . . .

[†]Please ensure that you use the most up to date class file, available from the WE Home Page at http://www3.interscience.wiley.com/journal/6276/home

1. INTRODUCTION

The European Technology Platform for Wind Energy (*TPWind*) estimates that wind power could account for 34 % of the electricity consumption in the European Union (EU) by 2030 [1], corresponding to as much as 800 thousand jobs in the EU alone [2]. The scenario in turn magnifies the extent to which inaccuracy in day ahead assessment of this highly fluctuating renewable energy source has adverse impact on the practical feasibility of wind power integration in the electrical grid. A handfull of institutions worldwide carry out global numerical weather prediction (NWP) on a daily basis, cyclically updated with the latest observations facilitated by the World Meteorological Organisation for each forecast initialisation, and thus provide essential input for competitive wind power trading on the energy markets. Although atmospheric dynamics relevant for renewable energy scheduling is well understood theoretically, the continuum of spatio-temporal scales involved, from largescale synoptic systems down to microscale cloud processes, cannot be represented fully in NWP models primarily due to practical limitations on computational resources. However, the resolution of general circulation model (GCM) forecasts can be indirectly increased by embedding one or more limited area model (LAM) computional domains covering geographical regions of interest, taking initialisation data and lateral boundary conditions from the GCM.

A recent study [3] briefly considers day ahead wind speed prediction accuracy subject to different LAM resolutions, using the LAM also employed in this work, yet decides to drop the 4 km horizontally resolved computational domain prior to further analysis due to performance inferior to that of the coarser 10 km parent domain. Another LAM study for wind energy scheduling [4] is based on the Eta model, cf. e.g. [5] for a recent description, though only considers the performance of the highest resolved computational domain and initialise forecasts with reanalysis data — which is not available in a realtime setting. An older study [6] compares performance of three LAMs and the Global Forecast System (GFS) [7] and find no wind speed forecast improvement for the high-resolution 1.33 km domain, though e.g. [8] reports more realistic meteorological structures when going from 10 km to 2.5 km resolution using the Regional Atmospheric Modeling System (RAMS) [9]. The present work supplements this literature, and several other related works not cited here, with a meticulous LAM performance study covering different model resolutions and terrain types, targetting application within wind energy scheduling.

The LAM employed is the Weather Research and Forecasting (WRF) model using the Advanced Research WRF (ARW) dynamics solver [10], and the GCM data derives from operational GFS forecasts. The WRF model is an open source community LAM in widespread use globally, the underpinning methodology relies on conformal map projection to a physical grid, and many contributors continually improve and test numerical schemes employed in the discrete formulation and sub-grid scale parameterisations for physical processes. GFS data at 0.5° and three-hourly resolution is freely available online, with approximately 5 hours delay from forecast initialisation time. Hence, the data and tools employed in this study are well-tested and readily available for realtime application.

The article is structured as follows. Section 2 specifies the details of the study, and results on forecast performance at different spatio-temporal NWP model resolutions are presented in Section 3, along with forecast enhancement assessment based on temporal smoothing and semi-parametric statistical modelling, respectively. The conclusion is given in Section 4.

2. EXPERIMENTAL DESIGN

The LAM study generating datasets analysed in Section 3 is summarised in Table I and Figure 1, and the governing equations for atmospheric dynamics discretised in the WRF model are integrated 48 hours beyond initialisation. Note that the lateral boundaries of the 10 km domains are not symmetric around the forecast target sites; the steep gradients of the Norwegian mountains may incur numerical artefacts superposing – and cf. [18] hence potentially distorting – meteorological features of interest translated from the 30 km domains into the 10 km domains, for the two WRF model configurations respectively, through the lateral boundary conditions (LBCs) [19] to finer discretisation. In severe cases

Wind Energ. 2015; **00**:1–16 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

Experiment aspect	Approach taken
Study period	23 May 2012 to 24 May 2013
LAM employed in the study	Advanced Research WRF model [10] version 3.4.1
Dataset generated	WRF model forecasts initialised daily at 00 and 12 UTC and time-integrated up to 48 hours from initialisation
Locations- and type of forecast reference data	Wind farm data from three sites offshore, on the coast in flat terrain, and in heterogenous topography inland
WRF model initialisation and lateral forcing	Extracted from operational GFS [7] data available at 0.5° horizontal resolution, lateral forcing is three-hourly
Configuration of WRF model runs	Two separate same-resolution configurations; 1) covering offshore and coastal sites, 2) covering the inland site
Horizontal resolution in WRF model runs	Four computational grids, one-way nested by increasing-resolution order, with 30 km, 10 km, 3.3 km, 1.1 km
Vertical resolution in WRF model runs	40 terrain-following hydrostatic pressure levels up to 50 mb, of which four are within $\sim$100 m above the ground
Integration time steps in WRF model runs	120 s on 30 km grids, 40 s (10 km), 13.3 s (3.3 km), and 4.4 s (1.1 km)
Sub-grid scale boundary layer dynamics	Parameterised with the Mellor-Yamada-Nakanishi-Niino (MYNN) level 2.5 local closure scheme [11, 12]
Modelling of land surface processes	Based on the Unified Noah Land Surface Model [13]
Representation of microphysics	By the improved Thompson scheme [14]
Simulation of solar radiation	Represented according to Dudhia's method for shortwave radiation [15] and the RRTM for longwave [16]
Cumulus parameterisation	Kain-Fritsch's approach [17] is enabled for 30 km and 10 km domains, no fine-grid cumulus parameterisation

Table I. Outline of the LAM experiment.

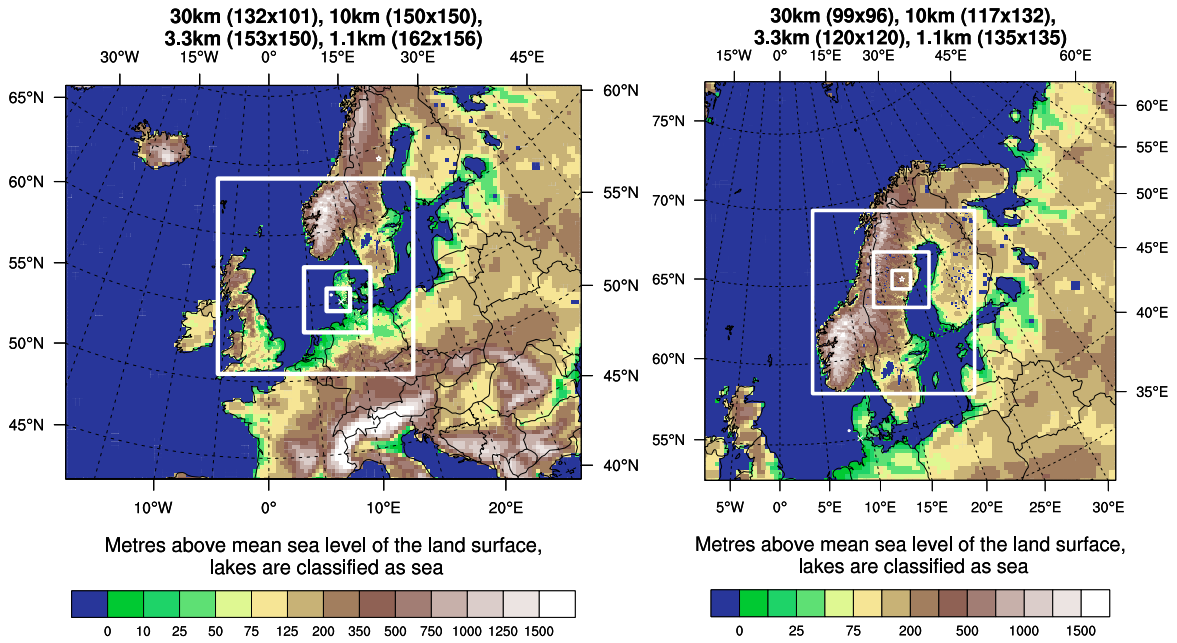


Figure 1. The two WRF model configurations employed in the study. Each horizontally-rectangular computational domain delimited in white is nested within the immediate parent in that lateral boundary conditions deriving from the latter is applied – one-way only – to the former as time integration progresses. The GFS forces the 30 km grids laterally, cf. Table I. The offshore wind farm, Horns Rev, is marked with a white bullet, the coastal Rejsby Hede is a white cross, and the inland Stor-Rotliden wind farm a white star.

simulations crash, which in turn exclude forecasts from the dataset and thereby bias conclusions drawn from this study as extreme weather situations would be underrepresented. Thus the lateral expansion of the 10 km computational domain boundary beyond the Norwegian mountains.

Further motivation for the particular WRF model configuration employed is given in [20]. An outline of recorded WRF model forecast data analysed in this study is given in Table II.

2.1. Forecast performance assessment strategy

On the world's largest day ahead market for power trading the energy auction closes at noon CET [21] and consequently only forecast leadtimes beyond 12 hours ahead of this deadline are relevant for wind energy scheduling. In order to resolve forecast leadtime dependency, two daily GFS initialisations are included in the study; 00 and 12 UTC, and hence three forecast leadtime horizon groups are available to resolve wind speed forecast performance dependency on the

VERTICAL LEVEL	METEOROLOGICAL VARIABLE
For each vertical WRF model layer	Wind speed
For each vertical WRF model layer	Wind direction
For each vertical WRF model layer	Turbulent kinetic energy estimated in the MYNN scheme, cf. Table I
Fluctuating	Planetary boundary layer (PBL) height estimated in the MYNN scheme
Interpolated to 2 m AGL	Temperature
Interpolated to 2 m AGL	Water vapour mixing ratio
Characterising vertical profile of surface layer flow	Monin-Obukhov Length derived from the MYNN surface layer parameterisation
At the surface	Pressure

Table II. Subset of WRF model variables extracted every 10th minute.

leadtime; namely (12,24], (24,36] and (36,48] hours ahead of forecast initialisation. N.B.: Leadtime-resolved performance assessment is relaxed in the wind *power* forecast evaluation, where a single forecast leadtime hour span is considered, namely hours 12 to 30. Also, in terms of power the 5 hour lag on GFS data availability is accounted for, albeit no additional delay is assumed for WRF model data availability as this depends on computational resources at disposal.

For a broad overview and accompanying in-depth discussions on forecast validation metrics, the reader is referred to e.g. [22, 23]. In the present context, two scalar accuracy metrics commonly used in wind speed and power forecast performance assessment are the root mean square error (RMSE), and the mean absolute error (MAE) [24, 25, 26, 27]. Forecast skill score in terms of mean square error can be shown [28] to be decomposable into bias (unconditional) and scaling errors (conditional) contributions, along with a third contribution; the squared Pearson correlation coefficient (PCC). In contrast to RMSE and MAE, the PCC metric is not sensitive to forecast bias and scaling errors and hence ordinary least square (OLS) fits to the normal theory [29] linear regression model for wind speed W,

$$\mathsf{W} = c + w\theta + \varepsilon, \tag{1}$$

with intercept c, slope θ, unexplained variation ε, and wind speed predictor w derived from NWP model forecasts and assumed error-free (cf. e.g. [30]), will have sample PCC equivalent to that for raw w data, when assessed based on a reference dataset of wind speed measurements w. This invariance to linear transformations is useful for obtaining information on forecast performance post statistical correction by Eq. (1) — based on the *unprocessed* forecast data w. In-sample fitted values for the Eq. (1) model response W are henceforth termed linear least square (LLS) wind speed, following the convenient naming convention employed in [27]. In summary, forecast performance accuracy metrics employed include

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - o_i)^2}{N}} \qquad \mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|x_i - o_i| \qquad \mathrm{PCC} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(o_i - \bar{o})^2}} \tag{2}$$

with N data points of simultaneously recorded forecast variable x and observation o representing wind speed or power, and their respective dataset means $\bar{x}$ and $\bar{o}$. As wind power is usually traded in hourly intervals [31] the wind speed average in the hour leading up to said time stamp is suitable for the application at hand and hence employed.

2.2. Forecast reference datasets and NWP model variable field extraction

Wind turbine anemometer readings of wind speed are averaged spatially, i.e. across individual-turbine time series data points, for the three wind farms marked in Figure 1, respectively. In contrast to the case for mast measurements, rapid smallscale fluctuations that no NWP model is able to accurately capture within a NWP model grid cell are hence evened out, resulting in a less noisy reference for forecast performance assessment. Out of the yearlong time period considered 99.1 %, 96.7 %, and 94.7 % of the dataset is available post quality control for the offshore Horns Rev (HR), coastal Rejsby

Wind Energ. 2015; **00**:1–16 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

Hede (RH), and Stor-Rotliden (SR) in complex terrain, respectively. In the GFS dataset consisting of 730 48 hour forecasts 11 of these are missing and hence 11 WRF model forecasts are unavailable in the study. When aligned with the observation datasets this renders 96.9 %, 94.1 %, and 92.9 % of the data points covered by the time period specified in Table I available in the study of forecast performance for HR, RH, and SR wind farms, respectively. Details on wind power production datasets are deferred to Section 3.2 as forecast performance in terms of power is not considered in Section 3.1.

The computational grids in the two WRF model configurations sketched in Figure 1 are designed with a horizontal grid point at RH and SR coordinates in their separate configurations, respectively, for all model resolutions. The wind turbines constituting the coastal site, RH, are spread across an area of roughly $1 \times 2\,\mathrm{km}^2$, and hence for this site only WRF model grid points at the mean-coordinate of the farm-constituent wind turbines are included for all NWP model resolutions. The offshore and inland sites HR and SR each cover an area around $4 \times 5\,\mathrm{km}^2$ and hence information from horizontal nearest neighbour computational grid points is considered for WRF model forecasts for these sites. The tricube weight function, typically an element in local smoothing techniques [32, 33, 34],

$$w_t(\boldsymbol{x}) = \begin{cases} \dfrac{\left(r - \|\boldsymbol{x} - \boldsymbol{y}\|\right)^3}{\sum\limits_{\|\boldsymbol{x}-\boldsymbol{y}\|<r} \left(r - \|\boldsymbol{x} - \boldsymbol{y}\|\right)^3} & \text{for } \|\boldsymbol{x} - \boldsymbol{y}\| < r \\ 0 & \text{for } \|\boldsymbol{x} - \boldsymbol{y}\| \geq r \end{cases} \tag{3}$$

is here used to process WRF model variable fields horizontally at the HR and SR sites for comparison to farm-averaged wind speed measurements. The tricube weight $w_t(\boldsymbol{x})$ is only nonzero when the smoothing radius r is greater than the Euclidean distance $\|\boldsymbol{x} - \boldsymbol{y}\|$ between coordinates $\boldsymbol{x}$ for the grid point to be assigned a weight and the wind farm centre coordinates $\boldsymbol{y}$. The r values for HR and SR are estimated by assessing a range of smoothing radii in terms of PCC performance for each forecast leadtime horizon group and choosing the r optimising performance for each computational grid resolution, cf. [35]. Horizontal WRF grids at the second, third and fourth vertical WRF model layers are suitable for comparison to datasets for RH, with wind turbine nacelle anemometers 45 m AGL, for HR this is 70 m AGL, and for SR at 95 m AGL.

W.r.t. the GFS data, the hub-height datasets for the four horizontally nearest neighbour GFS grid points are bilinearly interpolated to wind farm coordinates. For the RH site, GFS wind speeds 10 m and 80 m AGL are vertically interpolated to the 45 m AGL hub-height pertaining to this coastal wind farm according to the logarithmic wind profile [36].

3. RESULTS

Wind power prediction relies on power curve estimation [37] and statistical correction is typically employed, e.g. in terms of linear correction cf. Eq. (1) or more advanced model output statistics (MOS) [38], and hence RMSE and MAE based on LLS wind speed rather than raw forecasts are assessed here. For each of the three sites considered, regression modelling theory prescribes a static relation between the PCC and RMSE accuracy metrics for LLS wind speed forecasts via the coefficient of determination, cf. e.g. [29]. Table III supports the notion that the relative comparison of computational grid performance by the PCC accuracy metric for raw forecast data corresponds to that for RMSE post standard processing by linear correction according to Eq. (1), the latter relying on wind speed measurements. Hence, the PCC is exclusively applied for wind speed data, while MAE and RMSE are considered for wind power forecast performance assessment in sections 3.2 and 3.3 as these metrics are customary, cf. Section 2.1, and provide units roughly translatable to monetary value.

Computational grid	Leadtime horizon	Root mean square error deviation factor			Mean absolute error deviation factor		
		Horns Rev	Rejsby Hede	Stor-Rotliden	Horns Rev	Rejsby Hede	Stor-Rotliden
GFS 27 km (0.5°)	(12,24] h	–	–	2.0	–	–	2.1
WRF 30 km	(12,24] h	6.3	3.4	2.0	5.6	3.1	2.0
WRF 10 km	(12,24] h	5.7	3.5	2.1	5.4	2.9	2.2
WRF 3.3 km	(12,24] h	5.7	3.4	2.1	5.4	2.5	2.5
WRF 1.1 km	(12,24] h	5.7	3.4	–	5.3	2.5	–
GFS 27 km (0.5°)	(24,36] h	–	–	1.7	–	–	1.6
WRF 30 km	(24,36] h	3.8	2.5	1.7	3.8	2.4	1.6
WRF 10 km	(24,36] h	3.7	2.5	1.7	3.2	2.2	1.8
WRF 3.3 km	(24,36] h	3.6	2.5	1.8	3.2	2.0	1.9
WRF 1.1km	(24,36] h	3.6	2.5	–	3.1	2.1	–
GFS 27 km (0.5°)	(36,48] h	–	–	1.2	–	–	1.8
WRF 30 km	(36,48] h	2.7	2.1	1.2	3.2	2.3	1.2
WRF 10 km	(36,48] h	2.7	2.1	1.2	2.8	2.2	1.3
WRF 3.3 km	(36,48] h	2.6	2.1	1.3	2.7	2.1	1.2
WRF 1.1 km	(36,48] h	2.6	2.1	–	2.7	2.1	–

Table III. Percentagewise deviation from the best performing computational grid (best accuracy defines 100%) in terms of RMSE and MAE, respectively, divided by percentagewise deviation from the best performing computational grid in terms of PCC.

3.1. Forecast performance

The analysis strategy described in Section 2.1 hinges on the implicit assumption that performance of forecasts initialised 12 UTC is comparable to that of 00 UTC forecasts. In order to verify this, the PCC is computed separately for 00 and 12 UTC forecasts and the results are shown in Figure 2, along with performance for the dataset of chronologically augmented 00 and 12 UTC forecasts. Upon splitting the dataset in forecast initialisation times each forecast leadtime group becomes confined to mutually exclusive half-day time periods; namely the 00–12 and 12–24 UTC hours of the day. That is, e.g. forecast leadtime group (24,36] h for 12 UTC forecasts represents only daily UTC hours 12 to 24, never 00 to 12. Under the assumption of no systematic influence of interactions between forecast leadtime group L, time of the day T, and forecast initialisation time I on forecast PCC performance, the analysis of variance model M with no interactions (see e.g. [29])

$$\mathsf{M} = L_i + T_j + I_k + \varepsilon_{ijk} \qquad \begin{cases} i &= 1,2,3 \\ j &= 1,2 \\ k &= 1,2 \end{cases} \tag{4}$$

can provide a rough PCC assessment of the extent to which forecasts initialised 00 and 12 UTC can be used interchangably, as assumed when augmenting these in the dataset underlying all analyses carried out in both parts of this work, cf. Section 2.1. Defining the model baseline$\equiv L_1 + T_1 + I_1$, that is; forecast leadtime group (12,24] h, time of the day midnight to noon UTC, and initialisation time 00 UTC, the results shown in Table IV indicate whether L, T and I have significant impact on forecast performance in terms of the PCC. Note that the PCC contribution from the time of the day T is larger than the Eq. (4) model residual standard deviation for all three respective sites, while for forecast initialisation time I this is only the case for the GFS at the offshore (HR) and coastal (RH) sites, as well as the coarsest WRF model domain at the coastal wind farm RH. Hence, Eq. (4) OLS fits to forecast data derived from NWPs on each of the five computational grid resolutions *cannot* consistently support the assumption that 00 and 12 UTC forecast initialisations are distinguishable when assessed within the framework of Section 2.1.

For the offshore HR and coastal RH sites, Figure 2 exhibits performance decrease relative to the GFS data for all WRF model resolutions; to better understand the modelling challenges underlying the adverse LAM performance bulk metric assessment of Figure 2, three illustrative forecast examples are shown in Figure 3. Figure 3a demonstrates consistent WRF modelled wind speed phase and magnitude correction increase as model resolution improves, while Figure 3b, from a few days later, shows the opposite tendency. However, weather conditions are similar for both of these July datasets in

Wind Energ. 2015; **00**:1–16 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

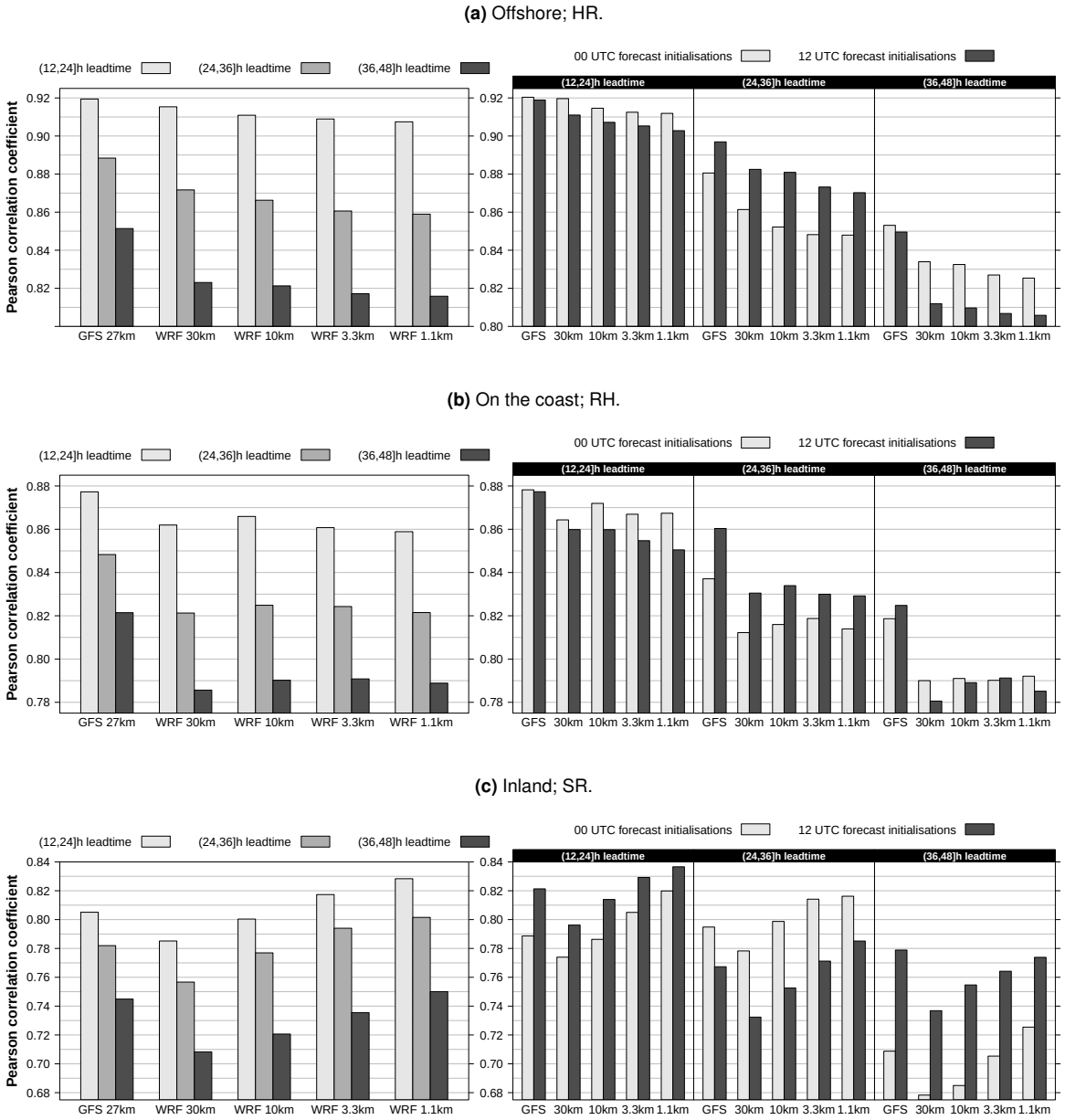


Figure 2. Wind speed forecast performance. **Left**: Assessment according to the three forecast leadtime groups, cf. Section 2.1. **Right**: Separate assessment for 00 and 12 UTC forecast initialisations, cf. Section 3.1.

that the temporal development at both occasions involves abrupt wind direction change from steady south-southeasterly to westerly around $\sim$20–25 hours forecast leadtime. As this study concerns LAM forecast performance assessment based on a yearlong dataset, a more detailed case study focusing on the performance-discrepancy identified for these days in July 2012 is not pursued. For Figure 3c each NWP model resolution increase brings additional wind direction distortion around northerly flow after leadtime hour 30. The high-resolution WRF model grids clearly pick up observed rapidly fluctuating dynamics not reproduced at coarser resolution, albeit slightly off in phase. As namely Figure 3c illustrates, scalar accuracy metrics like the PCC penalise the more detailed representation of atmospheric advection relative to the case for coarser, less volatile forecasts, due to relatively minor phase, amplitude, and temporal dilation errors. In this context, alternative LAM forecast value extraction concerning wind energy scheduling is discussed in Part II of this work [39].

Site	Modelled dataset	Baseline	$H = (24{,}36]$ h	$H = (36{,}48]$ h	$T =$ 12-24 h UTC	$I = 12$ UTC	Standard deviation for the six model residuals
	GFS 27km (0.5°)	0.911	−0.031	−0.068	0.009	0.007	0.001
	WRF 30km	0.905	−0.043	−0.092	0.018	0.003	0.007
HR	WRF 10km	0.897	−0.044	−0.090	0.022	0.007	0.008
	WRF 3.3km	0.897	−0.048	−0.092	0.019	0.006	0.006
	WRF 1.1km	0.896	−0.048	−0.092	0.018	0.004	0.005
	GFS 27km (0.5°)	0.866	−0.029	−0.056	0.010	0.013	0.003
	WRF 30km	0.853	−0.041	−0.077	0.013	0.006	0.002
RH	WRF 10km	0.857	−0.041	−0.076	0.013	0.005	0.005
	WRF 3.3km	0.855	−0.036	−0.070	0.008	0.003	0.007
	WRF 1.1km	0.851	−0.037	−0.070	0.014	0.002	0.005
	GFS 27km (0.5°)	0.819	−0.024	−0.061	−0.039	0.012	0.019
	WRF 30km	0.808	−0.030	−0.077	−0.043	−0.003	0.018
SR	WRF 10km	0.823	−0.024	−0.080	−0.047	0.001	0.021
	WRF 3.3km	0.839	−0.024	−0.082	−0.042	−0.001	0.017
	WRF 1.1km	0.844	−0.028	−0.079	−0.032	0.001	0.016

Table IV. Eq. (4) model fits to PCC data for the three sites, respectively.

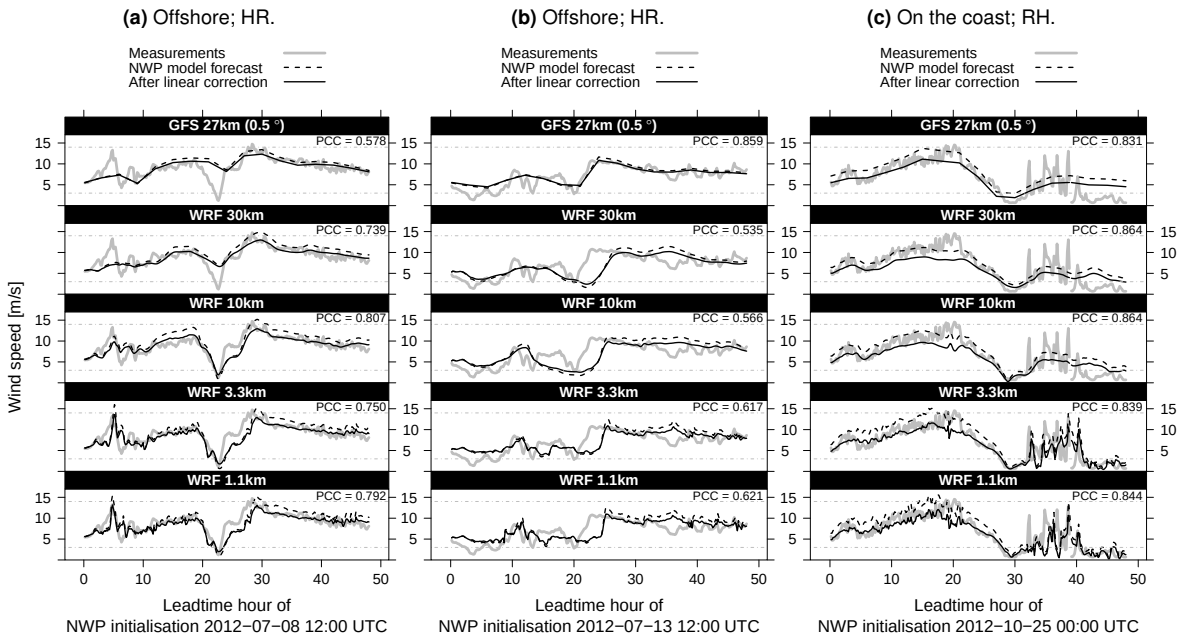


Figure 3. Time series of measured and forecasted wind speed illustrating **(a)** added value from WRF modelling at the HR site, **(b)** exhibits for the same site WRF model introduction of phase error not present in the GFS data, while **(c)** displays weather conditions subject to high-frequency fluctuations at RH. The grey dash-dotted horizontal lines roughly delimit full wind farm production and no production, respectively. Note that the raw 10 minute dataset is shown here for better illustration of rapid atmospheric dynamics represented in the WRF model simulations, rather than the hourly data employed in forecast performance assessment results.

3.2. The impact of temporal smoothing on forecast performance

Scalar accuracy metrics (2) assess the feasibility of NWP model input for wind energy scheduling, and thus GCM and LAM forecasts are typically processed to maximise performance in terms of such metrics prior to downstream decision making. For background literature on statistical modelling techniques for enhancing operational weather forecast value, the reader is referred to e.g. [40] for running-mean bias correction, [41] authors apply a Kalman filter (KF) approach, [42] study the performance of an analog scheme as well as the KF, and e.g. [43] employ MOS. Several studies compare these approaches [44, 24, 45], along with e.g. artificial neural networks modelling [27].

In contrast to the case in the studies just mentioned, the forecast performance enhancement approach employed in this section does not rely on previous data in that the performance impact of evening out slight temporal phase and dilation

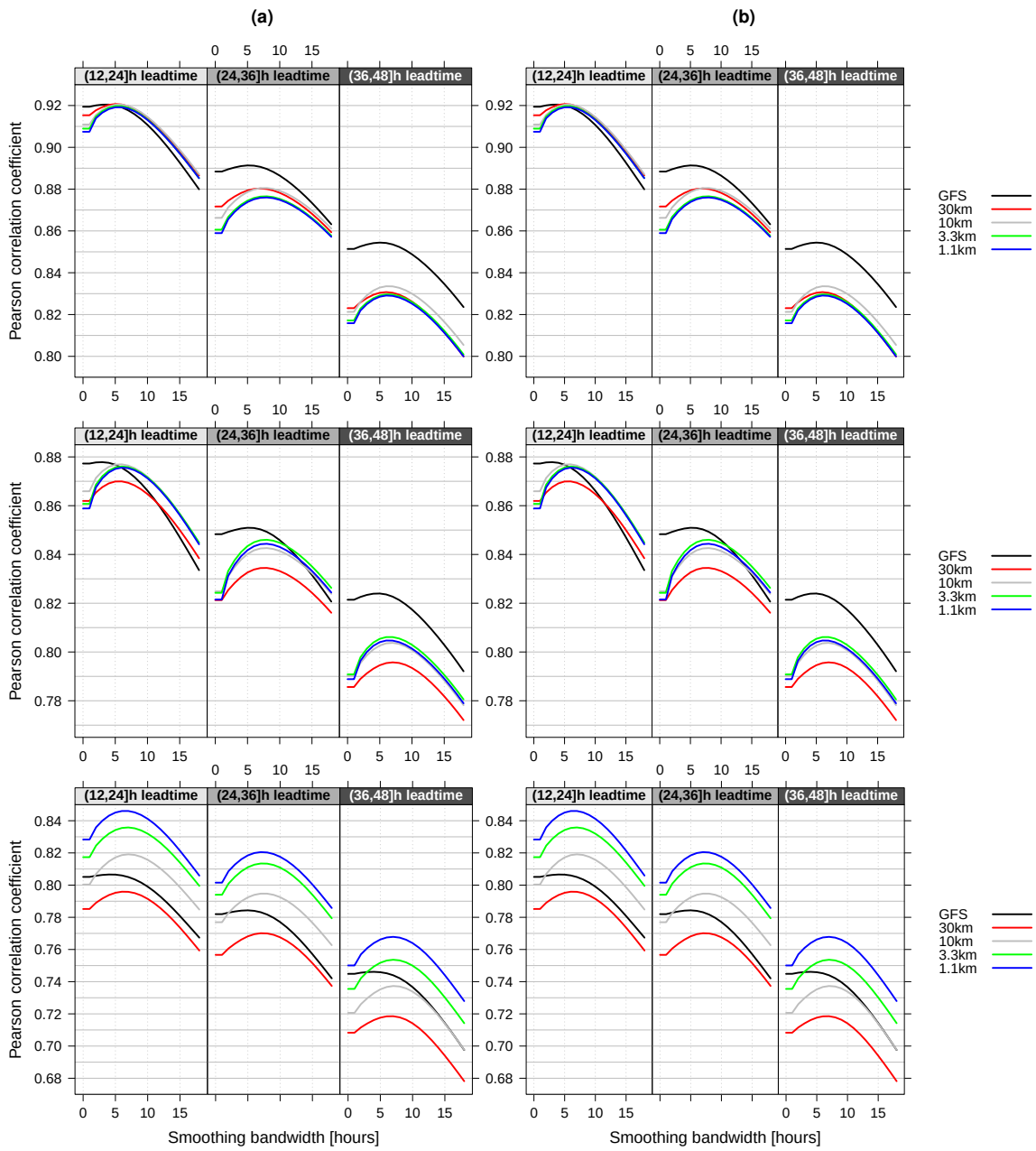


Figure 4. Forecast performance as a function of temporal smoothing bandwidth r according to **(a)** Eq. (3), i.e. zero order local polynomial smoothing, and **(b)** post first order local polynomial smoothing, cf. [33]. **Top row:** HR, **middle row:** RH, **bottom row:** SR. Data prior to leadtime hour 12 is included in the smoothing processes, while data after leadtime hour 48 is not available.

errors is assessed, which involves current forecast data only. This is motivated by the forecast performance results presented in Section 3.1, namely Figure 3, indicating that temporal phase and dilation errors distort value extraction in terms of PCC. Hence, subject to PCC and RMSE accuracy metrics, forecasts that largely capture observed dynamics – though with more or less severe timing errors – may only achieve performance similar to e.g. the naïve benchmark proposed in [46], gradually shifting weight from persistence to climatology as forecast leadtime increases.

The tricube weight function of Section 2.2 is now employed for temporal smoothing of individual wind speed forecasts. Hence, r of Eq. (3) here denotes a temporal smoothing window halfwidth, $\boldsymbol{y}$ is now a scalar y denoting the time stamp to be smoothed, and function argument $\boldsymbol{x}$ is a scalar x covering all time stamps of the forecast at hand. Figure 4 shows that

Site	Computational grid	Root mean square error [%]			Mean absolute error [%]		
		Raw	Smoothed	Improvement	Raw	Smoothed	Improvement
HR	GFS 27km (0.5°)	18.04	17.93	**0.11**	12.89	12.83	**0.06**
	WRF 30km	19.72	19.27	**0.45**	14.19	13.79	**0.40**
	WRF 10km	20.22	19.46	**0.76**	14.58	13.86	**0.72**
	WRF 3.3km	20.49	19.70	**0.79**	14.79	14.08	**0.71**
	WRF 1.1km	20.58	19.78	**0.80**	14.86	14.16	**0.70**
RH	GFS 27km (0.5°)	11.97	11.75	**0.22**	8.42	8.33	**0.09**
	WRF 30km	13.46	12.99	**0.47**	9.62	9.23	**0.39**
	WRF 10km	13.50	12.91	**0.59**	9.57	9.10	**0.47**
	WRF 3.3km	13.63	12.86	**0.77**	9.66	9.10	**0.56**
	WRF 1.1km	13.68	12.86	**0.82**	9.70	9.10	**0.60**
SR	GFS 27km (0.5°)	17.42	17.31	**0.11**	13.25	13.16	**0.09**
	WRF 30km	18.26	17.79	**0.47**	13.97	13.56	**0.41**
	WRF 10km	18.17	17.51	**0.66**	13.92	13.42	**0.50**
	WRF 3.3km	17.94	17.21	**0.73**	13.77	13.19	**0.58**
	WRF 1.1km	17.66	16.88	**0.78**	13.47	12.85	**0.62**

Table V. Capacity-relative wind power forecast [49, 50] performance on the (12,30] h leadtime subject to $r = 7$ hour temporal wind speed forecast smoothing by Eq. (3) for HR (capacity 160 MW), RH (23.4 MW), and SR (77.8 MW) wind farms. The first three months are discarded as spinup in this WPPT setup for GFS/WRF input. N.b. that the standard WPPT setup involves a handfull NWP sources.

optimal improvement from temporal smoothing by Eq. (3) is attainable for $r \sim 7$ hours, as do corresponding results for MAE and RMSE based on LLS wind speed forecast data (not shown). Note that first order local polynomial regression smoothing [33] provides slightly higher peak performance for all WRF model resolutions at horizon (36,48] h for the HR and RH sites, while the converse is true for SR, an effect the authors attribute to different boundary bias implications for the two smoothing approaches (no forecast data beyond 48 h).

Section 2.2 outlines the wind speed dataset used in this article and in [39]. By the same token, corresponding reference datasets for assessment of wind *power* forecast performance are based on 96.2 %, 95.8 %, and 77.7 % complete datasets for HR, RH, and SR wind farms, respectively. The relatively severe dataset reduction for the Swedish site is due to icing events – a current frontline research topic treated e.g. in [47] – reducing the power yield during winter, which in turn complicates power curve estimation [48, 37]; a crucial modelling step in the Wind Power Prediction Tool (WPPT) [49, 50] employed in this section and Section 3.3, as in all wind power forecast models relying on input derived from NWP models. Hence, data points digressing severely from the underlying power curve in wind turbine nacelle anemometer readings vs. power production scatter plots are attributed to icing and removed from the dataset, see [20] for further details on the processing of the wind farm data. As mentioned in the introduction, GFS forecasts are available with a ∼5 hour delay from forecast initialisation time operationally, and the recursive and adaptive WPPT generates forecasts every hour accounting for the latest wind farm measurements in a computationally efficient manner (recursive) and applying recent data with exponentially decaying weight as a function of temporal distance to the current time stamp (adaptive). With the additional lag due to unavailability of NWP model forecasts issued every hour, the upper forecast leadtime consistently evaluatable (no missing data points for each hourly power forecast) is limited to hour 30 ahead of wind power forecast initialisation. Furthermore, the first three months are discarded as WPPT spinup (unreliable output during initial WPPT parameter estimation) and hence the yearly period evaluated for wind speed becomes subyearly for wind power, yet fully covering the most productive part of the year for the sites studied; Northern Hemisphere autumn, winter and spring.

Table V shows the impact of temporal smoothing by Eq. (3) applied to wind speed forecasts (i.e. local zero order polynomial smoothing, cf. Figure 4a) in terms of wind power forecast accuracy. A small improvement is observed for GFS data, available at three-hourly resolution, and for WRF model data resolved as 10 minute output, and averaged to hourly data, the impact almost consistently (except for HR/SR MAE; likely due to the spatial smoothing for high-resolution WRF model grids for these sites, cf. [35, 20]) increases with weather model resolution. This substantiates the idea of severe performance penalty from slight phase errors and temporal dilation, as Figure 3 hints, in otherwise largely correct descriptions of atmospheric dynamics. As noted last in Section 3.1 alternative forecast evaluation w.r.t. wind energy scheduling is carried out in the Part II of this work [39].

Wind Energ. 2015; **00**:1–16 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

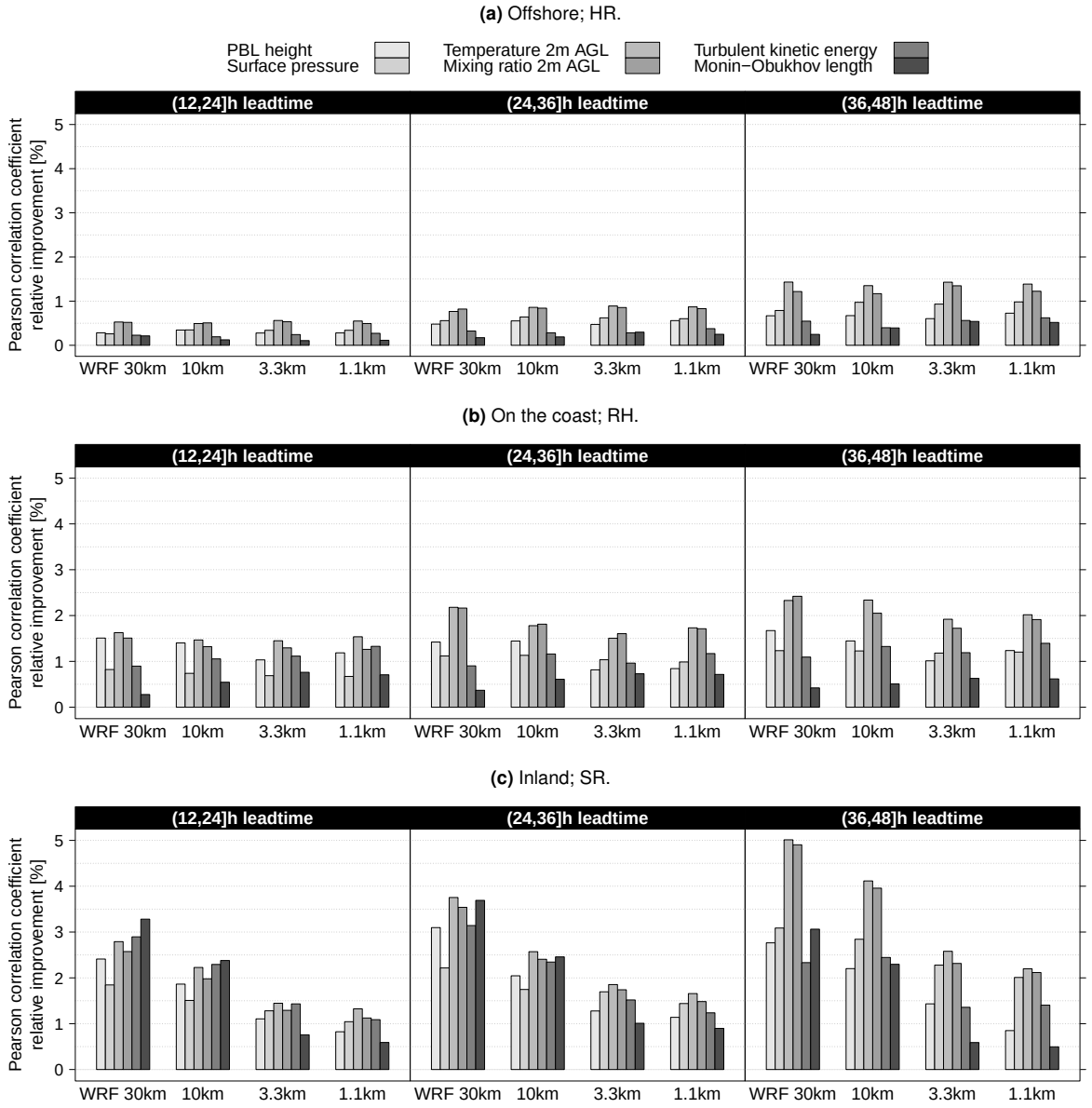


Figure 5. Forecast performance post processing according to Eq. (5) for six z candidates.

3.3. Potential forecast performance increase from conditional parametric modelling

As an extension to the multiple linear regression framework underlying MOS, cf. e.g. the work cited in the brief literature review opening Section 3.2, potential wind speed forecast value from accounting for nonlinear coupling to other LAM variables is here considered. The governing equations for atmospheric dynamics as represented in the WRF model involve complex mathematical relations between the prognostic and diagnostic variables of the discretised NWP formulation, and sub-grid scale parameterised processes such as e.g. atmospheric turbulence; composed of heat-, moisture- and momentum fluxes [51].

In traditional regression modelling for prediction purposes, estimates of intercept and coefficients are inferred from a model training dataset and coefficients such as θ in Eq. (1) are considered unknown constants, whereas conditional parametric modelling frameworks allow for θ dependencies. The conditional parametric analogue to Eq. (1) considered

Site	Comp.	Root mean square error improvement [%]						Mean absolute error improvement [%]					
	grid	PBLH	P	T	M	TKE	MOL	PBLH	P	T	M	TKE	MOL
HR	30 km	0.32	0.73	0.89	0.92	0.39	0.40	0.24	0.77	1.00	1.02	0.42	0.35
	10 km	0.44	0.63	0.90	0.56	0.32	0.24	0.37	0.69	0.91	0.75	0.36	0.18
	3.3 km	0.40	0.65	1.17	0.65	0.40	0.26	0.44	0.65	1.04	0.66	0.47	0.21
	1.1 km	0.52	0.62	0.95	0.66	0.43	0.22	0.64	0.58	0.91	0.74	0.48	0.18
RH	30 km	−1.04	0.69	−0.41	−0.66	−0.96	−1.06	−1.49	−0.15	−1.57	−2.02	−1.50	−2.09
	10 km	−0.62	1.67	−0.37	−0.47	−0.28	−0.66	−0.95	0.68	−1.38	−1.76	−0.83	−1.35
	3.3 km	−0.82	1.98	−0.13	−0.33	−0.37	−0.46	−1.22	0.97	−1.12	−1.49	−0.76	−1.12
	1.1 km	−0.86	1.67	−0.10	−0.48	−0.16	−0.37	−1.38	0.59	−1.20	−1.74	−0.60	−0.99
SR	30 km	0.93	1.24	0.80	0.67	2.13	2.03	0.93	1.20	1.08	1.00	1.94	2.06
	10 km	1.60	0.87	1.53	0.74	2.40	2.22	1.41	0.59	1.58	0.86	2.31	2.17
	3.3 km	1.74	0.80	1.84	1.18	1.94	1.44	1.68	0.75	1.87	1.12	1.92	1.43
	1.1 km	1.52	0.91	1.77	1.05	1.38	1.17	1.50	0.84	1.86	0.89	1.41	1.15

Table VI. Wind power forecast performance improvement from Eq. (5) preprocessed wind speed forecasts for WRF model derived z candidates PBL height (PBLH), surface pressure (P), temperature 2m AGL (T), moisture 2m AGL (M), turbulent kinetic energy (TKE), and Monin-Obukhov length (MOL). As in Table V the $(12,30]$ h leadtime is considered and the first three months of the dataset are discarded as spinup of the WPPT setup for GFS/WRF input. N.b. that the standard WPPT setup involves a handfull NWP sources.

here,

$$\mathsf{W} = x\theta(z) + \varepsilon, \tag{5}$$

involves the additional predictor z, on which θ is assumed to depend smoothly. Based on analytical $\theta(z)$ estimates [52], corresponding to OLS θ estimates for Eq. (1), the extent to which each prognostic or diagnostic LAM variable z couples to wind speed forecast error is gauged by means of local polynomial regression smoothing [33] for construction of a lookup table mapping z to θ. Note that the procedure applied does not allow for a constant intercept, cf. c in Eq. (1), so LLS wind speed is substituted for x rather than raw wind speed forecast data in order to avoid WRF model bias thwarting inference on z predictor value. A broad range of WRF model aspects of potential systematic explanatory value for advection predictability are studied, a brief description of the particular variables considered ensues.

In WRF model version 3.4.1 potential temperature and water vapour mixing ratio are prognostic variables, while Monin-Obukhov length (MOL) and turbulent kinetic energy (TKE) are respectively examples of a surface layer parameterisation diagnostic [53] and an additional prognostic variable introduced via the MYNN PBL parameterisation; cf. Table I and references therein. Potential temperature is nonlinearly coupled to thermodynamic (or absolute) temperature via atmospheric pressure, and the PBL height is a diagnostic of the MYNN scheme.

PCC improvement resulting from separate Eq. (5) fits for each of these z candidates is shown in Figure 5, illustrating systematic coupling to wind speed forecast error in the WRF model configuration employed. Zero, first, and second order local polynomial smoothing yields similar PCC improvement as a function of the nearest neighbour fraction bandwidth (not shown), cf. the alternative fixed-distance bandwidth of Eq. (3). Zero order local polynomial smoothing is employed, and the nearest neighbour fraction producing peak performance underlies results in Figure 5 and Table VI.

Cross validating [54] the modelling approach, e.g. according to the special case of reserving one year of data for model training (in-sample) and using data for the subsequent year for prediction (i.e. out-of-sample evaluation), is nontrivial since all possible z values must be covered during $\theta(z)$ fitting, although realtime remapping to z lookup table boundary values could provide a technical fix. However, with a relatively short single-year dataset and due to support found in the literature [55] pro an in-sample evaluation strategy for assessment of linear regression model predictive ability, inference on z value for wind speed forecast processing by Eq. (5) is based on training and testing on the same, full-year dataset, which entails that results shown in Figure 5 and Table VI reflect the upper bound on attainable performance increase for Eq. (5) assessment based on this particular dataset.

Note, in Figure 5, that the relative improvement from statistical modelling by Eq. (5) generally increases with forecast leadtime, and that prognostic WRF model variables appear to contribute more corrective value than do variables derived

Wind Energ. 2015; **00**:1–16 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

from sub-grid scale parameterisations. Inland, at the SR site, the relative improvement is larger for coarser WRF model resolutions. A curious discrepancy is observed for the coastal site, RH, in that Figure 5 suggests an Eq. (5) enhancement factor in between HR and SR, while all z candidates – except pressure – deteriorate forecast performance in terms of wind power, cf. Table VI. A key challenge in the WPPT, as in other statistical wind power forecast models [37], is to model accurately the relation between atmospheric flow and wind turbine generation. Figure 5 and Table VI together indicate that non-pressure z predictors in Eq. (5) result in inflated scatter around the RH wind farm power curve, thus impeding accurate power curve modelling which is most sensitive in the steep region for $\sim$7–10 m/s wind speeds (non-summer months), while the adverse wind speed spread is dominated by the explanatory effect for the in-sample assessment of wind speed forecast performance in Figure 5 (full year dataset).

4. CONCLUSION

For the offshore and coastal wind farms studied, the LAM configuration used is found to deteriorate forecast performance relative to that of the GCM providing initialisation and LBCs. In contrast, consistent LAM forecast improvement is found for the wind farm inland for each consequetive spatio-temporal resolution increase, superior to the forcing GCM data at the highest LAM resolutions. However, this picture is not preserved when assessing performance in terms wind *power* subject to operational conditions. Though, with temporal smoothing of forecasts, performance inland is found to leverage scalar accuracy metrics for the high-resolution LAM data beyond GCM performance, while LAM performance remains inferior to that of the GCM offshore and on the coast. Note, however, that proper comparative GCM/LAM performance assessment should be subject to an additional LAM forecast leadtime lag, depending on the wall-clock time attainable for the particular limited area forecasting system in question. Finally, conditional parametric modelling of nonlinear coupling of simulated atmospheric variables to the wind speed forecast error indicates that the prognostic variables considered – temperature and moisture – are most promising for offshore wind power forecast enhancement, while only surface pressure modelling has non-deteriorating impact for the coastal site. Inland, TKE and MOL appear more promising as predictors w.r.t. semi-parametric MOS modelling for wind energy scheduling.

GCM data represents the atmospheric fluid spanning the globe, and simulated meteorological fields are plausibly off by a few km in any geographical location now and then, as further substantiated e.g. in Part II of this work [39]. To improve upon LAM predictability by scalar accuracy metrics of interest in wind energy scheduling, studies of the extent to which realtime assimilation of atmospheric measurements to initialise forecasts for the target region should be carried out, as doing this would serve to nudge GCM forcing data towards the spatio-temporal trajectories observed locally, when translated to LAM grids.

ACKNOWLEDGEMENTS

The research presented here is funded by the Danish Transmission System Operator *Energinet.dk*, within the framework of public service obligation project *PSO 10464*. Wind farm datasets were kindly provided by *Vattenfall*, an international energy company primarily based in Sweden. Figure 1 is created using the NCAR Command Language [56], the data analysis is performed using R [57].

REFERENCES

1. Jensen PH, Knudsen S. *Wind energy - drivers and barriers for higher shares of wind in the global power generation mix*. DTU International Energy Report, Technical University of Denmark, 2014; Danish and European plans for wind

energy deployment.

2. Madsen PH, Hansen MH. *Wind energy - drivers and barriers for higher shares of wind in the global power generation mix*. DTU International Energy Report, Technical University of Denmark, 2014; Wind energy technology developments.
3. Deppe, A J and Gallus Jr, W A and Takle, E S. A WRF ensemble for improved wind speed forecasts at turbine height. *Weather and Forecasting* 2013; **28**:212–228.
4. Lazić, L and Pejanović, G and Živković, M. Wind forecasts for wind power generation using the Eta model. *Renewable Energy* 2010; **35**:1236–1243.
5. Mesinger F, Chou SC, Gomes JL, Jovic D, Bastos P, Bustamante JF, Lazic L, Lyra AA, Morelli S, Ristic I, *et al.*. An upgraded version of the Eta model. *Meteorology and Atmospheric Physics* 2012; **116**:63–79.
6. Rife, D L and Davis, C A and Liu, Y and Warner, T T. Predictability of low-level winds by mesoscale meteorological models. *Monthly Weather Review* 2004; **132**:2553–2569.
7. Kanamitsu M. Description of the NMC global data assimilation and forecast system. *Weather and Forecasting* 1989; **4**:335–342.
8. McQueen, J T and Draxler, R R and Rolph, G D. Influence of grid size and terrain resolution on wind-field predictions from an operational mesoscale model. *Journal of Applied Meteorology* 1995; **34**:2166–2181.
9. Pielke RA, Cotton WR, Walko RL, Tremback CJ, Lyons WA, Grasso LD, Nicholls ME, Moran MD, Wesley DA, Lee TJ, *et al.*. A comprehensive meteorological modeling system – RAMS. *Meteorology and Atmospheric Physics* 1992; **49**:69–91.
10. Skamarock WC, Klemp JB. A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal of Computational Physics* 2008; **227**:3465–3485.
11. Nakanishi M. Improvement of the Mellor-Yamada turbulence closure model based on large-eddy simulation data. *Boundary-Layer Meteorology* 2001; **99**:349–378.
12. Kitamura Y. Modifications to the Mellor-Yamada-Nakanishi-Niino (MYNN) model for the stable stratification case. *Journal of the Meteorological Society of Japan* 2010; **88**:857–864.
13. Tewari M, Chen F, Wang W, Dudhia J, LeMone MA, Mitchell K, Ek M, Gayno G, Wegiel J, Cuenca RH. Implementation and verification of the unified noah land surface model in the wrf model. *Proceedings of the 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, Seattle (WA), USA, January*, 2004.
14. Thompson G, Field PR, Rasmussen RM, Hall WD. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review* 2008; **136**:5096–5115.
15. Dudhia J. Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale twodimensional model. *Journal of Atmospheric Science* 1989; **46**:3077–3107.
16. Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlatedk model for the longwave. *Journal of Geophysical Research – Atmospheres* 1997; **102**:16 663–16 682.
17. Kain JS. The Kain-Fritsch convective parameterization: An update. *Journal of Applied Meteorology* 2004; **43**:170–181.
18. Warner TT, Peterson RA, Treadon RE. A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bulletin of the American Meteorological Society* 1997; **78**:2599–2617.
19. Davies HC. A lateral formulation for multi-level prediction models. *Quarterly Journal of the Royal Meteorological Society* 1976; **102**:405–418.
20. Rosgaard MH. Limited area forecasting and statistical modelling for wind energy scheduling. PhD Thesis, Technical University of Denmark 2015.

21. Orths AG, Eriksen PB. *Wind Power in Power Systems*, chap. Wind Power in the Danish Power System. 2 edn., John Wiley & Sons, Ltd., 2012; 536.
22. Jollife IT, Stephenson DP. *Forecast Verification: A Practitioner's Guide in Atmospheric Science, 2nd edition*. Wiley, 2011.
23. Wilks D. *Statistical Methods in the Atmospheric Sciences, 3rd edition*. Academic Press/Elsevier, 2011.
24. Müller MD. Effects of model resolution and statistical postprocessing on shelter temperature and wind forecasts. *Journal of Applied Meteorology and Climatology* 2011; **50**:1627–1636.
25. Pinson P. Adaptive calibration of (u,v)-wind ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society* 2013; **138**:1273–1284.
26. Madsen H, Pinson P, Kariniotakis G, Nielsen HA, Nielsen TS. Standardizing the performance evaluation of short-term wind power prediction models. *Wind Engineering* 2005; **29**:475–489.
27. Sweeney C, Lynch P, Nolan P. Reducing errors of wind speed forecasts by an optimal combination of post-processing methods. *Meteorological Applications* 2013; **20**:32–40.
28. Murphy AS. Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review* 1988; **116**:2417–2424.
29. Olsson U. *Generalized Linear Models. An Applied Approach*, chap. General Linear Models. Studentlitteratur, 2001.
30. Jonsson B. Prediction with a linear regression model and errors in a regressor. *International Journal of Forecasting* 1994; **10**:549–555.
31. Jónsson, T and Pinson, P and Nielsen, H Aa and Madsen, H. Exponential smoothing approaches for prediction in real-time electricity markets. *Energies* 2014; **7**:3710–3732.
32. Birkhoff G. Tricubic polynomial interpolation. *Proceedings of the National Academy of Sciences of the United States of America*, 1971.
33. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 1979; **74**:829–836.
34. Lekien F, Marsden J. Tricubic interpolation in three dimensions. *International Journal for Numerical Methods in Engineering* 2005; **63**:455–471.
35. Rosgaard MH, Hahmann AN, Nielsen TS, Giebel G, Sørensen PE, Madsen H. Evaluation of dynamical downscaling resolution effect on wind energy forecast value for a wind farm in central sweden. *Proceedings of the 3rd International Lund Regional-Scale Climate Modelling Workshop: 21st Century Challenges in Regional Climate Modelling, Lund, Sweden, 16-19 June*, 2014.
36. Monin AS. The atmosperic boundary layer. *Annual Review of Fluid Mechanics* 1970; **2**:225–250.
37. Lydia M, Kumar SS, Selvakumar AI, Kumar GEP. A comprehensive review on wind turbine power curve modeling techniques. *Renewable and Sustainable Energy Reviews* 2014; **30**:452–460.
38. Glahn HR, Lowry DA. The use of Model Output Statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology* 1972; **11**:1203–1211.
39. Rosgaard MH, Hahmann AN, Nielsen HA, Giebel G, Sørensen PE, Skov TS. Limited area forecasting for wind energy scheduling. Part II: Predictability of wind speed variability and uncertainty. *Wind Energy* 2015; .
40. Stensrud DJ, Yussouf N. Bias-corrected short-range ensemble forecasts of near surface variables. *Meteorological Applications* 2005; **12**:217–230.
41. Cassola F, Burlando M. Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output. *Applied Energy* 2012; **99**:154–166.
42. Monache LD, Nipen T, Liu Y, Roux G, Stull R. Kalman filter and analog schemes to postprocess numerical weather predictions. *Monthly Weather Review* 2011; **139**:3554–3570.
43. Hart KA, Steenburgh WJ, Onton DJ, Siffert AJ. An evaluation of mesoscale-model-based Model Output Statistics (MOS) during the 2002 Olympic and Paralympic Winter Games. *Weather and Forecasting* 2004; **19**:200–218.

44. Cheng WYY, Steenburgh WJ. Strengths and weaknesses of MOS, running-mean bias removal, and Kalman filter techniques for improving model forecasts over the western United States. *Weather and Forecasting* 2007; **22**:1304–1318.
45. Sweeney C, Lynch P. Adaptive post-processing of short-term wind forecasts for energy applications. *Wind Energy* 2011; **14**:317–325.
46. Nielsen, T S and Joensen, A K and Madsen, H. A new reference for wind power forecasting. *Wind Energy* 1998; **1**:29–34.
47. Davis N, Hahmann AN, Clausen NE, Žagar M. Forecast of icing events at a wind farm in Sweden. *Journal of Applied Meteorology* 2014; **53**:262–281.
48. Messner JW, Zeileis A, Broecker J, Mayr GJ. Probabilistic wind power forecasts with an inverse power curve transformation and censored regression. *Wind Energy* 2013; **17**:1753–1766.
49. Nielsen TS, Nielsen HA, Madsen H. Prediction of wind power using time-varying coefficient-functions. *Proceedings of the 15th Ifac World Congress on Automatic Control, Barcelona, Spain, 2126 July*, 2002.
50. Nielsen HA, Nielsen TS, Madsen H. An overview of wind power forecast types and their use in large-scale integration of wind power. *Proceedings of the 10th International Workshop on Large-Scale Integration of Wind Power into Power Systems, Aarhus, Denmark, 25-26 October*, 2011.
51. Hartmann DL. *Global Physical Climatology*, chap. Global Climate Models. Academic Press, 1994; 257–264.
52. Nielsen HA. *LFLM version 1.0, an S-PLUS/R library for locally weighted fitting of linear models.* Technical Report 22. Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark. 1997.
53. Jiménez, PA and Dudhia, and González-Rouco, JF and Navarro, J and Montávez, JP and García-Bustamante, E. A revised scheme for the WRF surface layer formulation. *Monthly Weather Review* 2012; **140**:898–918.
54. Efron B, Gong G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *American Statistician* 1983; **37**:36–48.
55. Roecker, E B. Prediction error and its estimation for subset-selected models. *Technometrics* 1991; **33**:459–468.
56. The NCAR Command Language. Version 6.2.0. Boulder, Colorado: UCAR/NCAR/CISL/VETS. URL http://dx.doi.org/10.5065/D6WD3XH5.
57. R Core Team. R: A language and environment for statistical computing. Version 3.0.2. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

DOI: 10.1002/we
Prepared using weauth.cls

Appendix B

Journal publication II — *draft*

WIND ENERGY
Wind Energ. 2015; **00**:1–18
DOI: 10.1002/we

RESEARCH ARTICLE

Limited area forecasting for wind energy scheduling. Part II: Predictability of wind speed variability and uncertainty

Martin H. Rosgaard[1,2], Andrea N. Hahmann[1], Henrik Aa. Nielsen[2], Gregor Giebel[1], Poul E. Sørensen[1], Torben S. Nielsen[2]

[1]DTU Wind Energy, Technical University of Denmark, Risø Campus, DK-4000 Roskilde, Denmark.
[2]ENFOR A/S, DK-2970 Hørsholm, Denmark.

ABSTRACT

Aspects of the simulated wind speed fluctuations are assessed for different model resolutions in terms of spectra, phase, atmospheric stability, ramp events, and predictability of uncertainty, based on a year's twice-daily 48 hour forecasts for three wind farm sites offshore, coastal and inland. Forecast time series are generated using the Weather Research and Forecasting (WRF) model forced by Global Forecast System (GFS) data. Modelled fluctuation spectra are found to approach spectra based on measurements as WRF model resolution increases, yet it is also shown that wind speed forecasts lag behind measurements as a function of model integration time. In terms of the Critical Success Index ramp event forecast performance improves with WRF model resolution, and wind speed uncertainty is shown to relate near-linearly to the logarithm of forecasted wind speed variability offshore and on the coast, albeit not inland. Copyright © 2015 John Wiley & Sons, Ltd.

KEYWORDS

numerical weather prediction; ramp forecasting; uncertainty predictor; wind speed fluctuation; phase error; atmospheric stability

Correspondence

Martin Haubjerg Rosgaard, Technical University of Denmark, Department of Wind Energy. Email: mhros@dtu.dk

Contract/grant sponsor
Danish transmission system operator *Energinet.dk* public service obligation

Contract/grant number
PSO 10464

Received . . .

Atmospheric Circulation Effects on Wind Speed Variability at Turbine Height

1. INTRODUCTION

The feasibility of wind energy as the power supply backbone of the future hinges on day ahead predictability of this renewable resource [1]. Point forecasts for wind energy scheduling concern power predictions based on numerical weather

†Please ensure that you use the most up to date class file, available from the WE Home Page at
http://www3.interscience.wiley.com/journal/6276/home

prediction (NWP) data, and knowledge of the associated forecast uncertainty constitute valuable input for efficient trading of green energy on spot markets [2], modelling of necessary power reserves for the electrical grid [3], and optimal power auction bidding strategies [4]. Extensive reviews of different strategies for associating point forecasts of wind power with uncertainty estimates can be found in [5, 6, 7], involving e.g. the concepts of an uncertainty risk index [8], probabilistic prediction [9], and wind power production scenarios [10]. In this work, the potential of forecasted wind speed variability as a wind power uncertainty predictor is established, and the performance of high-resolution limited area forecasting of wind speed found quantified in [11] in terms of scalar accuracy metrics is readdressed in terms of wind speed ramp object evaluation — another forecasting aspect of interest for wind energy scheduling. In the past, the adverse impacts of such events were less important, though with the present day large average wind farm capacities onshore, and offshore in particular, the wind power feed into the electrical grid can suddenly increase or decrease by several GW [7].

The analysis strategy outlined in Section 3 starts out with a framework for preliminary probing of NWP model deficiency, before discussing wind speed ramp objects and forecast uncertainty in more detail. Results are presented in Section 4, followed by the conclusion in Section 5.

2. DATA

The dataset studied is described in detail in [11], a brief outline ensues. Two configurations of the Weather Research and Forecasting (WRF) model [12] focus on 1) two Danish wind farms; Horns Rev (HR) and Rejsby Hede (RH), and 2) a wind farm in central Sweden; Stor-Rotliden (SR), and forecasts are validated against wind farm averaged turbine anemometer readings from the three sites, respectively. Aside from slightly differing numbers of horizontal grid points and geographical regions covered, the two WRF model configurations are identical. The separate configurations have four computational grids, each with 40 vertical levels covering up to 50 mb altitude ($\sim$20 km above ground level), propagating dynamics one-way from the outermost domain at 30 km horizontal resolution to a finer grid of 10 km, onwards to a grid of 3.3 km, to the innermost grid at 1.1 km horizontal resolution. As the horizontal resolution, the governing equations are integrated with a time step that increases by a factor of three for each WRF model nesting level; from 2 minutes on 30 km grids down to $\sim$4 s for the 1.1 km grid. 48 hour WRF model forecasts for 00 and 12 UTC are initialised daily from operational Global Forecast System (GFS) [13] forecasts, and the two 30 km computational domains are forced laterally every three hours with GFS forecast data. WRF model output is saved at 10 minute temporal resolution.

The WRF model configuration employed in this study parameterises sub-grid scale planetary boundary layer processes according to the Mellor-Yamada-Nakanishi-Niino (MYNN) level 2.5 local closure scheme [14, 15], which extends the governing equations of the WRF model core [12] with an additional prognostic equation describing the evolution of the turbulent kinetic energy in the atmosphere. Cf. e.g. [16] for a concise overview and application of different closures.

WRF model output analysed in this study is outlined in Table I.

VERTICAL LEVEL	METEOROLOGICAL VARIABLE
For each vertical WRF model layer	Wind speed
For each vertical WRF model layer	Wind speed standard deviation based on data points at each integration step between output times
Characterising vertical profile of surface layer flow	Monin-Obukhov Length derived from the MYNN surface layer parameterisation

Table I. Subset of WRF model variables extracted every 10th minute.

3. PRACTICAL ASPECTS OF WIND SPEED VARIATION — AN ANALYSIS STRATEGY

Variability and uncertainty can be viewed as two aspects of variation [17]. Throughout this work,

Wind Energ. 2015; **00**:1–18 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

- wind speed *variability* is taken to refer to spectral, dichotomous, and standard deviation measures of temporal fluctuations in forecast and observation data separately, whereas
- wind speed *uncertainty* refers to the extent to which forecasts are accurate, as quantified by the standard deviation of the forecast error.

Sections 3.1 through 3.3 concern variability only and extends the forecast performance assessment based on scalar metrics in [11] by evaluation in terms of spectra and a meteorological object type of interest for wind energy scheduling; ramp events. Last, Section 3.4 addresses whether simulated variability can predict forecast uncertainty, cf. Section 1.

3.1. Decomposition of temporal variability in terms of spectral properties

Power spectral density for each forecast is estimated in order to probe the extent to which fluctuation frequency and amplitude of simulated wind speed time series matches observations. Such discrete spectra are termed periodograms [18], here denoted $\mathcal{P}$, and the evaluation in the present context relies on interpretation of wind speed datasets for individual forecasts, with temporal resolution t spanning the time period T covered by leadtime hours 12 to 48, as truncated (infinite) stationary signals (i.e. periodicity is assumed), each with $N = \frac{T}{t}$ wind speed data points. For each of these datasets the unitary discrete Fourier transform (DFT),

$$\mathfrak{F}(\omega_k) = N^{-\frac{1}{2}} \sum_{n=0}^{N-1} x_n \exp\left(-\mathrm{i}\omega_k t_n\right) \quad \text{with} \quad \omega_k = \frac{2\pi k}{Nt}, \quad t_n = tn, \quad \text{and} \quad k = 0, 1, \ldots, N-1 \tag{1}$$

is evaluated for each spectral data point counter (i.e. fluctuation frequency index) k. That is, the DFT quantifies the similarity of the wind speed time series x_n to sinusoid waves with angular frequencies ω_k, respectively, cf. e.g. the underlying mathematics of the sample Pearson correlation coefficient (PCC; r) for measured m_n and forecasted f_n wind speed;

$$r = \frac{\sum_{n=1}^{N} \left(f_i - \overline{f}\right)\left(m_i - \overline{m}\right)}{\sqrt{\sum_{n=1}^{N} \left(f_i - \overline{f}\right)^2} \sqrt{\sum_{n=1}^{N} \left(m_i - \overline{m}\right)^2}}$$

widely employed for forecast performance assessment in [11]. With Eq. (1) evaluation by the fast Fourier transform algorithm [19], periodograms for individual-forecast datasets of observed and forecasted wind speed can be estimated, cf. e.g. [20] (n.b. only a single signal segment is considered in this context);

$$\mathcal{P}(\omega_k) = \left|\mathfrak{F}(\omega_k)\right|^2, \qquad \mathfrak{F}(\omega_k) \in \mathbf{C}^N, \tag{2}$$

and power spectral densities for observed and forecasted wind speed datasets are estimated by averaging individual spectra for complete measurement-forecast dataset pairs. Note that the discrete spectrum resulting from Eq. (2) evaluation is equivalent to the DFT of the autocovariance corresponding to the respective x_n time series.

3.2. Forecast phase offset relative to measurements as a function of weather model integration time

The extent to which the WRF modelled fluctuation timing reflects observed fluctuations is assessed based on PCC calculation for forecasts vs. measurements subject to measurement time stamp displacement by ± 3 hours. The investigation is supplemented with inspection of the modelled boundary layer stability (see e.g. [21]), as quantified in terms of the Monin-Obukhov length [22];

$$L = -\frac{u_*^3 \overline{T_v}}{\kappa g \, \overline{w' T_v'}} \tag{3}$$

with u_* denoting the friction velocity, $\overline{T_v}$ average virtual potential temperature, κ the von Kármán constant, and $\overline{w'T_v'}$ the surface virtual potential temperature flux; i.e. the covariance of T_v with vertical velocity w, the upper single quote $'$ denoting perturbations from the mean quantities $\overline{T}_v$ and $\overline{w} = 0$, respectively. The classification employed is displayed in Table II

Monin-Obukhov length	Classication of the boundary layer
$0 < L < 500$	stable
$\lvert L \rvert > 500$	neutral
$-500 < L < 0$	unstable

Table II. Classification of atmospheric stability.

Roughly put, atmospheric stability pertains to the balance between downwards mechanically hurled turbulent flow due to friction induced by topography-dependent surface drag on horizontal advection, and upwards directed turbulent flow forced by thermal buoyancy as daytime heating of the ground is re-emitted as infrared radiation. When the former prevails over the latter the atmospheric boundary layer is said to be stable and surface processes have little effect on wind speeds aloft, while dominating thermal buoyancy characterises unstable conditions during which laminar flow aloft is disrupted to a greater extent. In neutral conditions the two vertical forcings approximately cancel such that an air parcel at a certain height will by and large remain at said height as a function of time. Hence, any dependence of simulated L on WRF model integration time may provide additional insight.

3.3. Predictability of ramp type fluctuations with implications for wind energy scheduling

In this work *ramp* refers to a temporally steady wind speed increase (up-ramp) or decrease (down-ramp), occuring within a time span of up to eight hours duration. Ramps traversing parts of the steepest region of typical wind farm power curves (i.e. ∼5–12 m/s) pose a critical challenge for transmission system operators (TSOs) and stakeholders alike, and day ahead knowledge of ramp events constitutes valuable input for wind energy scheduling and hence supports power grid stability [7]. A study related to this work [23], likewise employing the WRF ARW core [12] for wind energy scheduling applications, indicates that the MYNN 2.5 order local closure parameterisation is among the PBL schemes best capable of representing fast ramp events as observed in practice; a contributing factor to the selection of this scheme in the present study.

Based on ideas and empirical parameter choices put forth in previous work, a ramp detection algorithm is here developed and evaluated for the dataset at hand, in order to extend with object-based verification the scalar accuracy metric assessment of WRF model forecast value presented in [11]. Features of the approach employed are highlighted as bullet points. One study [24] restricts ramp event duration to 2 hours and find no consistent superiority of the computationally higher resolved MM5 model (the WRF model predecessor) forecasts in detecting ramps accurately in their hourly dataset. Upon relaxing the requirement of exact temporal overlap the high-resolution forecasts' performance improved for both meridional and zonal wind speed components.

- In the context of this work, observed and forecasted ramps must occur within a margin of a few hours apart as prior knowledge of ramps much off in time of occurence are of little use in operational wind energy scheduling.

Furthermore, the study [24] employs separate MM5 model runs for performance comparison of different resolutions.

- Resolution comparison here is based on WRF configurations with one-way interacting nested computational domains, cf. Section 2.

This ensures that the exact same GFS forcing is applied at the lateral boundaries of the outermost domains of the WRF model configurations. Another study [25] defines a wind power ramp object as a change of more than 50 % of full wind farm capacity within 4 hours, while a third study [26] specifies a change of 75 % within a 3 hours time span.

Wind Energ. 2015; **00**:1–18 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

- The aim here is not to evaluate ramp forecast performance for particular end-user requirements, e.g. by strictly confining attention to ramps of a singly-defined duration. Rather, a more broad insight is pursued by validating ramp object simulation across several spatio-temporal WRF model resolutions for a range of ramp durations.

Further sophistication in the form of having the ramp wind speed threshold scale with the event duration, as in the approaches taken e.g. in [27, 28], would entail selection of a suitable scaling parameter and is therefore avoided. The fixed change of 50 % wind farm capacity used e.g. in [25, 29] is adopted here as this appears to be a common threshold choice. Based on power curves derived from historical nacelle measurements of wind speed and power production for the three wind farms studied here,

- a ramp threshold of 5 m/s is employed, which roughly corresponds to a power change of at least 50 % of full wind farm capacity within the finite-slope power curve regime.

Where [24] used the wind speed difference between end-points of sliding time-windows 2 hours wide to identify ramp objects, an alternative approach is employed here due to the following technical implications. For a sinus curve signal $\sin(\omega)$ spanning $\omega \in [0, \pi]$, a window width differing from $\frac{\pi}{2}$ would detect no ramp objects of threshold 1, while for window width $\frac{\pi}{2}$ an up-ramp and a down-ramp would be identified for $\sin(\omega)$, provided the time series has a data point exactly at $\frac{\pi}{2}$. In theory, the precise ramp-duration filtering of the end-points approach employed in [24] is an unambiguous advantage, though in this study of relatively high (10 minute) temporal resolution fluctuations near the ramp beginning and end may lead to identification of multiple ramp objects closely spaced in time, in practice representing the same event. Which temporal distance should be used to infer whether ramp events occuring temporally close are in fact the same event? Given that durations of such forecasted ramp objects do not temporally match exactly, which one applies in the event of only a single observed corresponding ramp? Should an observed ramp object of 2 hours duration be considered missed by a forecasted ramp occuring at overlapping time stamps, yet of 3 hours duration? In order to simplify the complexity entailing the end-points approach, an idea implemented e.g. in [27] is pursued here;

- rather than letting the sliding ramp-screening window end-points define ramp bounds, the difference between maximum and minimum wind speed values instead defines the ramp magnitude and is compared to the threshold defined to determine whether it is an event of interest.

This choice of ramp definition consequently relaxes the precise ramp-duration filtering exercised in the end-points approach to inclusion of all ramps of duration up till the specified time window width. However, the issues discussed in the previous paragraph vanish, in that multiple forecasted ramps for a single observed event can no longer occur and a forecasted ramp of 3 hours duration covering the temporal span of an observed ramp of only 2 hours duration is now recorded as a correct forecast, in contrast to the case subject to the end-points classification algorithm. Despite these theoretical concerns, a study evaluating the end-points and maximum/minimum methods, albeit with ramp threshold scaling with event duration, arrived at similar results for the two ramp definitions applied to a dataset of 5-minute wind power time series [30]. That said, many authors do seem to prefer the definition involving detection of a maximum/minimum above the end-points definition for wind energy ramp research [25, 31, 26, 27].

One way to determine whether an observed ramp event is forecasted is to assign a single time stamp to the observed ramp and any forecasted ramps, respectively, and determine the temporal distance(s), Δt, between forecasted events and the observed event. Hence, an empirical choice must be made with regard to the magnitude of Δt. Furthermore, if the ramp durations of observed and forecasted ramps are different the long-duration object median time stamp may be far from that of the short duration object, despite overlapping time stamps prior to single-time stamp assignment. Alternatively, using the intersection of observed and forecasted ramp event time stamps spanning each object avoids the need for an empirical Δt choice as well as ramp duration dependencies. Hence,

- an observed ramp is assessed correctly forecasted if the time stamps constituting the object reccur in a forecasted ramp object, provided observed and forecasted ramps are both either up- or down-ramps.

The forecast assessment in terms of ramp object data is summarised as dichotomous tri-category counts according to the procedure outlined in Table IV, cf. Table III.

	OBSERVED	NOT OBSERVED
FORECASTED	Hits	False alarms
NOT FORECASTED	Misses	Correct misses (**n/a**)

Table III. Contingency table of dichotomous categories employed for ramp object forecast performance assessment.

The results are further assessed in terms of Eqs. (4-6).

$$\text{Accuracy} = \frac{\text{Hits}}{\text{Hits} + \text{False alarms}} \tag{4}$$

$$\text{Capture} = \frac{\text{Hits}}{\text{Hits} + \text{Misses}} \tag{5}$$

$$\begin{array}{c}\text{Critical Success Index}\\ (a.k.a.\ \text{Threat Score})\end{array} = \frac{\text{Hits}}{\text{Hits} + \text{False alarms} + \text{Misses}} \tag{6}$$

Data: *Wind speed dataset @10 minute resolution:*

- Wind farm average of turbine anemometer readings
- WRF model forecasts @ horizontal resolutions 30 km, 10 km, 3.3 km, and 1.1 km

Result: *Quantification of ramp object hits, misses, and false alarms, cf. Table III*

---

function A: IDENTIFY RAMP TIME STAMPS(input time series, screening-window width, threshold=5 m/s)
Assess for each data point, ± the nearest-neighbour data points within and including a temporal distance of half the screening-window width, whether the absolute value of maximum minus minimum wind speed exceeds the specified threshold, distinguishing up- and down-ramps with separate labels.
return time stamps for which ramps are identified
end function
function B: COLLAPSE OBSERVED COHERENT RAMP TIME STAMPS TO THEIR MEDIAN(**function** A output)
Time stamps spanning observed ramps can extend across the temporal forecast limits at leadtime hours 12 and 48. Hence, this function is used in **function** C to determine whether the forecast in question contains the majority of an observed ramp object; i.e. whether the **function** A output should be assigned occurence within the leadtime hour span 12 to 48 of said forecast. If not, the observed ramp object is disregarded when the forecast in question is assessed. **function** B is also applied in **function** C during ramp object classification as hit, miss or false alarm, cf. Table III.
return assignment of a single time stamp to coherent ramp time stamps
end function
function C: ASSIGN DICHOTOMOUS CATEGORIES FOR RAMP OBJECT DATASETS(**function** A output)
Based on **function** A output for the full-year wind speed measurements sliced to match the 36-hour (leadtime hours 12 to 48) time span covered by **function** A output for an individual-forecast dataset, determine whether an observed ramp object shares time stamps with a forecasted object (hit) or not (miss), and whether each forecasted ramp object has no time stamp in common with any observed objects (false alarm), cf. Table III. Once this has been determined, **function** B is applied to assign a single time stamp to the event based on the corresponding observed ramp time stamps (hits and misses) or forecasted ramp time stamps (false alarms).
return output from **function** A condensed to a dataset of dichotomously categorised events
end function
loop over a range of ramp object screening-window widths spanning 1 hour to 8 hours
...
end loop
function D: POSTPROCESS THE RESULTING DATASET(output from **loop** over **function** C)
Count the number of ramp object hits, misses, and false alarms and evaluate the corresponding accuracy, capture, and threat scores, cf. Eqs. (4-6).
return output from **function** A condensed to a dataset of dichotomously categorised events
end function

Table IV. Algorithm for assessment of wind speed ramp object predictability

Wind Energ. 2015; **00**:1–18 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

Note that the nomenclature for these quantities varies in the literature; e.g. the terminology for Eqs. (4-5) is also employed in [25], while [32] denotes these quantities *precision* and *recall*, respectively. More sophisticated skill scores focus e.g. on separating events from non-events, like the Hanssen and Kuiper skill score [33], while the Equitable Threat Skill Score [34] used e.g. in [24]; a study related to this work, extents Eq. (6) to account for all four Table III categories (n.b. [35]). However, such scores are not applicable here as there is no reasonable way to define a non-event, i.e. the lower right element in Table III, within the presented framework.

3.4. Quantification of uncertainty predictability in terms of forecasted variability

Wind power prediction uncertainty modelling adds additional sophistication to the decision-maker toolbox for trading wind energy on markets such as the day ahead Elspot auction [36], facilitating operational power trading across borders in several northern European countries. Wind speed forecast uncertainty is here defined as the standard deviation of the forecast error, cf. the introductory paragraph of Section 3. The corresponding *sample* standard deviation based on forecasted and measured time series reads

$$s_\epsilon = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})}, \qquad \text{with} \qquad \epsilon_i = m_i - f_i \quad \text{and} \quad \bar{\epsilon} = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i, \tag{7}$$

forecast error ϵ representing deviation of forecasted (f) from measured (m) wind speed, and n denoting the number of time series data points, in contrast to the case in Section 3.1 where n was a data point index (to avoid i or j confusion with $\sqrt{-1}$). Deviation between measurements and model estimates of the same continuous variable can reasonably be assumed Gaussian and hence ϵ is taken to represent forecast error samples from a normal population $N\left(\mu_\varepsilon,\sigma_\varepsilon^2\right)$. With the additional assumption of ϵ data point independency, it can be shown that the quantity $\frac{(n-1)s_\epsilon^2}{\sigma_\varepsilon^2}$ follows a χ^2 distribution with $n-1$ degrees of freedom. By the traditional frequentist approach to estimating confidence intervals [37], the $(1-\alpha)100\%$ confidence interval for s_ϵ can be estimated by

$$s_\epsilon\sqrt{\frac{n-1}{\chi^2_{\alpha/2}}} < \sigma_\varepsilon < s_\epsilon\sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}}}, \tag{8}$$

see e.g. [38]. The variance of a χ^2 distributed variable with k degrees of freedom is $2k$, hence

$$\mathrm{Var}\left[\frac{(n-1)s_\epsilon^2}{\sigma_\varepsilon^2}\right] = 2(n-1)$$
$$\Updownarrow$$
$$\mathrm{Var}\left[s_\epsilon^2\right] = \frac{2\sigma_\varepsilon^4}{n-1},$$

i.e. the variance of s_ϵ^2 is inversely proportional to the number of data points n, which implies that the confidence interval becomes narrower as n increases. Figure 1 illustrates the rate with which the s_ϵ factor $\sqrt{\frac{n-1}{\chi_x^2}}$ of the (8) limits approach 1 asymptotically as a function of n from below $\left(x = \frac{\alpha}{2}\right)$ and above $\left(x = 1 - \frac{\alpha}{2}\right)$, and shows that the 95% confidence band for uncertainty estimates by Eq. (7) for individual forecasts of only 216 data points (the dashed white line; corresponds to 10-minute resolution for leadtime hours 12 to 48) is 10–20 times wider in terms of $\sqrt{\frac{n-1}{\chi_x^2}}$ than if a full year of 10-minute data points (solid white line) are used to assess s_ϵ. This, in turn, roughly implies that uncertainty estimation of individual forecasts by Eq. (7) are 10–20 times more noisy than estimates based on a full year's data, and hence any underlying uncertainty coupling to a potential predictor will be much less thwarted when Eq. (7) estimates are based upon a long dataset. In addition, the data point autocorrelation is not negligible as assumed in (8) and followingly the *effective* number of data points (and thus degrees of freedom) is smaller than n. Therefore the confidence bandwidth factors $\sqrt{\frac{n-1}{\chi_x^2}}$ for $n = 216$ and $n = 52560$ are in fact wider than sketched in Figure 1.

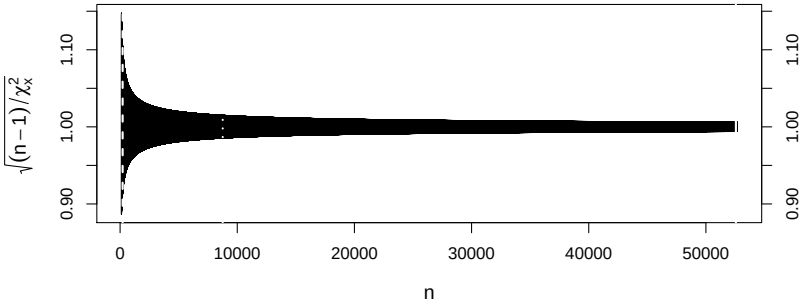


Figure 1. The confidence interval (8) bandwidth in terms of the s_ϵ factor $\sqrt{\frac{n-1}{\chi^2_x}}$ as a function of the number of data points n, for a 95% ($\alpha = 0.05$) confidence band, i.e. within (8) interval limits $\chi^2_x = \chi^2_{0.025}$ and $\chi^2_x = \chi^2_{0.975}$. Data point count $n = 216$ is marked with a white dashed line, $n = 24 \cdot 365 = 8760$ is marked with a dotted line, and $n = 6 \cdot 24 \cdot 365 = 52560$ is marked with a solid white line.

Intuitively, the temporal volatility of NWP modelled wind speed plausibly relates systematically to wind speed uncertainty in terms of Eq. (7). One metric pursuing this notion is the 'meteo-risk index' (MRI) proposed in [8], quantifying current wind power prediction uncertainty based on a weighted sum of the population standard deviation of the deviation between a subset of the current wind speed forecast and corresponding time stamp values from previous wind speed forecasts. Although demonstrated to be a powerful tool [8], the reliability of the MRI arguably depends on the frequency of NWP model initialisations available and the forecast subset-interval of interest. In this work, the feasibility of an alternative approach is investigated, relying only on current forecast data and providing uncertainty predictions at point forecast temporal resolution. As hinted in the Section title, the uncertainty predictor examined is the WRF modelled wind speed variability. This method does, however, involve customisation of the NWP model source code such that wind speed variability is added as a model output variable (not a part of standard output), that is; the sample standard deviation of wind speed values at WRF model time integration resolution is evaluated at each 10 minute output time. The algorithm first proposed in [39] is employed;

$$\mathcal{M}_i = \mathcal{M}_{i-1} + \frac{w_i - \mathcal{M}_{i-1}}{i}, \qquad \mathcal{S}_i = \mathcal{S}_{i-1} + (w_i - \mathcal{M}_{i-1})(w_i - \mathcal{M}_i), \qquad i = 2, 3, \ldots, n,$$

with $\mathcal{M}_i$ denoting running-mean of forecasted wind speed w_i for the ith data point, with $\mathcal{M}_1 = w_1$ and $\mathcal{S}_1 = 0$. At each 10 minute output time the standard deviation is then estimated as

$$s_w = \sqrt{\frac{\mathcal{S}_i}{i-1}}.$$

The integration time step employed for each of the computational grids must coincide with the 10 minute output time, increasing with a factor of three, as do the spatial resolution, with each nesting level. Denoting the outer domain (30 km) $D = 0$, the first nest (10 km) $D = 1$, the second nest (3.3 km) $D = 2$ and the innermost nest (1.1 km) $D = 3$; the number of data points constituting 10 minute forecast variability is given by

$$n = 5 \cdot 3^D. \tag{9}$$

As Eq. (9) indicates, integration of the WRF discretisation of the governing equations is performed at 2 minute temporal resolution for the outermost computational domain, cf. Section 2, corresponding to 5 data points over a 10 minute period, etc. by Eq. (9) for the nested domains.

For wind energy scheduling purposes, the hourly-mean wind speed and power is typically of interest, cf. e.g. the discussion on the Elspot market for day ahead energy trading in [40]. In order to expand the 10-minutely variability forecasts to hourly values, observe that for N groups with n data points (of variable w) in each group, the total sum of

Wind Energ. 2015; **00**:1–18 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

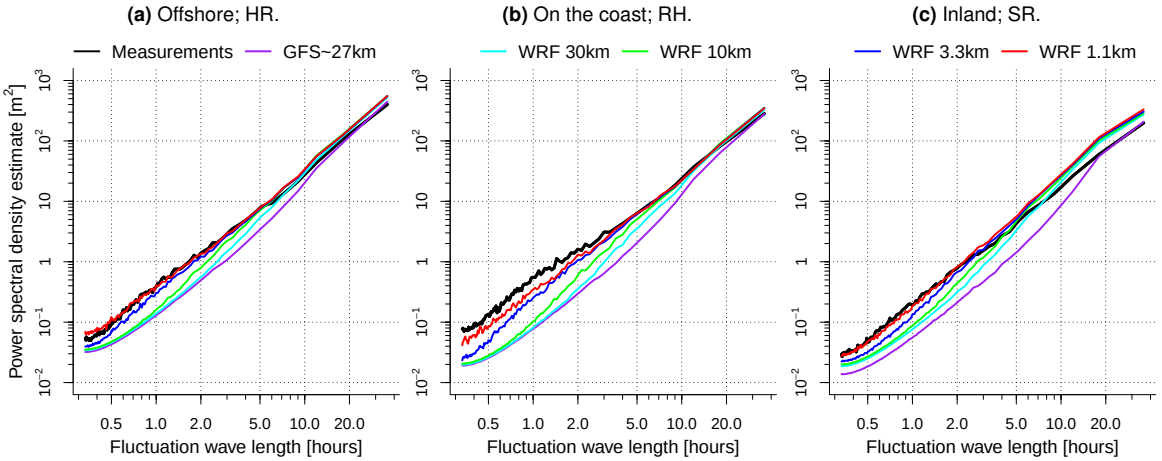


Figure 2. Wind speed periodograms for three wind farms based on 580, 475, and 531 forecasts, respectively. Out of the total 719 forecasts no time stamp in these forecast datasets contains any missing wind speed measurements, and for each of the NWP computational domains the respective 580, 475 and 531 individual spectra are averaged. Forecast leadtime hours 12 to 48 are considered, hence the longest wave length of 36 hours corresponding to the fundamental fluctuation frequency.

squares;

$$\sum_{i=1}^{N}\sum_{j=1}^{n}(w_{ij}-\overline{w})^2 = \sum_{i=1}^{N}\sum_{j=1}^{n}(w_{ij}-\overline{w_i}+\overline{w_i}-\overline{w})^2,$$
$$= \left[\sum_{i=1}^{N}\sum_{j=1}^{n}(w_{ij}-\overline{w_i})^2\right] + n\left[\sum_{i=1}^{N}(\overline{w_i}-\overline{w})^2\right] + 2\left[\sum_{i=1}^{N}\underbrace{\sum_{j=1}^{n}(w_{ij}-\overline{w_i})}_{=0}(\overline{w_i}-\overline{w})\right].$$

Hence, the sample standard deviation of wind speed time series blocks, w_{ij}, at WRF model integration time step resolution, spanned by N consequetive groups each with the same number of data points (n), is

$$S_w = \sqrt{\frac{\sum_{i=1}^{N}(n-1)s_{w,i}^2 + n(\overline{w_i}-\overline{w})^2}{nN-1}} \tag{10}$$

when expressed in terms of the corresponding group mean $\bar{w}$, as well as the mean $\bar{w}_i$ and standard deviations s_i for individual i groups, each of which containing n data points. That is, to obtain *hourly* wind speed variability (S_w) from 10 minute wind speed (w_i) and variability ($s_{w,i}$) forecast time series with data point index i, Eq. (10) is evaluated with $N = 6$.

4. RESULTS

In this section, the analysis strategy outlined in Section 3 is applied to the dataset described in Section 2.

4.1. Variability spectra and phase errors

Figure 2 shows the estimated power spectral density spectra for each wind farm site, cf. Section 3.1. N.b. that approximately one third of the forecasts spanning the study period are excluded due to missing measurements and 11 unavailable GFS forecasts, though inspection of the included forecasts verify that the yearlong dataset period is covered fairly evenly for all three sites (not shown). Hence, the Figure 2 indication that wind speed forecast spectra monotonously approach

corresponding observations for all three sites suggests that the WRF 1.1 km runs do indeed reproduce the observed spectral intensity of wind speed, in spite of the – cf. [11] – inferior performance in terms of scalar accuracy metrics. In general, the WRF modelled amplitude of low frequency modes is overpredicted for all wind farm sites. Performance in terms of PCC is evaluated for the three augmented horizon groups described and investigated in [11]. In Figure 3 optimal WRF model performance appears to be shifting right as the leadtime increases, indicating that atmospheric dynamics as simulated in this particular configuration of the WRF model is gradually lagging behind observed dynamics as time integration proceeds away from forecast initialisation time. The lag appears to be introduced in the coarsest WRF model computational domain and is then gradually corrected as resolution increases; see e.g. WRF 30 km and 1.1 km for horizon group (12,24] h in Fig. 3a, or the same resolutions for horizon group (24,36] h in Fig. 3c. Note in Table V that the atmospheric conditions migrate consistently from neutral to unstable for HR forecasts as a function of leadtime, while for the land-based sites the trend is rather from stable to neutral. Held together with Figure 3, this could indicate that thermally driven buoyancy gradually dominates downward momentum flux as the WRF model simulations progress (increased stability), either due to slowing advection as Figure 3 seems to suggest or due to thermal flux overprediction. However, the WRF modelled wind speed bias (overprediction) steadily increases by 0.08–0.15 m/s (not shown; SR is in between extremes $\sim$0.08 m/s for RH and $\sim$0.15 m/s for HR) for each horizon group into the future, contradicting advection slowdown and hence the gradually increasing phase lag in Figure 3 and transition towards unstable conditions in Table V is instead partly attributed to erroneous vertical fluxes due to overheating of the ground that accumulates with each time integration step in the WRF model simulations.

4.2. Predictability of rapid changes in atmospheric flow intensity

The algorithm of Table IV is applied to the dataset outlined in Section 2 in order to assess the WRF model configurations' ability to reproduce a class of meteorological events of interest for wind energy scheduling; ramp objects.

Table VI shows the number of ramps observed vs. screening window-width. In a wind energy scheduling context, ramp events spanning up to $\sim$3–4 hours are usually considered [26, 25], though longer durations are included here in order to illustrate a drawback of the approach due to merging events. Note, in Table VI, e.g. event counts for ramp durations 6 and 7 hours for HR ramp events in the Autumn; the decrease of 3 up-ramp events when expanding the screening window-width from 6 to 7 hours cannot be explained solely by conversion to down-ramp events in that only 2 additional down-ramp events are counted when going from 6 to 7 hours max. ramp duration. Rather, the issue is that same-sign events closely spaced in time will overlap for wide enough durations such that the Table IV procedure cannot distinguish the events and consequently assign these a single time stamp. This ramp-merging issue is arguably less pronounced for shorter ramp durations of 1–4 hours, as is commonly studied in the wind energy research literature; cf. e.g. [26, 25, 30, 28, 27].

Figure 4 demonstrates that false alarms, as a function of WRF model resolution, outweight the count of correctly forecasted ramp objects for the Danish sites, and hence the accuracy consistently worsens as WRF model resolution increases. However, for SR the accuracy for each computational grid is more similar, though still with slightly inferior

Site	Leadtime	Data point percentage **unstable** [%]				Data point percentage **neutral** [%]				Data point percentage **stable** [%]			
	horizon [hours]	30km	10km	3.3km	1.1km	30km	10km	3.3km	1.1km	30km	10km	3.3km	1.1km
	(12,24]	63.0	63.9	64.5	64.6	14.0	15.2	14.6	14.3	23.0	20.9	20.9	21.2
HR	(24,36]	64.5	65.5	65.5	65.8	14.1	14.6	14.3	14.1	21.4	19.9	20.2	20.1
	(36,48]	66.0	66.2	67.1	67.2	12.5	13.7	13.1	13.1	21.5	20.1	19.8	19.8
	(12,24]	28.3	25.5	24.2	25.6	31.7	35.7	40.7	35.7	39.9	38.8	35.1	38.7
RH	(24,36]	29.0	26.1	25.8	27.5	33.6	38.1	41.9	37.2	37.5	35.8	32.3	35.3
	(36,48]	27.9	25.7	25.2	26.6	35.6	38.8	43.4	38.5	36.5	35.6	31.4	34.9
	(12,24]	27.3	25.9	21.7	19.8	41.8	43.1	47.0	52.4	30.9	31.0	31.3	27.8
SR	(24,36]	27.9	27.4	22.5	20.1	43.2	44.3	47.6	54.3	28.9	28.2	29.9	25.6
	(36,48]	26.9	27.3	22.8	20.0	44.7	45.0	48.8	55.0	28.5	27.6	28.4	25.0

Table V. Data point percentage for which the WRF modelled Monin-Obukhov length indicates unstable, neutral, and stable weather conditions, respectively, cf. Eq. (3) and Table II.

Wind Energ. 2015; **00**:1–18 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

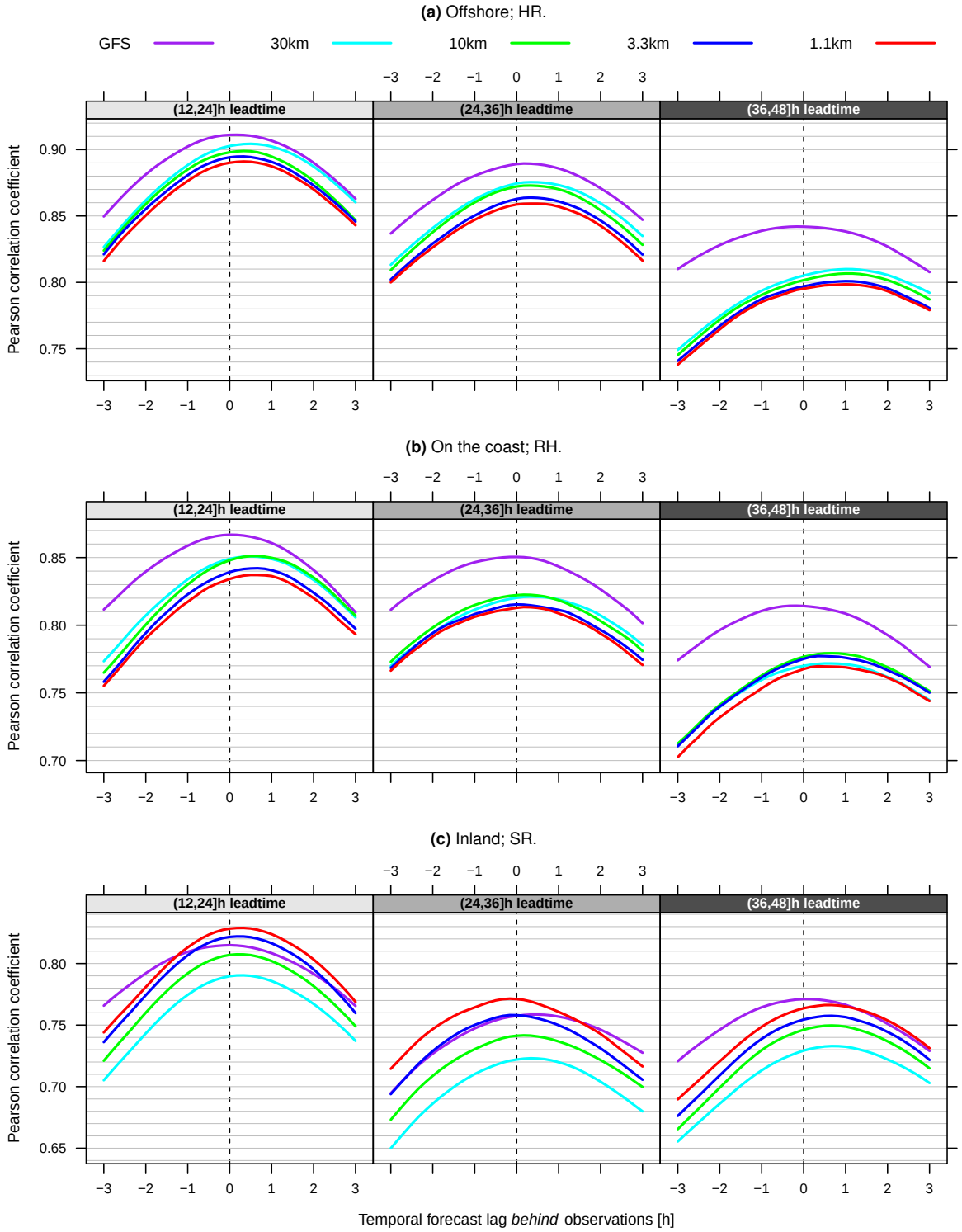


Figure 3. Forecast performance in terms of PCC vs. temporal displacement of wind farm measurements for the three sites.

performance of the finest resolutions. On the other hand, note that high-resolution forecasts better capture observed ramps for all sites. Figure 5 summarises ramp object forecast performance and shows that the WRF model configurations

Site	Maximum duration	Down-ramps					Up-ramps				
		Spring	Summer	Autumn	Winter	**In total**	Spring	Summer	Autumn	Winter	**In total**
HR	1 hour	2	12	33	16	**63**	1	6	33	10	**50**
	2 hours	9	15	55	26	**105**	9	17	58	20	**104**
	3 hours	21	22	70	33	**146**	24	36	72	43	**175**
	4 hours	28	31	77	36	**172**	35	51	76	54	**216**
	5 hours	37	39	76	45	**197**	38	57	85	58	**238**
	6 hours	41	45	88	57	**231**	43	60	92	55	**250**
	7 hours	55	45	90	58	**248**	59	63	89	54	**265**
	8 hours	59	52	91	54	**256**	63	56	82	59	**260**
RH	1 hour	3	16	36	9	**64**	3	19	41	8	**71**
	2 hours	8	20	59	23	**110**	10	32	69	24	**135**
	3 hours	19	32	77	40	**168**	19	40	70	34	**163**
	4 hours	23	35	85	41	**184**	27	45	87	46	**205**
	5 hours	36	47	88	48	**219**	41	52	80	53	**226**
	6 hours	40	65	83	55	**243**	47	59	84	61	**251**
	7 hours	46	65	74	57	**242**	52	57	74	61	**244**
	8 hours	49	67	66	56	**238**	54	64	71	69	**258**
SR	1 hour	1	3	0	0	**4**	4	4	3	0	**11**
	2 hours	3	7	5	2	**7**	6	8	6	5	**25**
	3 hours	12	11	6	7	**36**	11	11	12	10	**44**
	4 hours	15	15	12	10	**52**	18	18	31	18	**85**
	5 hours	22	22	18	19	**81**	29	18	29	23	**99**
	6 hours	22	27	27	20	**96**	32	25	29	28	**114**
	7 hours	28	27	27	23	**105**	32	28	33	30	**123**
	8 hours	31	28	32	28	**119**	31	30	37	33	**131**

Table VI. Number of ramp events identified from application of **function**s A and B of Table IV to 10 minute wind farm data as a function of the four seasons.

consistently improve predictions with each WRF model nesting level, when assessed in terms of the Critical Success Index Eq. (6).

Upon restricting (i.e. increasing) the wind speed change threshold for forecasted ramp objects, while keeping the threshold for observed ramps fixed at 5 m/s, generally reduces the number of false alarms more effectively than the number hits, such that accuracy by Eq. (4) improves for high-resolution forecasts in particular (not shown). On the other hand, the capture ratio Eq. (5) decreases for all WRF model resolutions as observed ramps are now missed to a greater extent. By the same token, temporal smoothing cf. [11] dramatically reduces the number of false alarms relative to the number of hits, again improving accuracy and reducing capture slightly (not shown). Finally, artificially extending the ramp time stamps identified by **function** A of Table IV by a extended tolerance of e.g. 30 minutes or a few hours at the start and end of coherent ramp time stamps consistently improves all three performance metrics; Eqs. (4–6) (not shown). The three suggested methods can be used to tune Algorithm IV towards the criteria of interest to the ramp forecast end-user assessing ramp forecast performance.

4.3. Forecasted variability as an uncertainty predictor

Figure 6 hints the extent to which there is an underlying near-linear relation between the logarithm of WRF modelled wind speed variability and wind speed uncertainty expressed by Eq. (7). Notice how the fairly well-separated dependency traces for each horizon group becomes more "noisy" when the raw 10 minute data is transformed to hourly in Figure 7, by Eq. (10) and hour-averaged ϵ input for Eq. (7). Keep in mind that this corresponds to one sixth (8760) of the 10 minute dataset n (52560), cf. Figure 1, and hence the s_ϵ evaluation within each 10-percentile WRF model variability bin, based on n equal to one tenth of 8760/52560 (hourly/10-min. data), becomes less confident accordingly. The ϵ mean is fairly constant across the 10-percentile bins (not shown), which in turn enables fair comparison of s_ϵ estimates between bins. That is; the $\bar{\epsilon}_{\text{bin}}$ offset for ϵ deviation within each bin is comparable between bins.

Overall, the offshore site HR exhibits the strongest relation between forecasted wind speed variability and forecast uncertainty, while the sought-after connection is more questionable for the inland SR site. For the high-variability

Wind Energ. 2015; **00**:1–18 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

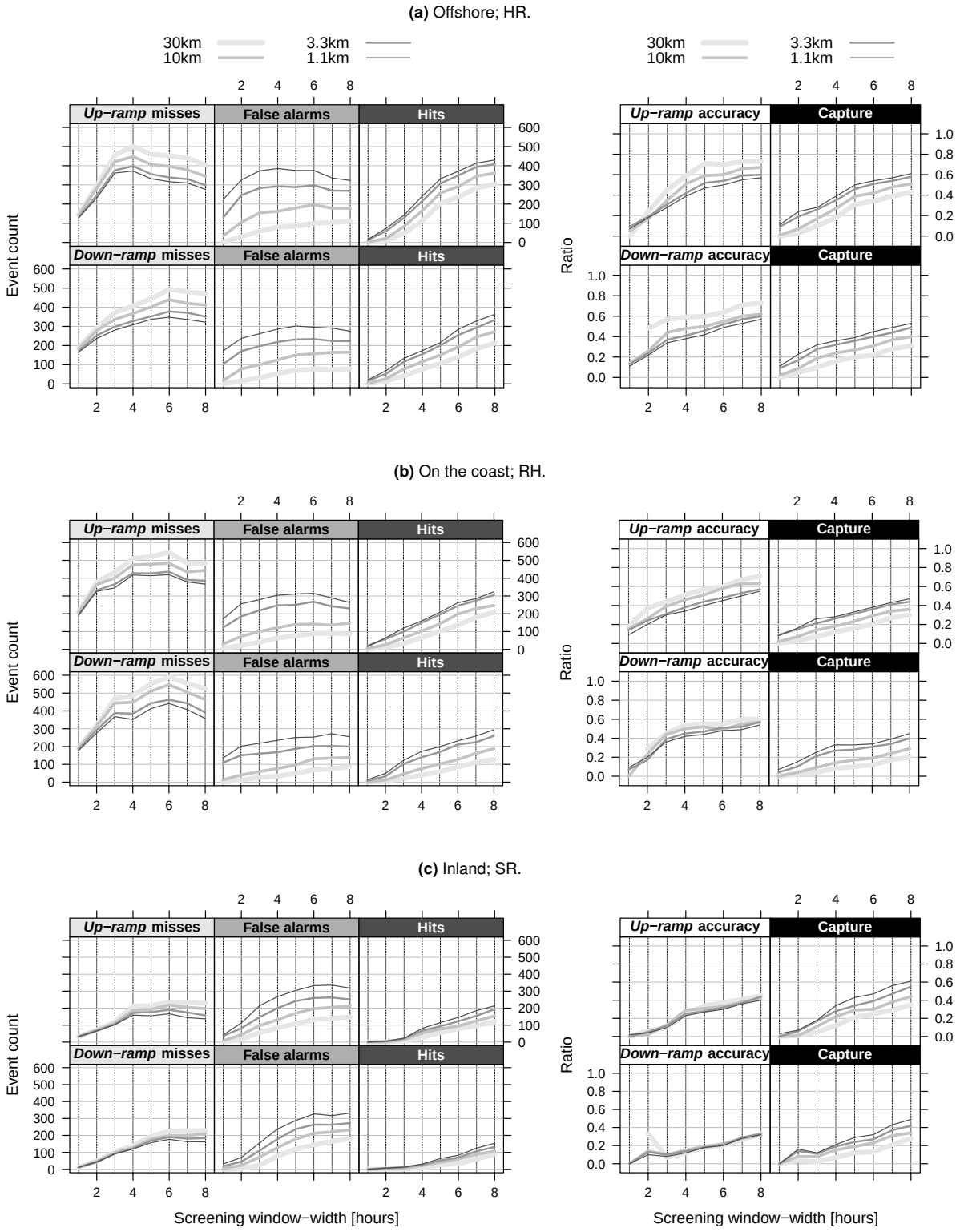


Figure 4. Count of wind speed ramp object forecast misses, false alarms and correct forecasts, along with derived forecast accuracy and capture ratios; Eqs. (4-5).

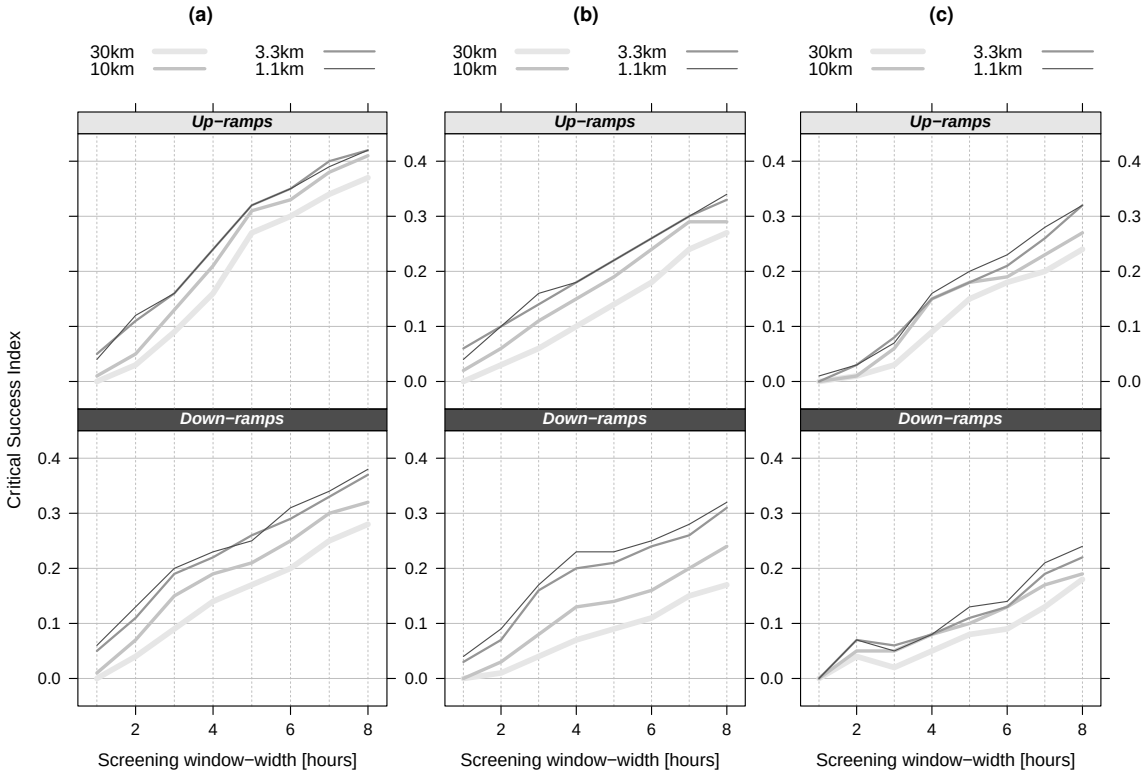


Figure 5. The Critical Success Index, Eq. (6), for wind speed ramps at **(a)** HR, **(b)** RH, **(c)** SR.

5. CONCLUSION

The work presented here supplements the traditional bulk metric performance assessment and enhancement studies for limited area NWP forecasts based on the WRF model [12], as carried out in [11] for four different WRF model resolutions, by alternative quantification of forecast value in terms of wind speed ramp objects and predictability of wind speed uncertainty by forecasted variability. Subject to simplifying assumptions discussed in detail, the simulated wind speed fluctuation intensity within the day ahead leadtimes typically of interest for energy trading [36]; hours 12–48 beyond forecast initialisation time, is found to gradually approach the spectra for observed fluctuations as WRF model resolution increases. However, temporally offsetting measured wind speed relative to corresponding forecast data shows that atmospheric dynamics represented by the WRF model configuration employed in this work tends to temporally fall behind observed dynamics. Inspection of simulated atmospheric stability as a function of forecast leadtime reveals a gradual shift from stable towards unstable conditions. The GFS data forcing the WRF model forecasts laterally exhibits less severe leadtime-dependent phase lag.

A generic approach for wind speed ramp event detection is developed and WRF model forecasts assessed based on a year of data, rather than limiting attention to a few case studies. Ramp forecast performance is found to improve with WRF model resolution for all three sites studied when quantified in terms of the Critical Success Index. It is noted that higher accuracy can achieved, at the expense of the number of ramp events captured in forecasts, by increasing the wind speed threshold in the ramp definition, or by smoothing wind speed forecasts temporally cf. [11]. Furthermore, the authors remark that all ramp performance metrics can be tuned with artificial temporal dilation of ramp event durations, thus controlling the phase error tolerance.

Finally, wind speed forecast uncertainty is defined and the predictive ability of this quantity by WRF modelled wind speed variability is established, based on yearlong datasets for two wind farm sites offshore and on the coast. However, for the third site in complex terrain inland, the log-linear relation established for – namely the high-resolution forecasts

Wind Energ. 2015; **00**:1–18 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

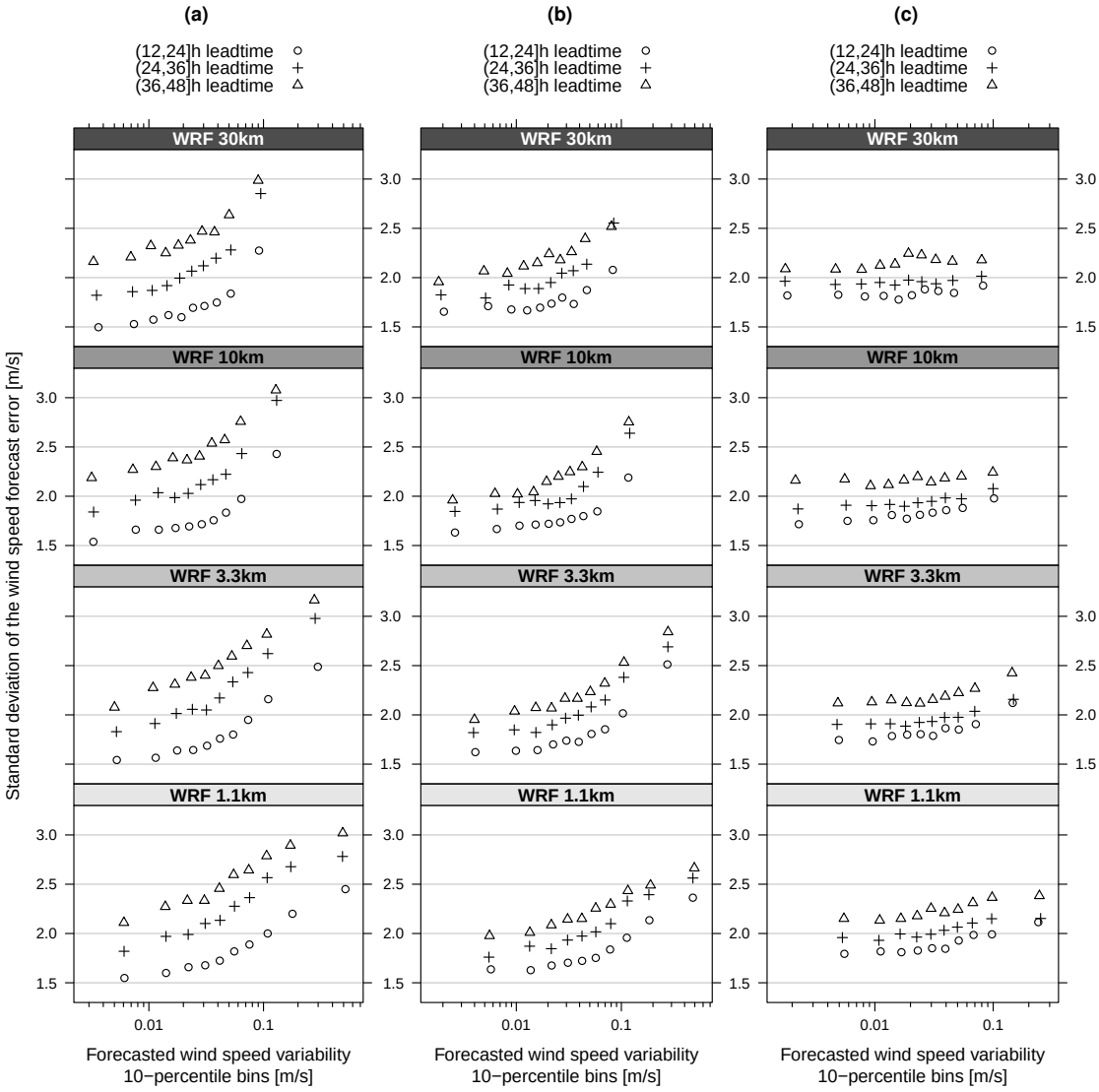


Figure 6. Wind speed uncertainty in terms of Eq. (7) vs. WRF modelled wind speed variability based on the raw 10 minute datasets for **(a)** HR, **(b)** RH, **(c)** SR.

of – the former two sites is not as evident. As such, the uncertainty study constitutes a precursor for future integration of NWP modelled wind speed variability as a predictor in wind power uncertainty forecasting models, and suggests that this endeavour will be most feasible for sites offshore or in flat terrain.

ACKNOWLEDGEMENTS

The research presented here and in [11] is funded by the Danish Transmission System Operator *Energinet.dk*, within the framework of public service obligation project *PSO 10464*. Wind farm datasets were kindly provided by *Vattenfall*, an international energy company primarily based in Sweden. The authors thank Stig Mortensen of ENFOR A/S for helpful comments on the uncertainty study. The data analysis is performed using R [41].

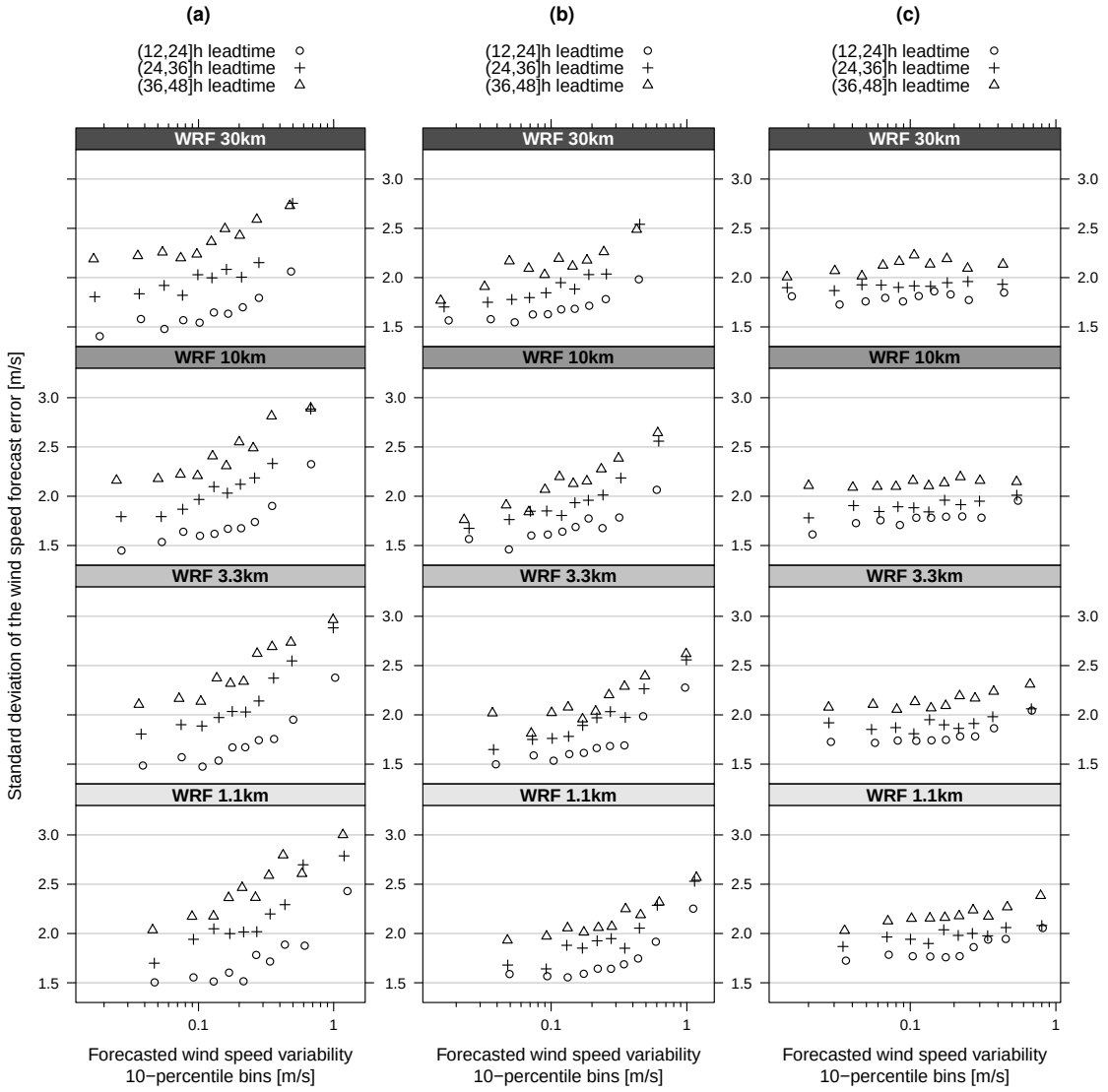


Figure 7. Wind speed uncertainty in terms of Eq. (7), with hourly-averaged ϵ, vs. WRF modelled wind speed variability scaled to hourly resolution, cf. Eq. (10), for **(a)** HR, **(b)** RH, **(c)** SR.

REFERENCES

1. Pinson P. Wind energy: Forecasting challenges for its operational management. *Statistical Science* 2013; **28**:564–585.
2. Usaola JG, Angarita JM. Combining hydro-generation and wind energy. *Electric Power Systems Research* 2007; **77**:393–400.
3. Doherty R, O'Malley M. A new approach to quantify reserve demand in systems with significant installed wind capacity. *IEEE Transactions on Power Systems* 2005; **20**:587–595.
4. Pinson P, Chevallier C, Kariniotakis G. Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Transactions on Power Systems* 2007; **22**:1148–1156.
5. Pinson P, Nielsen HA, Madsen H, Lange M, Kariniotakis G. Methods for the estimation of the uncertainty of wind power forecasts. *Technical Report*, ANEMOS project workpackage 3 deliverable. March 2007.
6. Monteiro C, Bessa R, Miranda V, Botterud A, Wang J, Conzelmann G. Wind power forecasting: State-of-the-art 2009. *Technical Report ANL/DIS-10-1*, Argonne National Laboratory. November 2009.

7. Giebel G, Brownsword R, Kariniotakis G, Denhard M, Draxl C. The state of the art in short-term prediction of wind power: A literature overview, 2nd edition (p. 60). *Technical Report DELIVERABLE D-1.2*, ANEMOS.plus and Safewind projects, sponsered by the European Commission. January 2011.
8. Pinson P, Kariniotakis G. On-line assessment of prediction risk for wind power production forecasts. *Wind Energy* 2004; **7**:119–132.
9. Nielsen HA, Madsen H, Nielsen TS. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind Energy* 2006; **9**:95–108.
10. Pinson P, Madsen H, Nielsen HA, Papaefthymiou G, Klöckl B. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy* 2009; **12**:51–62.
11. Rosgaard MH, Hahmann AN, Skov TS, Giebel G, Nielsen HA. Limited area forecasting for wind energy scheduling. Part I: Performance assessment and enhancement studies. *Wind Energy* 2015; .
12. Skamarock WC, Klemp JB. A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal of Computational Physics* 2008; **227**:3465–3485.
13. Kanamitsu M. Description of the NMC global data assimilation and forecast system. *Weather and Forecasting* 1989; **4**:335–342.
14. Nakanishi M. Improvement of the Mellor-Yamada turbulence closure model based on large-eddy simulation data. *Boundary-Layer Meteorology* 2001; **99**:349–378.
15. Kitamura Y. Modifications to the Mellor-Yamada-Nakanishi-Niino (MYNN) model for the stable stratification case. *Journal of the Meteorological Society of Japan* 2010; **88**:857–864.
16. Alapaty, K and Pleim, J E and Raman, S and Devdutta, S N and Byun, D W. Simulation of atmospheric boundary layer processes using local- and nonlocal-closure schemes. *Journaly of Applied Meteorology* 1997; **36**:214–233.
17. van Belle G. *Statistical Rules of Thumb*, chap. Environmental Studies. second edn., Wiley, 2008; 99–100.
18. Schuster, A. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestial Magnetism* 1898; **3**:13–41.
19. Cooley JW, Tukey JW. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation* 1965; **19**:297–301.
20. Welch, P D. The use of fast Fourier transforms for the estimation of power spectra: A method based on time averaging over short modified periodograms. *IEEE Transactions on Audio and Electroacoustics* 1967; **15**:70–73.
21. Mohan, M and Siddiqui, T A. Analysis of various schemes for the estimation of atmospheric stability classification. *Atmospheric Environment* 1998; **32**:3775–3781.
22. Obukhov, A M. Turbulence in an atmosphere with a non-uniform temperature. *Boundary-Layer Meteorology* 1971; **2**:7–29.
23. Deppe, A J and Gallus Jr, W A and Takle, E S. A WRF ensemble for improved wind speed forecasts at turbine height. *Weather and Forecasting* 2013; **28**:212–228.
24. Rife, D and Davis, C. Verification of temporal variations in mesoscale numerical wind forecasts. *Monthly Weather Review* 2005; **133**:3368–3381.
25. Greaves B, Collins J, Parkes J. Temporal forecast uncertainty for ramp events. *Proceedings of the European Wind Energy Conference and Exhibition (EWEC), Marseille, France, 15-19 March*, 2009.
26. Cutler, N and Kay, M and Jacka, K and Nielsen, T S. Detecting, categorizing and forecasting large ramps in wind farm power output using meteorological observations and WPPT. *Wind Energy* 2007; **10**:453–470.
27. Kamath C. Associating weather conditions with ramp events in wind power generation. *Proceedings of the IEEE/PES Power Systems Conference and Exposition (PSCE), Phoenix, Arizona, USA, 20-23 March*, 2011.
28. Zack JW, Young S, Cote M, Nocera J, Aymami J, Vidal J. Development and testing of an innovative short-term large wind ramp forecasting system. *Proceedings of the European Wind Energy Conference and Exhibition (EWEC), Warsaw, Poland, 20-23 April*, 2010.

29. Bossavy A, Girard G, Kariniotakis G. Forecasting uncertainty related to ramps of wind power production. *Proceedings of the European Wind Energy Conference and Exhibition (EWEC), Warsaw, Poland, 20-23 April*, 2010.
30. Kamath C. Understanding wind ramp events through analysis of historical data. *Proceedings of the IEEE/PES Transmission and Distribution Conference and Exposition, New Orleans, Louisiana, USA, 19-22 April*, 2010.
31. Sevlian, R and Rajagopal, R. Detection and statistics of wind power ramps. *IEEE Transactions on Power Systems* 2013; **28**:3610–3620.
32. Ferreira C, Gama J, Matias J, Botterud A, Wang J. A survey on wind power ramp forecasting. *Technical Report ANL/DIS-10-13, p. 9*, Argonne National Laboratory. December 2010.
33. Hanssen AW, Kuipers WJA. *On the Relationship Between the Frequency of Rain and Various Meteorological Parameters*. Staatsdrukkerij- en Uitgeverijbedrijf, 1965.
34. Gandin KS, Murphy AH. Equitable scores for categorical forecasts. *Monthly Weather Review* 1992; **120**:361–370.
35. Hogan RJ, Ferro CJT, Jollife IT, Stephenson DB. Equitability revisited: Why the ”Equitable Threat Score” is not equitable. *Weather and Forecasting* 2010; **25**:710–726.
36. Orths AG, Eriksen PB. *Wind Power in Power Systems*, chap. Wind Power in the Danish Power System. 2 edn., John Wiley & Sons, Ltd., 2012; 531–532.
37. Neyman, J. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society A* 1937; **236**:333380.
38. Johnson RA. *Miller & Freund's Probability and Statistics for Engineers*, chap. Inferences Concerning Variances. seventh edn., Prentice Hall, 2004; 283.
39. Welford BP. Note on a method for calculating corrected sums of squares and products. *Technometrics* 1962; :419–420.
40. Jónsson, T and Pinson, P and Nielsen, H Aa and Madsen, H. Exponential smoothing approaches for prediction in real-time electricity markets. *Energies* 2014; **7**:3710–3732.
41. R Core Team. R: A language and environment for statistical computing. Version 3.0.2. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Wind Energ. 2015; **00**:1–18 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

Appendix C

Journal publication III — *draft*

Correction of NWP model deficiencies by statistical postprocessing

Martin H. Rosgaard[1,2]*, Henrik Aa. Nielsen[2], Torben S. Nielsen[2], Andrea N. Hahmann[1]

[1]*DTU Wind Energy, Technical University of Denmark, DK-4000 Roskilde, Denmark.* [2]*ENFOR A/S, DK-2970 Hørsholm, Denmark.*

*Correspondence to: Martin Haubjerg Rosgaard, Technical University of Denmark. Email: mhros@dtu.dk

The objective in this article is twofold. On one hand, a postprocessing framework for improved operational numerical weather prediction (NWP) of wind speed is described, developed and evaluated. On the other hand, the approach explored hints potential formulation issues in an operational NWP model generating global weather forecasts four times daily, with numerous users worldwide. The analysis is based on two years of hourly wind speed time series measured at three locations; offshore, in coastal and flat terrain, and inland in complex topography, respectively. Based on the statistical model candidates inferred from the data, the lifted index diagnostic is consistently found among the NWP model predictors of the best performing statistical models across sites.

Key Words: numerical weather prediction; statistical forecasting; wind energy scheduling; general linear modelling; model output statistics; backward elimination; backward stepwise selection; bayesian information criterion

Received . . .

1. Introduction

Modern-day studies of the atmosphere heavily rely upon data produced by numerical weather prediction (NWP) models and the study of systematic deviations of weather simulations from the observed atmosphere contributes to the continued improvement of NWP model formulations, which in turn better enables researchers to make candid statements on issues relating to climate and weather. Many forecast users worldwide depend critically on the accuracy of operational NWP models, with implications for a diverse range of applications from natural disaster risk assessment to efficient integration of renewable energy sources in the electrical grid.

As the governing equations describing the spatio-temporal continuum of observed atmospheric dynamics must be discretised in practice to enable temporal integration such that future atmospheric states can be inferred, Judd *et al.* (2008) argue that NWP models can be viewed as occupying a phase space of much lower dimensionality than the observed atmosphere. Leith (1974) distinguishes the subaspect of inadequate representation of atmospheric dynamics on numerically unresolved scales as a separate issue, further comminuted e.g. by Holton and Hakim (2013) into key physical process classes pertaining to 1) radiation, 2) clouds and precipitation, and 3) turbulent mixing and exchange, respectively. With the insight gained from the pioneering work of Courant *et al.* (1928, 1967), among others, Charney (1947, 1948, 1949) recognised, from the very outset of NWP on electronic computers, the need to filter solutions to the coupled nonlinear partial differential equations describing the temporal evolution of the atmosphere. Inadequate representation of dynamics on the sub-grid scale in NWP models, in addition to filtering of the governing equations for atmospheric dynamics, may spur systematic deviation between simulated time series for atmospheric quantities and corresponding observations. Furthermore, each numerical discretisation scheme has implications for prediction accuracy and numerical stability, e.g. Bénard (2003) finds stability properties of time-integration schemes commonly employed for NWP to be highly dependent on fundamental choices, such as the vertical coordinate.

This article concerns detection of systematic coupling of various NWP modelled quantities to atmospheric flow, and to this end outlines a postprocessing methodology for assessing explanatory value within a Model Output Statistics (MOS) framework, cf. Glahn and Lowry (1972). Systematic error in NWP models is usually quantified in terms of forecast variable bias, which is formally defined as the expectation of the forecast error; see e.g. Dee and Da Silva (1998) for a description of a theoretical framework for bias estimation. In practice, bias is estimated as the average error for a sufficiently long historical dataset of NWP model simulations and observations, while Dee (2005) relaxes the definition somewhat by accepting running-mean error approaches for short time periods as estimators of bias. E.g. Hacker and Rife (2007) adopt this definition in constructing and evaluating a running-mean algorithm for estimation of systematic errors for near-surface temperature on a limited area NWP model (LAM) grid, and for another LAM Sweeney and Lynch (2011) compare running-mean window lengths spanning the previous 3 and 28 days, respectively, to a Kalman (1960) filter approach, and a weighted combination of the three, as statistical postprocessing methods for wind speed NWP data. In a subsequent study, Sweeney *et al.* (2013) increase the resolution of the LAM domains and evaluate performance of seven different postprocessing approaches, including linear regression (MOS), running-mean bias correction, Kalman filtering, artificial neural networks, as well as forecasting techniques involving wind direction dependency, diurnal effects, and mean and variance nudging. They find none of these superior, and that

it is difficult to improve performance beyond simple running-mean bias correction. Stensrud and Yussouf (2005) find bias correction forecast performance competitive or better than MOS for their NWP multimodel ensemble approach, using data from the previous 12 days for bias correction, though Cheng and Steenburgh (2007) arrive at general MOS superiority in their comparison to running-mean and Kalman filter techniques, using data from 145 observation stations scattered across the western United States of America, except during a persistent cold air pool event in February 2005 where the Kalman filter approach is found to perform better. In the study by Hart *et al.* (2004) MOS forecasts based on coarsely resolved LAM data outperform high-resolution data from the same LAM configuration. Thorarinsdottir and Gneiting (2010) introduce wind speed as response variable in extension to the nonhomogeneous Gaussian regression forecasting approach first suggested by Gneiting *et al.* (2005), and find substantial improvement relative to reference forecasts. Pinson (2012) proposes another single-model NWP ensemble MOS approach for wind forecasting, namely adaptive and recursive Maximum Likelihood (ML) estimation of parameters in bivariate models for wind component ensemble mean and variance, and demonstrate better deterministic and probabilistic scores upon applying the method on a three-year dataset. Wilks (2014) describes a framework for multivariate ensemble MOS using empirical copulas and examines how recently proposed variations of these nonparametric functions, linking multivariate distribution functions to their constituent univariate marginal distributions, perform when the method is applied for probalistic heat wave forecasting. Statistical postprocessing that accounts for multivariate joint forecast distributions are not pursued here, albeit this would be a natural extension of the study presented in the event the problem extension or reformulation at hand warrants a multivariate approach.

The article is organised as follows. Section 2 outlines the dataset on which the regression modelling framework presented in Section 3 is applied. In particular, Section 3 specifies the underlying rationale for each step in the construction of statistical model candidates. Section 4 describes the findings, and the conclusion follows in Section 5.

2. Data

Kanamitsu (1989) describes an early version of the Global Forecast System (GFS), one in a handfull spectral NWP models used to generate global forecasts operationally, developed and maintained by the National Centers for Environmental Prediction (NCEP) in the United States of America. Four GFS forecast initialisations are available daily on a 0.5° global grid, downloadable with a few hours delay. Since the model upgrade 27 July 2010 the time integration of the governing equations was carried out at horizontal resolution T574, which roughly corresponds to a 27 km horizontal distance between computational grid points, though 14 January 2015 the horizontal resolution of operational GFS forecasts was upgraded to T1534 (13 km), and the number of vertical model layers increased from 64 to 90. Forecasts at the locations studied in this work are operational T574 GFS datasets spanning October 2012 through September 2014, and hence derive from the *static* model formulation available in a realtime setting during the time period 5 September 2012 up till 14 January 2015. Results presented in Section 4 thus reflect forecast improvement, achievable in practice operationally. A two-year dataset is also considered e.g. in Cassola and Burlando (2012); Sweeney *et al.* (2013); two studies related to this work.

Geographical locations of the three wind farm sites studied are sketched in Figure 1. The wind speed reference data underlying

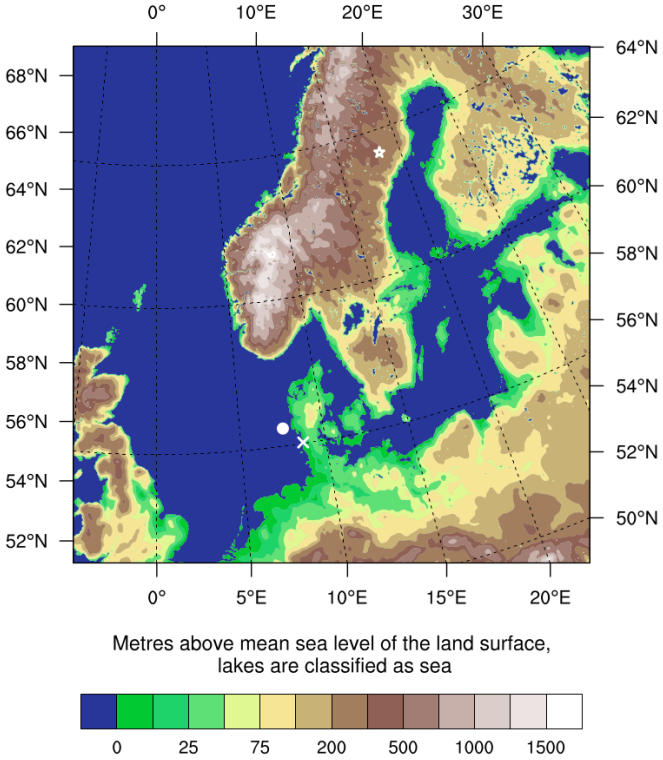


Figure 1. Geographical locations of the three wind farm sites studied, each marked with a white symbol. Horns Rev (bullet), Rejsby Hede (X), Stor-Rotliden (star).

results presented in Section 4 are based on the spatial average of wind speed time series recorded at the wind turbine nacelles of Horns Rev, Rejsby Hede and Stor-Rotliden wind farms, respectively. That is; the wind speed forecast reference datasets represent $\sim 1 \times 2\,\mathrm{km}^2$ (Rejsby Hede) and $\sim 4 \times 5\,\mathrm{km}^2$ (Horns Rev and Stor-Rotliden) horizontal wind speed fields. Wind power is usually traded on an hourly-mean basis, cf. e.g. Jónsson *et al.* (2014), and forecasts and measurements are hence analysed as hour-averaged time series, respectively, thus further smoothing rapid fluctuations in atmospheric flow intensity measurements not resolved explicitly in any NWP model. An overview of the dataset is given in Table I.

Forecasted wind components and temperature 80 m above ground level (AGL) are used for Horns Rev as this is near the height above mean sea level of the wind turbine anemometers recording the wind speed, while 100 m AGL forecasts are relevant for comparison to Stor-Rotliden data. Rejsby Hede wind turbine hub-height is 45 m AGL so 10 m AGL and 80 m AGL wind speed forecasts are vertically interpolated to this height based on

Table I. Statistics for wind speed measurement time series used for model training and evaluation at Horns Rev (HR), Rejsby Hede (RH), and Stor-Rotliden (SR). Wind farm acronyms are followed by the year number, separated by *slash*. Year 1: October 2012 till October 2013. Year 2: October 2013 till October 2014. *Complete pairs* refers to the percentage of datapoints for which both observation and forecast is nonmissing. Other quantities are in m s^{-1}. Wind speed time series mean is denoted $\hat{\mu}$, the median $\hat{m}$, and standard deviation $\hat{\sigma}$. The usual meteorological bias sign convention applies, i.e. forecast *minus* observation.

Site/ year	Complete pairs [%]	OBSERVATIONS $\hat{\mu}$	$\hat{m}$	$\hat{\sigma}$	FORECASTS $\hat{\mu}$	$\hat{m}$	$\hat{\sigma}$	Bias
HR/1	96.0	8.55	8.25	3.74	8.85	8.42	3.96	0.30
HR/2	97.7	9.30	9.07	4.25	9.52	9.21	4.60	0.25
RH/1	92.4	5.87	5.35	3.19	7.39	7.06	3.28	1.68
RH/2	95.4	6.46	5.98	3.51	7.97	7.68	3.72	1.66
SR/1	93.5	6.55	6.46	2.39	5.63	5.45	2.55	-0.94
SR/2	92.6	6.67	6.43	2.76	5.94	5.59	2.92	-0.76

the logarithmic wind profile relation according to Monin (1970), assuming neutral atmospheric stability conditions and vanishing zero plane displacement (the terrain is flat cropland with a few scattered patches of trees and houses nearby). Wind direction and temperature are linearly interpolated. Forecast data from the closest four computational grid points is bilinearly interpolated horizontally to the coordinates of the three wind farm centres, respectively. GFS predictions of pressure, the friction velocity at the surface, and measures of air mass rotation in the atmospheric boundary layer and the rest of the troposphere are included in the analysis, as well as heat fluxes, different indicators of atmospheric stability, and variables relating to humidity and precipitation. Also, a scalar estimate of the local geopotential height gradient at 700 mb, quantified by the sum of gradient magnitudes in the two horizontal dimensions, is included to approximate dependence on the geostrophic wind in the modelling framework. In this respect, the geopotential height anomaly – here understood to be the deviation from the *average over time*, cf. e.g. Yuan *et al.* (2008) – may contribute explanatory value, effectively coupling synoptic scale weather system dynamics to local conditions. The surface pressure tendency is closely coupled to this quantity and is used in the model development, alas the available GFS output is temporally coarse.

3. Method

The MOS approach outlined in the following sections is based on general linear modelling, which involves mathematically relating one (univariate modelling), two (bivariate) or several (multivariate) dependent variable(s) mathematically to one or more independent variables assumed to explain variation observed for the dependent variable(s). The terminology of denoting a dependent variable the model response, as employed e.g. in Hastie *et al.* (2009); Olsson (2001); Madsen and Thyregod (2010) is adopted here. Alternatively, the response is also known as *predictand* in Termonia and Deckmyn (2007); Wilks (2011); Warner (2011) or *outcome* in Pawitan (2001), while an explanatory variable is usually termed *regressor* denoting continuous data only, or *predictor* which is a broader term also denoting e.g. categorical data, etc. In contrast to the case for deterministic models, e.g. weather and climate models, statistical models account for the possibility that a relationship is not perfectly as modelled by allowing for unexplained (stochastic) response variation ε for which independence across observations and a distribution is typically assumed for model coefficient estimation and inference purposes. Deviation between statistical model predictions from corresponding observations are usually termed model *residuals*, denoted ϵ in this article.

In the present context, the basic premise is that measured wind speed can be univariately modelled as linearly dependent on additive forecast variable interaction terms. That is; in a multiple linear regression framework the response is related to a sum of predictor terms, each with an unknown coefficient to be estimated analytically by the method of ordinary least squares (OLS), cf. e.g. Hayashi (2000), based on a sufficiently large dataset of response and predictor variables. Statistically modelled quantities are in upper case to distinguish these stochastic variables from their corresponding realisations, which are in lower case. The study is based on the general linear model for hourly wind speed W,

$$\mathsf{W} = w_{LLS}\,\Theta(\boldsymbol{x}) + \varepsilon, \quad \text{with} \quad \Theta(\boldsymbol{x}) = \begin{bmatrix} 1 & \boldsymbol{x}^{\mathrm{T}} \end{bmatrix}\boldsymbol{\theta}, \qquad (1)$$

where ε denotes the model error, $\boldsymbol{\theta}$ is a column vector containing the coefficients to be estimated based on NWP data and measurements w for the sites studied, cf. Section 2, and $\boldsymbol{x}$ is a column vector containing quantities derived from NWP modelled quantities. Termonia and Deckmyn (2007) propose to infer viable predictors based on direct inspiration from NWP formulation and sub-grid scale parameterisations, yet in this work the choice of $\boldsymbol{x}$ elements in Eq. (1) is more vaguely guided by physical intuition, thus placing the MOS predictor range considered here somewhere in between the mechanistic-inspired strategy of Termonia and Deckmyn (2007) and purely empirical (data mining) statistical modelling, cf. e.g. Pawitan (2001) p. 5 on the respective merits of scientific formula and empirical models. Physical characteristica of $\boldsymbol{x}$ element candidates are briefly outlined last in Section 2; n.b. $\boldsymbol{x}$ is understood to represent continuous as well as derived *categorical* (binary only) variables, coded as dummy variables cf. e.g. Olsson (2001) pp. 12–21. The central predictor of Eq. (1), w_{LLS}, denotes forecasted wind speed after removal of bias and scaling errors, as recommended by Fritsch *et al.* (2000) and employed e.g. in Woodcock and Engel (2005); Engel and Ebert (2007, 2012), i.e. w_{LLS} data points are linear least square (LLS) predictions from the model

$$\mathsf{W}_{LLS} = \begin{bmatrix} 1 & w \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} + \varepsilon, \qquad (2)$$

with raw forecasted wind speed w, cf. e.g. Sweeney *et al.* (2013) from which the LLS acronym is adopted in this article. With all additive terms proportional to w_{LLS} the Eq. (1) modelled wind speed approaches zero when w_{LSS} approaches zero.

In contrast to the binning approach employed in Sweeney *et al.* (2013), forecasted wind direction is here implemented in Eq. (1) as Fourier expansion terms

$$f(w_d) = \sum_{k=1}^{\mathsf{N}} a_k \cos(kw_d) + b_k \sin(kw_d), \quad w_d \in (-\pi, \pi] \qquad (3)$$

up to $5th$ order; $\mathsf{N} = 5$, to approximate wind speed dependence on a periodic function of forecasted wind direction w_d. Parameters a_k and b_k are estimated as $\boldsymbol{\theta}$ coefficients in Eq. (1), i.e. $\cos(kw_d)$ and $\sin(kw_d)$ are $\boldsymbol{x}$ elements, respectively.

3.1. *Forecast performance assessment*

One of the two key purposes of the study is to explore the extent to which NWP model forecasts can be postprocessed empirically to advance efficient operational integration of wind energy in the electrical grid. Madsen *et al.* (2005) discuss two commonly employed scalar metrics for wind power prediction accuracy assessment, namely the *mean absolute error* (MAE) and the *root mean square error* (RMSE), each addressing slightly different aspects of forecast accuracy. As the expansion of the OLS acronym suggests, the procedure for estimation of $\boldsymbol{\theta}$ elements in Eq. (1) involves minimisation of the squared residuals ϵ^2. Hence, the modelling framework is tuned towards RMSE minimisation. Compared to RMSE, the MAE accuracy metric is less sensitive to severe forecast errors and is therefore arguably more relevant if the assessment should reflect normal operational conditions for day ahead wind energy scheduling. When used for performance evaluation along with RMSE, the MAE constitutes a baseline for inference on the variation in the magnitude of errors. Since both systematic and random errors contribute to MAE and RMSE, these metrics are sensitive to NWP model bias in raw (unprocessed) forecasts. In contrast, the absolute value of the *Pearson correlation coefficient* (PCC) is invariant under linear transformations of one or both constituent variables and hence not sensitive to bias and scaling errors.

As mentioned, RMSE and, to a lesser extent, MAE are commonly used for wind speed and power forecast performance assessment, cf. e.g. Madsen *et al.* (2005); Pinson (2012); Sweeney *et al.* (2013); Müller (2011), though e.g. Ranaboldo *et al.* (2013)

also include the PCC for evaluation of wind power predictions. Although rarely employed in wind energy forecasting, the inclusion of the PCC as estimator of forecast performance for the statistical models developed here – alongside MAE and RMSE – is indirectly motivated by the work of Murphy (1988). He meticulously highlights the deficiency of the PCC squared as a weather forecast assessment metric *when used by itself*, in that he demonstrates that forecast skill (in terms of mean square error) can be broken into a sum of the squared PCC and two bias-sensitive terms. Murphy (1988) argues that the PCC squared resemble *potential* skill attainable post elimination of systematic errors, while bias-sensitive performance metrics such as the RMSE and MAE describe *actual* performance. Due to wind speed forecast bias-removal cf. Eq. (2) prior to inclusion in Eq. (1), RMSE and PCC will exhibit the same relative performance between Eq. (1) model candidates studied for a given site, yet *between sites* the balance between "potential skill indicator" PCC and actual performance in terms of MAE/RMSE *approximately* emphasises differences in atmospheric flow distortion at each site, a feature in part subject to the local terrain due to the distinct surroundings of the offshore, coastal and inland sites selected in the experimental design. This picture is *approximate* as the interpretation above is subject to viewing $\Theta(\boldsymbol{x})$ in Eq. (1) as an ordinary linear model coefficient, i.e. disregarding the $\boldsymbol{x}$-dependence. It then follows from the definition of the coefficient of determination for Eq. (1) – which is in fact equal to the PCC squared and usually denoted R^2 cf. e.g. Olsson (2001) p. 5 – that

$$\sum_{\text{all } i} \epsilon_i^2 = \left(1 - R^2\right) \sum_{\text{all } i} \left(\mathsf{w}_i - \overline{\mathsf{w}}\right)^2 , \tag{4}$$

in which RMSE squared times the number of time series data points represents the left hand side, and the sum on the right hand side represents the variation in wind speed measurements, i.e. $\overline{\mathsf{w}}$ denotes the average of wind speed measurements. The superposed bar retains this denotation in the following.

Hence, by including PCC alongside RMSE (and MAE) in the results pertaining to wind *speed*, presented in Figure 6, the reader is indirectly reminded of the site statistics of Table I, namely $\hat{\sigma}$. However, with regard to wind *power* only RMSE and MAE are applied in Table III, as is customary cf. Madsen *et al.* (2005).

3.2. *Model development strategy*

The dataset available spans two years, cf. Table I. Where e.g. Müller (2011); Sweeney and Lynch (2011) use moving window recursive approaches for one-year datasets, the first year of data is here reserved for model development, while data for the second year is solely used to assess performance of the final wind speed model candidates based on the structure outlined in Eq. (1) — a special case of the cross validation technique, cf. e.g. Efron and Gong (1983). This stategy substantiates out-of-sample any systematic wind speed forecast error coupling to GFS variables identified in-sample. Furthermore, all seasons are represented in both of the datasets for model development and evaluation, respectively, and the results presented are therefore more generic than if a subyearly period, sensitive to seasonal effects, had instead been analysed. The dataset is binned according to forecast leadtime groups of width equal to the time period between NWP model forecast initialisations. Data within each forecast leadtime group is then augmented such that all wind speed measurement data points occur exactly once in each group. In this way all hours of the day are represented in each forecast performance metric evaluation and inference on forecast performance leadtime dependency is thus readily assessed.

As mentioned in Section 2, forecasts are issued four times daily, and wind energy scheduling typically concerns the forecast leadtime horizon from 12 to 48 hours ahead, depending on application requirements and the latency with which forecast data can be obtained. E.g. Messner *et al.* (2013) consider forecast leadtime hours 12 to 48 for studying forecasted wind speed transformation to wind turbine power production, a topic briefly considered in Section 4.4. To simplify execution of the final, empirical model development step explained in Section 3.5 only two daily forecast initialisations are included in the model development phase, namely Coordinated Universal Time (UTC) midnight and noon forecast initialisations. Thus, for each site Eq. (1) modelling cf. sections 3.4 and 3.5 is carried out separately for the three sites considered and forecast leadtime hour groups $12 \leftrightarrow 24$, $24 \leftrightarrow 36$ and $36 \leftrightarrow 48$, respectively, and hence all wind speed measurements of the resulting nine datasets are represented exactly once, respectively. For the evaluation of the NWP postprocessing framework in Section 4 all four daily GFS initialisations are employed.

Before carrying out the Eq. (1) development steps described in sections 3.4 and 3.5 separately for the nine datasets, the GFS-derived quantities included in the analysis, cf. the last paragraph of Section 2 and the Eq. (1) description first in Section 3, are preliminarily screened for explanatory value subject to the Eq. (1) framework in Section 3.3. As noted above, the procedure outlined in sections 3.3, 3.4, and 3.5 is based on year 1 data, cf. Table I.

3.3. *Preparatory screening of predictor candidates*

This section describes a few diagnostic steps to be carried out for the dataset used for model structure identification and coefficient estimation (year 1; Table I) in order to preprocess the numeric predictor candidates for modelling. In accordance with the model development strategy outlined in Section 3.2, the quantification of the w_{LLS} data used in Eq. (1) model development is based on forecast data spanning leadtime hours 12 to 48. That is, Eq. (2) fits are based on data for GFS forecast hours 12 to 48 in this section.

Following a standard regression model building approach one may choose to plot Eq. (2) model residuals against potential additional explanatory variables, i.e. $w_{LLS}x$ terms — n.b. that said residuals; $\mathsf{w} - w_{LLS}$, are henceforth referred to as *LLS error*. However, each Eq. (1) model predictors can be written $w_{LLS}(x - \overline{x}) + w_{LLS}\overline{x}$, and note that the latter term introduces variation purely due to w_{LLS}. Therefore, the Eq. (2) model residuals are instead plotted against $w_{LLS}(x - \overline{x})$. In the event a linear relation appears a plausible approximation to the underlying GFS intervariable coupling to advection, the corresponding Eq. (1) term is kept unchanged as a quantitative predictor. Local linearity within regimes divided by a $w_{LLS}x$-threshold is accommodated by adding separate linear terms that mutually excludes one-another depending on a threshold value. That is, depending on the value of $w_{LLS}x$ one or the other term is active, always one of them and never both. Finally, if no linear relation is evident, though a regime-like distribution can be inferred from the scatter density plot, or perhaps more clearly from a histogram for the numeric values of the $w_{LLS}x$ term, the predictor candidate is included as a categorical variable with regime-dividing values inferred from the scatter plot, cf. $w_{LLS}x$ regime-dependent linear relationships discussed before. In case the threshold value differs considerably for the three reference data sites studied the predictor term is discarded to keep the methodology generic across terrain types.

Continuous NWP modelled quantities describe different aspects of the the atmosphere via a set of coupled partial differential equations and sub-grid scale parameterisations, and are hence correlated to some extent. The degree of multicollinearity in the Eq. (1) model is examined by inspection of the PCC matrix for the predictor terms, which provides a rough estimate of the extent to which redundant information is

included in the model. Since the PCC is only sensitive to linear relationships, the interpredictor PCC matrix is supplemented by inspection of interpredictor scatter density plots in order to infer whether systematic nonlinear relations between $w_{LLS}x$ terms are present. However, as noted e.g. by Williams *et al.* (2013) the adverse effects of multicollinearity are typically not severe in a prediction context, as pertains to this work.

3.4. *Model reduction based on Maximum Likelihood criteria*

Eq. (1) predictor terms containing the least explanatory value are removed sequentially according to a stepwise process known as *backward elimination*, e.g. in Wilks (2011), or *backward stepwise selection* in Hastie *et al.* (2009). The backward elimination strategy is employed as the full model is thus subject to study, in contrast to the case for *forward stepwise selection*, where each predictor is added and tested sequentially. Note that both techniques have advantages and drawbacks, depending on the modelling task and the dataset at hand. Cf. Section 3.3, scatter density plots of LLS error against $w_{LLS}\boldsymbol{x}$ terms can indicate linear coupling for certain variables, while others may look less promising as explanatory variables in the assumed response-relation structure Eq. (1). The objective is then to identify and exclude redundant information to avoid suboptimal and inconsistent model coefficient estimates, as discussed e.g. in Birgé (2006); Bahadur (1958).

The Akaike Information Criterion (AIC), first proposed by Akaike (1974), and the Bayesian Information Criterion (BIC), Schwarz (1978), are two penalised ML criteria, both dependent on the number of coefficients to estimate and the latter also explicitly depends on the number of data points available for estimation;

$$\mathrm{AIC} = -2\ell(\mathrm{W}) + 2n, \quad \mathrm{BIC} = -2\ell(\mathrm{W}) + n\log(N),$$

where $\ell(\mathrm{W})$ denotes the maximised value of the log-likelihood function for the model specified by Eq. (1), n is the number of coefficients $\boldsymbol{\theta}$ to be estimated, and N is the number of observations. As e.g. pointed out in Olsson (2001) p. 3, OLS estimation of Eq. (1) coefficients corresponds to ML estimation if the errors ε are assumed independent and zero-mean normal distributed, which in turn enables the AIC and BIC as Eq. (1) model reduction tools. Results discussed in Section 4.2 motivates the Gaussian-ε assumption, while the issue of finite autocorrelation in part motivates the choice of backward elimination scheme, cf. the next paragraph. The AIC and BIC are common tools for model selection within various scientific disciplines, see e.g. Kuha (2004); Alfaro and Huelsenbeck (2006); Jones (2011); Vrieze (2012), while alternatives such as bootstrap criteria, cross validation criteria, Mallow's C_p, etc. are also found in the literature, see e.g. Hastie *et al.* (2009); Zucchini (2000); Bulteel *et al.* (2013); Birgé (2006). Vrieze (2012) argues that the BIC is *consistent* under certain assumptions, i.e. this criterion will asymptotically select the "true" model, given that this is a subset of the full model space investigated, whereas the AIC is not consistent. Hastie *et al.* (2009) p. 135 notes that the BIC is more strict in that it penalises the inclusion of redundant model predictors more severely than the AIC. However, given that the full true model is not a part of the model space studied the AIC is said to be *efficient* in that it will asymptotically choose the model that minimises the mean square error, while the BIC is not efficient if the true model is not a subset of the investigated model space. Model selection based on the AIC is always subject to the risk of selecting too large a model (i.e. with redundant information) regardless of the number of observations in the model development dataset, while the corresponding risk when applying the BIC is very small given a sufficient number of observations are available. On the other hand, the risk of selecting too small a model is higher with the BIC than with the AIC, regardless of the number of observations. The merits of the two model selection criteria can roughly be summarised in the following points.

- If a false negative finding is considered more misleading than a false positive, the AIC is preferable.
- If a false positive finding is considered more misleading than a false negative, the BIC is preferable.

Inclusion of redundant information in the coefficient estimation process can lead to poor predictive ability for the Eq. (1) model candidates when applied for postprocessing of future NWP data. Furthermore, measured wind speed and NWP modelled quantities are time series of smoothly varying atmospheric dynamics aspects and as such exhibit finite autocorrelation, which suggests that Eq. (1) model errors are not chronologically independent. Therefore the *effective* number of (independent) data points is lower than the actual number of data points N. Since the width of confidence intervals for Eq. (1) coefficient estimates scales inversely with $\sqrt{N}$, the test statistic used to accept/reject the null-hypothesis stating non-value for individual predictor terms is prone to attribute significance (rejection of null-hypothesis) to terms for which the test statistic would otherwise infer non-significance, subject to the effective number of independent observations. Thus, more predictor terms may survive the backward elimination process than warranted according to the assumption of independent observations. Hence, the BIC employs stricter predictor-penalisation than the AIC and the former is therefore preferred for backward elimination of Eq. (1) predictor terms. With the assumption of zero-mean Gaussian Eq. (1) errors ε with variance σ_ε^2, Hastie *et al.* (2009) p. 233 note that the BIC can be expressed in terms of the Eq. (1) residual mean $\bar{\epsilon}$,

$$\mathrm{BIC} = \frac{N\bar{\epsilon}}{\sigma_\varepsilon^2} + n\log(N)$$

Since the BIC, although more restrictive than the AIC, may yet keep too near-redundant predictor terms in Eq. (1), the model reduction process is extended with additional model selection criteria inferred from the models resulting from BIC reduction.

The BIC model reduction process allows for elimination of the standalone w_{LLS} term in Eq. (1), in which case the predictor is reinstated posthumously as the quantity corresponds directly to the model response. Likewise, not all Eq. (1) predictor involving wind direction terms, cf. Eq. (3), are found significant post BIC model reduction, and all wind direction terms up to the highest Fourier expansion order surviving the BIC reduction process are reinstated if eliminated during the process.

3.5. *Further model reduction based on empirical assessment*

Statistical significance of a predictor identified cf. the Eq. (1) model development approach outlined in Section 3.4 is no guarantee for notable influence on model predictions. E.g. the discussed issue of model residual autocorrelation may lead to exaggerated significance attribution for some predictors, which in turn increases the risk of sustaining a predictor set subject to severe multicollinearity. Hence, the Eq. (1) model development approach outlined in Section 3.4 is extended with assessment of the contribution of each predictor term to model prediction variation. This is done by inspecting distribution histograms for $x\hat{\theta}$ values, i.e. each element-wise $\boldsymbol{x}\hat{\boldsymbol{\theta}}$-product between GFS predictor candidates and the coefficient estimate corresponding to the relevant $w_{LLS}x$ predictor term in Eq. (1). The final Eq. (1) model development step is based on 95 % quantile ranges for said histograms. Based on this method, 1) the explanatory value for each predictor is gauged directly, and 2) any potential adverse

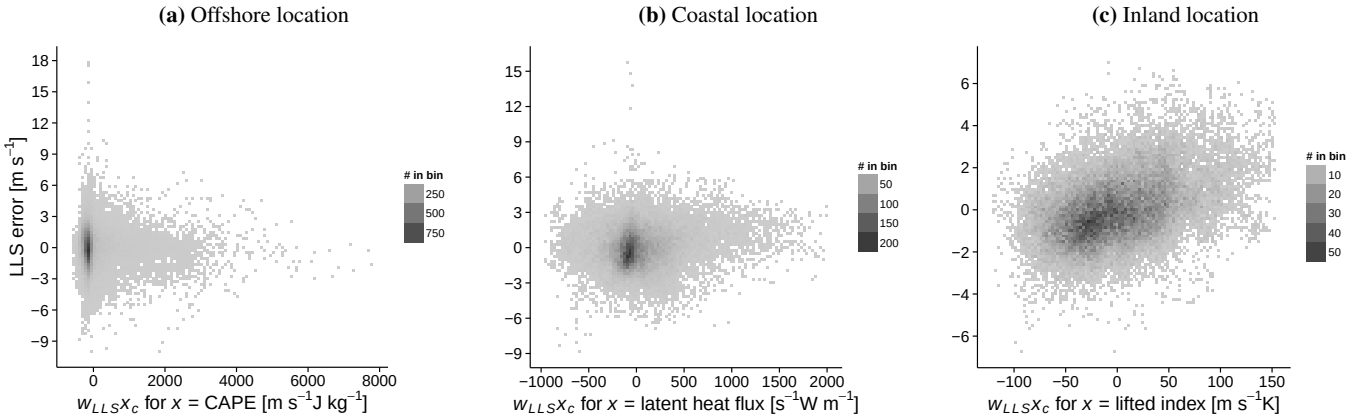


Figure 2. Scatter density plots for examples of GFS variables included as **(a)** binary categorical in modelling based on Horns Rev data,**(b)** two-regime linear regressor at Rejsby Hede, and **(c)** ordinary linear regressor for the Stor-Rotliden site.

effects of severe multicollinearity may be detected. The latter point is elaborated upon in the following paragraph.

Provided two or more predictor terms are highly correlated, the model design matrix $\boldsymbol{M}$, containing predictors along the column dimension and observations along the row dimension, multiplied with it's transpose, is near-degenerate since $\boldsymbol{M}$ itself is near-degenerate; $\det(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}) = \det(\boldsymbol{M}^{\mathrm{T}})\cdot\det(\boldsymbol{M})$. The OLS estimator for Eq. (1) coefficients

$$\hat{\boldsymbol{\theta}} = \left(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}\right)^{-1}\boldsymbol{M}^{\mathrm{T}}\mathbf{w} \tag{5}$$

has covariance

$$\mathrm{Cov}\left[\hat{\boldsymbol{\theta}}\right] = \sigma_{\varepsilon}^{2}\left(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}\right)^{-1} \tag{6}$$

cf. e.g. Pawitan (2001) p. 151. Model coefficient estimates become uncertain, cf. e.g. Valentine (1969), and hence the precise effect of a predictor severely collinear with one or more other predictors is thwarted. Recall from Section 3.3 that the overall fit of the model is not affected cf. e.g. Williams *et al.* (2013), though for this to remain true for varying degrees of uncertainty in model coefficient estimates a scenario with relatively large coefficient estimates – of opposite sign – being attributed to two strongly correlated predictors is plausible, in spite of the fact that neither term contains explanatory value. As modelling is carried out separately for each of the nine datasets discussed in Section 3.2, severe multicollinearity may result in pairwise wide $x\hat{\theta}$ histograms for some datasets subject to two strongly collinear terms, while in other cases the impact may be modest or said terms may have been eliminated in the BIC reduction process.

In order to effectively detect and deal with such illusive explanatory value signals, $x\hat{\theta}$ variation histograms and associated quantile ranges are supplemented by ordering according to PCC forecast performance reduction due to removal of a given predictor. That is; the predictor term causing the most severe reduction in prediction performance when removed from the model is at the top of the hierarchy, and so forth as the adverse performance impact from predictor removal decreases. If a predictor with a relatively wide 95 % quantile range assumes a low ranking this indicates strong interpredictor coupling that may be confirmed from the interpredictor correlation matrix and scatter density plots motivated in the last paragraph in Section 3.3. Valentine (1969) proposes to substitute the individual terms with a single linear combination of the collinear predictors, though also emphasises the entailing challenge of choosing appropriate weights for each term. Hence, his other suggestion of simply leaving out one of the collinear predictors is applied here.

As a very last screening for true explanatory value all predictor terms remaining at this point are left out by turn, the model re-assessed, and severety-ranking of adverse impact on model prediction performance is established in a final attempt to root out predictors of deceptive value for Eq. (1) model predictions. Inconsistency between $x\hat{\theta}$-variation contribution vs. the same terms ranking leads exclusion of said predictor.

4. Results

This section describes the assessment of the models resulting from carrying out the Eq. (1) predictor reduction steps of Section 3 for the dataset presented in Section 2. As prescribed in Section 3.3, the exploratory data analysis outlined in Section 4.1 preprocesses the data for the backward elimination of Eq. (1) predictor terms. The model predictor reduction process described in sections 3.4 and 3.5 is executed in Section 4.2, leading to the final model candidates evaluated in sections 4.3 and 4.4.

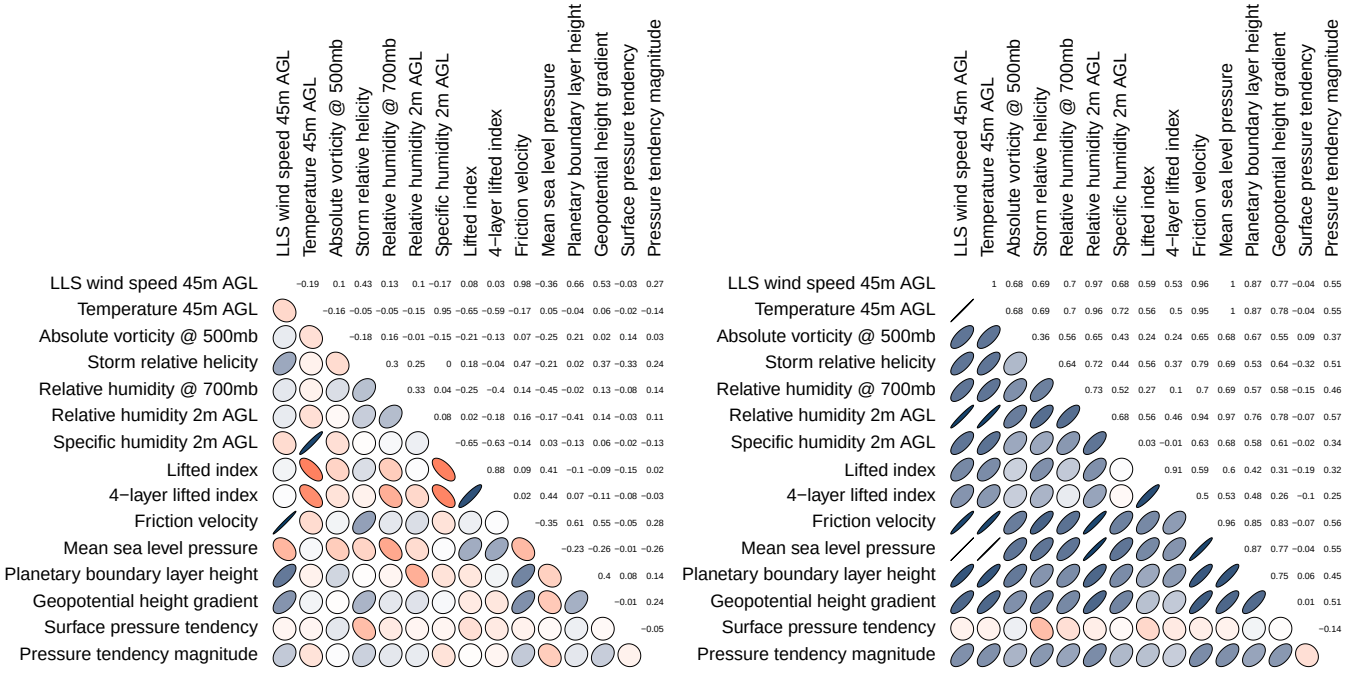
Note that sections 4.1 and 4.2 concern only 00 and 12 UTC forecast initialisations pertaining to year 1 of Table I, cf. the model development strategy laid out in Section 3.2, while all four daily GFS forecast initialisations are used in sections 4.3 and 4.4.

4.1. Exploratory data analysis prior to statistical modelling

Provided Eq. (1) holds, scatter density plots of LLS error against $w_{LLS}x_c$ terms constitute approximate guidance as to whether the predictor term should be 1) excluded, 2) included as a binary categorical variable, further substantiated by predictor term histograms cf. Section 3.3, 3) as a x-regime dependent term, or 4) as an ordinary regressor. Figure 2 includes examples of GFS-derived quantities processed as categorical, two-regime regressors, and ordinary continuous regressors prior to Eq. (1) predictor reduction based on the BIC carried out in Section 4.2.

1) The thermal buoyancy, and hence atmospheric stability, in the lowest part of the boundary layer arguably depends on variables such as daily sunshine duration and downward short wave radiation. Though based on scatter density plots cf. Section 3.3 these GFS quantities are excluded in the further analysis, as inclusion as neither categorical, regime-dependent linear, nor ordinary regressor appear suitable due to fairly evenly spread-out scatter with no hint of linear trends (not shown).

2) Figure 2a indicates for Horns Rev – as for the other two sites (not shown) – that convective available potential energy (CAPE) is likely not suitable for inclusion as continuous x in an Eq. (1) type predictor term, as no linear relation is evident from the scatter density plot. However, viewing the horizontal axis as a histogram – i.e. data points per $w_{LLS}x$ bin (not shown) – further highlights a tendency already apparent in Figure 2a for values to be either concentrated around zero or with values far exceeding zero, with approximately half the data points belonging to each



(a) Eq. (1) predictor w_{LLS} and standalone x variables.

(b) Eq. (1) predictors w_{LLS} and $w_{LLS}x$.

Figure 3. PCC matrices for the set of continuous GFS-derived x candidates included in the modelling process for the coastal Rejsby Hede site. As a correlation matrix is symmetric the lower diagonal is transformed to display visually interpretative PCC objects. Hence, colour coded representation substitutes numeric values in the lower diagonal; fading gradually from red (inverse linear relation), over white (no linear relation), to blue (linear relation).

binary class. From a physical point of view, whether potential energy of a parcel of air is present is one indicator of atmospheric stability, with implications for the vertical wind profile, and hence the GFS variables CAPE and the related convective inhibition (CIN) are implemented as binary categorical variables in the modelling process of Section 4.2. Forecast variables related to precipitation are indicators of water phase shift in the atmosphere, potentially associated with weather fronts, and are processed in a similar manner due to CAPE/CIN-like scatter density plot traits. Respective data point counts in the two binary factor levels are ensured to be distributed somewhere in between 30–70 % and 50–50 %; if one factor level is sparsely populated the Eq. (1) model coefficient estimation becomes unreliable and may crash if the design matrix $\boldsymbol{M}$ becomes singular, cf. Eqs. (5–6).

3) Moving on to Eq. (1) predictor terms based on regime-dependent x, Figure 2b shows the scatter density plot for latent heat flux simulated by the GFS. This variable, along with the related *sensible* heat flux, has a natural threshold at zero; separating condensation from evaporation for the former and vertical propagation direction for the latter.

4) Finally, the GFS-simulated lifted index vs. LLS error displayed in Figure 2c do suggest a linear relation, in contrast to the case for variables like CAPE and latent heat flux. Similarly, the scatter density plots for all GFS-derived quantities appearing in Figure 3 indicate no processing necessary, and these are therefore included as continuous x in the Eq. (1) model.

Figure 3a quantifies the linear coupling between the various quantities predicted from the GFS formulation of the governing equations for atmospheric dynamics. As all Eq. (1) predictor terms involve proportionality to LLS wind speed severe multicollinearity is expected and verified in Figure 3b, namely for temperature and mean sea level pressure. Notice that these terms are, in fact, correlated with LLS wind speed to two digits precision — and are hence likely not adding value to Eq. (1) prediction accuracy, though neither deteriorating model performance *in-sample* cf. e.g. Williams *et al.* (2013). Out-of-sample model performance consequences of inclusion of these terms are assessed in Section 4.3. As the final modelling step of Section 3.5 involve direct assessment of the influence of individual predictor terms on the model, all the GFS-derived quantities listed in Figure 3 are included in the backward elimination process of Section 4.2. The PCC matrices corresponding to Figure 3b for the offshore Horns Rev and inland Stor-Rotliden sites exhibit similar behaviour for all predictor terms (not shown here, though see Figure 7 for Stor-Rotliden). As remarked in Section 3.3 the PCC is only an estimator of linear dependence, though no conspicuous nonlinear relations are identified when inspecting interpredictor scatter density plots (not shown).

4.2. *Backward elimination of Eq. (1) predictor terms*

As prescribed in Section 3.2, the Eq. (1) predictor reduction based on the BIC is carried out separately for each of the nine datasets considered, each of which including all wind speed measurement data points exactly once. Which of the resulting predictor candidates to preserve in the final models is determined from combining the information obtained in the nine separate studies.

However, first the assumptions underlying the method proposed in Section 3.4 need to be substantiated. In order to employ the ML framework on which model predictor reduction using the BIC is based, Eq. (1) errors ε must be approximately normal distributed. Quantile-quantile plots for the nine datasets indicate fairly symmetrical distributions, yet a bit heavy-tailed (not shown). There are alternative approaches for nonnormal theory models, see e.g. Yuan and Bentler (1999), though the QQ-plots are not severely discouraging and hence the model development proceeds according to Section 3.4. The assumption that the ε distribution can be parameterised by two constants (the mean population error, $\mu_\varepsilon = 0$, and standard deviation σ_ε), which also underlies Eq. (5) – i.e. OLS estimation – implies homoscedasticity. In order to infer the extent to which this assumption is met, scatter density plots of model residuals against predictions are inspected (not shown), and no apparent dependence is found. Also, the cumulative squared model residuals are computed for the three sites studied. If σ_ε^2 is constant $\sum_{i=1}^{N} \epsilon_i^2$ vs. the time stamp will have near-constant slope. As data for the offshore site, Horns Rev, exhibits the most

severe sudden ϵ^2 increases disrupting this picture, the cumulative ϵ^2 for this site is shown in Figure 4, along with two examples of weather events underlying the ϵ^2-jumps; storms. Note that removing the two 24-hour storm periods results in what looks like a steady, constant-slope development which in turn implies homoscedasticity. However, removing the 4–5 events for which cumulative ϵ^2 exhibits sudden increase with time in Figure 4c has no impact on the final model candidates arrived at later in this section, and the impact on OLS estimates for $\boldsymbol{\theta}$ elements of Eq. (1) only influence model predictions up to the fourth decimal in the Section 4.3 analysis.

As argued for in Section 3.5, additional model reduction upon having BIC-reduced the nine Eq. (1)-type models separately based on the nine datasets described last in Section 3.2, is likely necessary to achieve consistent results. Thus quantile ranges displayed in Figure 5c, and similar plots for the other eight datasets (not shown), are used in order to better secure that out-of-sample performance (second year data) is consistent with in-sample performance (first year data). Cf. the discussion in Section 3.5 this information is applied as a tool for direct assessment of explanatory value of predictor terms remaining post backward elimination processing of Eq. (1) for the nine datasets. The $x\hat{\theta}$ histogram width, cf. Figure 5b, quantifies the extent of influence on the model response. Note that in comparison with Figure 5a the lifted index (x) histogram is wider than the model response coefficient histogram. This apparent discrepancy can be understood in light of Figure 3a, exhibiting similar traits as the corresponding PCC matrix for the Stor-Rotliden site, in that specific humidity is inversely correlated with lifted index and hence acts to counter the stronger influence of the lifted index. Thus the specific humidity histogram (not shown) is enveloped by the histogram for lifted index and acts to "contract" the variation

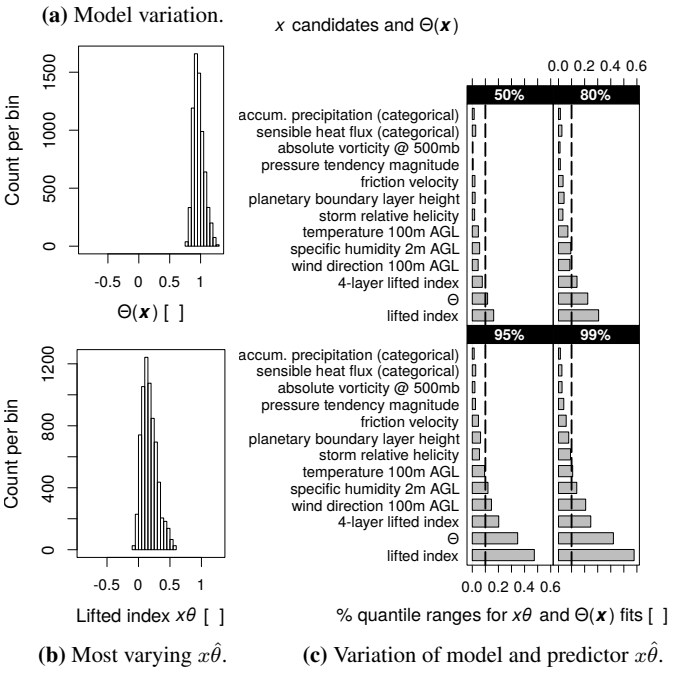


Figure 5. **(a)** and **(b)** show histograms of the coefficient estimates for the full Eq. (1) model and the fitted predictor term $x\hat{\theta}$ with greatest variation contribution ($\sim$ widest histogram), respectively. **(c)** shows quantile ranges for all predictor terms considered for this particular dataset, which is forecast leadtime hour group 12 to 24 for the Stor-Rotliden site.

of the lifted index. The variation histogram for temperature is narrower than that for specific humidity, cf. Figure 5c, and centered in the far negative around -1.6 (not shown). The lifted index variation, mainly of positive sign, is also balanced by the 4-layer lifted index (not shown), looking like a narrower mirror image around zero – i.e. of opposite sign – of the lifted index variation histogram.

The empirical reduction steps proposed in Section 3.5 and executed here conclude with recursively leaving one predictor out of the model in order to quantify the negative impact and construct a predictor-hierarchy of explanatory value. A clear example of the adverse effect of multicollinearity arise when including both pressure at the surface and at the mean sea level in the model development. For some of the nine datasets modelled, cf. the last paragraph in Section 3.2, both appear to be dominating predictors in terms of Figure 5-like histogram width, while for other datasets they do not even survive BIC-reduction. In the former case, inspection of the $\boldsymbol{\theta}$ estimates reveals very large values of opposite sign — the two corresponding predictor terms all but cancel when added in the full model. Or put differently, subject to severe predictor multicollinearity the framework presented in this article may provide false assessment of true explanatory value if this important step is omitted.

Section 3 frames the general guidelines for the analysis steps, though at this point model candidates must be inferred from the data based on the preliminary results. The modelling process produces nine quantile range plots cf. Figure 5c, and in the final screening predictor terms are tested according to the hierarchy-listing approach sketched above in order to rule out false signals. As a final criteria, at least 0.1 variation within the 95 % $x\hat{\theta}$-quantile range is required for one or more of the nine datasets; see the dashed vertical line in Figure 5c. When multiplying $x\hat{\theta}$ with w_{LLS} this corresponds to a 95 % quantile bandwidth of $\sim$1 m s^{-1} for LLS wind speeds around $\sim$10 m s^{-1}. The resulting model is denoted *Model 2* in Table II of Section 4.3.

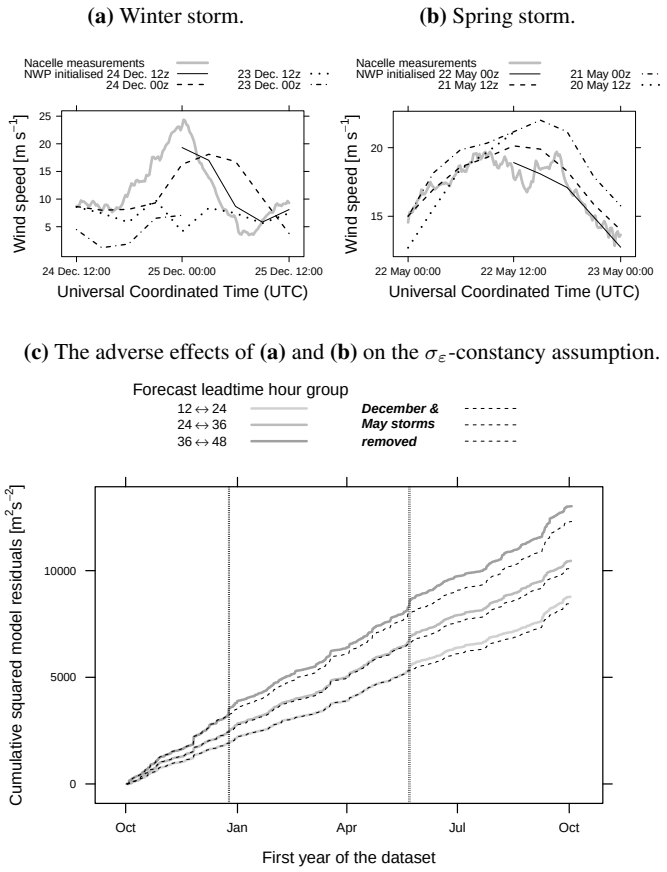


Figure 4. Quantitative illustration of how two poor storm event forecasts for Horns Rev challenge the assumption of unchanging σ_ε, which underlies the modelling framework. The time of the year when the two events, each of 24 hours duration, are removed from the dataset are marked with vertical dotted lines in **(c)**.

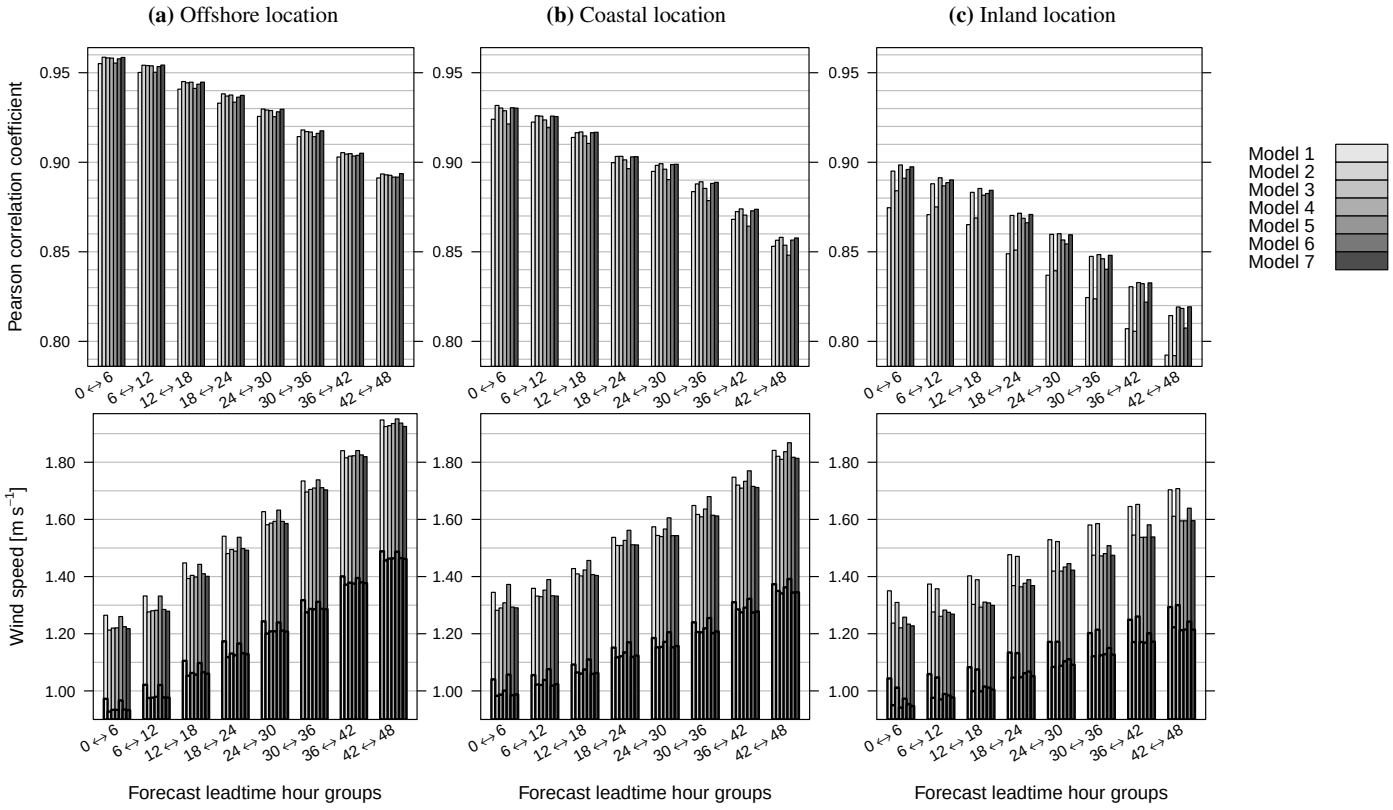


Figure 6. Top row: Forecast performance in terms of PCC. **Bottom row:** Forecast performance in terms of RMSE superposed by forecast performance in terms of MAE (bold type). **(a)** Horns Rev, **(b)** Rejsby Hede, **(c)** Stor-Rotliden.

4.3. Model performance assessment

In contrast to the case during model development steps carried out in sections 4.1 and 4.2, all four daily GFS initialisations; 00, 06, 12, and 18 UTC, are included in Section 4.3 as the 12 hours wide forecast leadtime groups were enforced to reduce the complexity of the empirical assessment by variation of quantile ranges. As such, the nine datasets visually assessed cf. Figure 5 would have been split in twice as many 6 hours wide leadtime group datasets if all four daily GFS initialisations had been included, and the benefit would arguably be negligible. In this section, the $\boldsymbol{\theta}$ elements of the final model candidates are estimated based on year 1 data and model performance evaluated based on year 2 data, cf. Table I. Thus the dependence of performance on the forecast leadtime is resolved twice as fine as in datasets underlying construction of the model candidates, while staying within the forecast performance assessment framework of Section 3.2.

As mentioned last in Section 4.2, the model arrived at there is denoted Model 2. A few additional model candidates are evaluated in order to nuance the performance investigation. A natural baseline is, of course, the LLS model, Eq. (2), which corrects for wind speed bias (see Table I) and scaling errors. This is the aptly named Model 1, a prerequisite for the other models in that Model 1 predictions are essential for modelling by Eq. (1). An additional perspective on performance in the present context; Model 3, could be to assess the effect of adding wind direction as is typically, if not always, done in statistical wind power prediction models. It turns out that, besides wind direction, the only other x variables present in all nine models post BIC-reduction are the lifted index and hub-height temperature, the further addition of which makes Model 4. Model 5 predictions quantify the effect of leaving out the wind direction from Model 4, while Model 6 and 7 assess the individual contribution to Model 4 of temperature and lifted index, respectively.

Model performance results are shown in Figure 6. Recall cf. Eq. (4) that the high forecast accuracy in terms of PCC and relatively poor RMSE in Figure 6a, compared to the corresponding metric balance in Figure 6c, highlights severe fluctuations in atmospheric flow off the Danish west coast relative to corresponding weather conditions in central Sweden, cf. Table I. Leaving out the wind direction (Model 1 and 5) has relatively severe impact on performance for the two Danish sites, while Model 7 for the inland Swedish site suffer a minor penalty gradually diminishing with leadtime until Model 5 achieves Model 4 performance around 30 hours. This is in line with expectation as namely the coastal site is situated $\sim$ 1–2 km from the Danish west coast and thus experience winds from the North Sea for westerly flow and winds subject to heterogenous land surface drag for easterly flow, which in turn makes direction-dependent forecast errors plausible. In fact, Model 5 (4 without wind direction) performs inferior to Model 1 (LLS wind speed) for the coastal site, i.e. the inclusion of temperature and lifted index deteriorates the out-of-sample wind speed prediction at this location. The Swedish site is situated on the top of a hill and surrounded by terrain of some heterogeneity, and hence the wind represented on the coarse $\sim$ 27 km horizontal computational grid of the GFS is more

Table II. Overview of models evaluated. Wind direction, Eq. (3), includes up to four Fourier expansion terms. Wind speed, wind direction, and temperature are at wind turbine nacelle hub-height, specific humidity is at 2 m AGL. † superscript denotes Eq. (2) modelling, while ‡ marks an Eq. (1) type model.

Model #	Constituent GFS-derived quantities included in predictor terms
Model 1†	Raw wind speed
Model 2‡	LLS wind speed, wind direction, temperature, lifted index, 4-layer lifted index, specific humidity, friction velocity, planetary boundary layer height
Model 3‡	LLS wind speed, wind direction
Model 4‡	LLS wind speed, wind direction, temperature, lifted index
Model 5‡	LLS wind speed, temperature, lifted index
Model 6‡	LLS wind speed, wind direction, temperature
Model 7‡	LLS wind speed, wind direction, lifted index

undisrupted by the topography in the simulation than is observed surface winds. This hypothesis is supported by the fact that Model 2 quickly drops to Model 1 performance as a function of leadtime, i.e. forecasted wind direction do not contribute much explanatory value here. Note also that Model 3 appears to be on par with, and occasionally better than, Model 6. And for the coastal site, in particular, it is even an advantage to further leave out temperature (Model 5). For the Swedish site Model 4 and 5 start out with similar performance, though temperature (Model 6) gradually drops as the leadtime increases, while the lifted index (Model 5) sustains the leading position.

Given that two or more models perform equivalently well the one with the least complexity is preferable. In that respect, the wind direction model (Model 3) appears to be the best candidate (yet cf. Section 4.4) for the coastal site, while slightly surpassed by Model 2 for the offshore site. Inland, Model 4 consistently outperforms the other model candidates, though approximately one day into the forecast temperature contributes very little explanatory value as Model 7 (Model 4 without temperature) performance merges with that of Model 4. Overall, predictability in terms of MAE and RMSE drops more rapidly for the offshore site than for the onshore sites and the former starts out better and ends up poorer, performancewise. Also, the relatively high PCC for the offshore site, albeit similar MAE/RMSE across sites, may indicate that severe GFS wind speed magnitude errors influence offshore prediction to a greater extent than weather system phase errors relative to locations on land, cf. Section 3.1 and the discussion on conditional (scaling errors) and unconditional (bias) contributions to forecast skill in Murphy (1988).

4.4. *Practical implications for renewable energy feasibility*

Motivation for considering all available forecast data up to 48 hours ahead is given first in Section 4.3. By the same token, the four daily GFS initialisations are applied in Section 4.4 where coefficient estimation is instead recursive and adaptive. Also, the focus on forecast leadtime resolution is relaxed in order to recast the dataset underlying the conclusion given in Section 5, and for the sake of a more summarising (clear-cut) comparison of model candidates within an alternative framework for coefficient estimation than the cross validation, cf. e.g. Efron and Gong (1983), employed in Section 4.3.

In this section explanatory value of the GFS-derived quantities summarised in Table II is quantified in terms of wind power forecast improvement for the three sites studied. Nielsen *et al.* (2001, 2012) describe the Wind Power Prediction Tool (WPPT) to estimate the collective electric power production of wind farm turbines, transforming kinetic energy in the moving air into electricity, for use in operational planning of wind energy on a daily basis. The stochastic simulation process in the WPPT involves wind direction dependent power curve modelling and the interested reader is referred to e.g. Lydia *et al.* (2014) for a recent overview of the topic. As Orths and Eriksen (2012) remark in their description of the world's largest day ahead market for trading power, the relevant forecast leadtime span for wind energy scheduling is 12–36 hours ahead of the auction deadline. In a realtime setting the four daily GFS forecasts are available with a 5 hours delay and this is accounted for in the statistical modelling underlying results in Table III, showing relative improvement in WPPT forecasts issued each hour, based on the GFS-derived quantity sets underlying models 2, 4, 6 and 7 of Table II, respectively.

By virtue of the pronounced differences in the rationales underlying the two employed evaluation strategies, results of Figure 6 and Table III indicate that modelled lifted index may be key to understanding formulation deficiencies in the GFS,

Table III. Performance of wind power forecasts by the WPPT, adapted to include the four predictor sets of Table II models 2, 4, 6 and 7 (referred to as *post* additional modelling), and the corresponding change Δ relative to performance of the standard WPPT setup for GFS (*prior* to additional modelling) relying on NWP data for wind speed and wind direction alone. Performance evaluation is based on forecast leadtime between hours 12 to 36 ahead of initialisation for year 2 data, cf. Table I. Horns Rev is abbreviated HR and has capacity 160 MW (80 wind turbines), Rejsby Hede is RH with capacity 23.4 MW (39), and Stor-Rotliden is denoted SR and has 77.8 MW (40). RMSE and MAE are given as percentage of respective wind farm capacities. N.b. in the standard WPPT setup input from multiple NWP models is employed; the pure GFS-based setup performs inferior by 10–15 %.

Site	Model #	RMSE [%] Prior	Post	Δ	MAE [%] Prior	Post	Δ
HR	2	16.50	16.14	−0.36	11.41	11.18	−0.23
	4		16.08	−0.42		11.14	−0.27
	6		16.23	−0.27		11.24	−0.17
	7		16.23	−0.27		11.24	−0.17
RH	2	12.39	11.97	−0.42	8.51	8.25	−0.26
	4		11.92	−0.47		8.21	−0.30
	6		12.05	−0.34		8.29	−0.22
	7		12.01	−0.38		8.25	−0.26
SR	2	15.81	16.12	+0.31	11.66	11.90	+0.24
	4		15.50	−0.31		11.48	−0.18
	6		15.96	+0.15		11.85	+0.19
	7		15.54	−0.27		11.52	−0.14

as this variable contains systematic information to consistently reduce advection forecast error based on two years of hourly data. Note, for this site also, that the inclusion of predictor terms involving temperature and friction velocity, both of which exhibit severe collinearity with LLS wind speed cf. Figure 7, results in forecast performance reduction relative to the performance of the WPPT setup for GFS input. This prompts a review of the fundamental difference between model coefficient estimation techniques underlying results of Figure 6 and Table III. The sharp division of fixed training and test datasets employed to arrive at the results of Figure 6 is – as mentioned in the first paragraph of this section – replaced by recursive and adaptive coefficient estimation in the WPPT. *Recursive* in the sense that coefficient estimates are updated in a computationally efficient manner as wind farm measurements and forecast data becomes available,

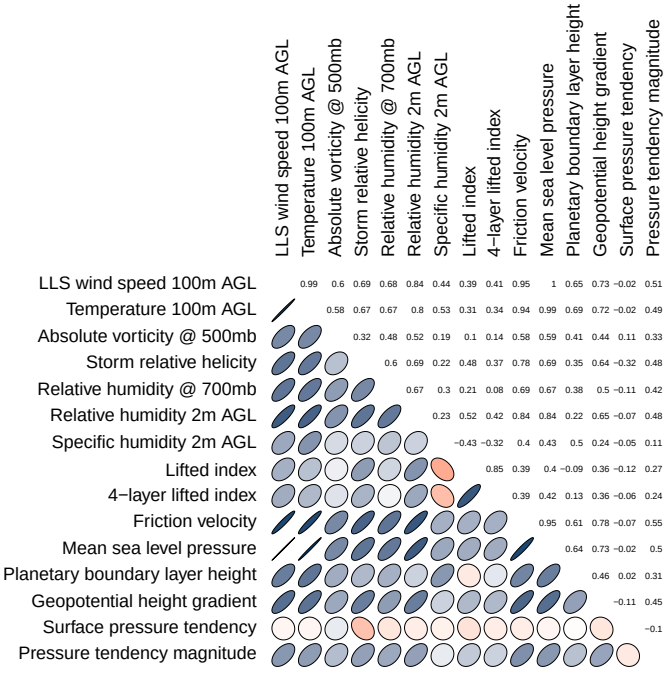


Figure 7. GFS intervariable PCC matrix for the inland site, Stor-Rotliden, post multiplication with LLS wind speed, cf. Figure 2b.

and *adaptive* due to exponentially decaying weight on previous data points as a function of the temporal distance to the current time stamp. In the WPPT configuration employed here, coefficient estimates for all models are based on fewer data points than in the Section 4.3 approach, and namely the poorly performing Model 2 and Model 6 are subject to strong collinearity with the standalone LLS wind speed predictor, which in turn renders the coefficient estimation unreliable cf. Eqs. (5–6). While all wind direction inclusive models perform similarly for Rejsby Hede, cf. Figure 6b, all GFS-derived quantities improve wind power forecasts for this coastal site relative to the standard WPPT configuration, which can roughly be viewed as representing Model 3 (the one with only LLS wind speed and wind direction). This may indicate strong seasonal dependence of the model coefficients which is accommodated for in the WPPT, though thwarted in the full-year model training dataset determining the model coefficient estimates underlying subsequent-year model predictions in Figure 6.

5. Conclusion

The purpose of a model is to provide a basis for understanding data and, namely in a weather forecasting context, to predict future measurements. The feasibility of various approximations to the governing equations for atmospheric dynamics, numerical discretisations for dynamical cores, and physical process schemes have been investigated since the dawn of the scientific discipline of computer-aided NWP 65 years ago, and new research to improve the status quo is continually carried out. As a tool for supporting this endeavour, the work presented in this article offers a methodology for indirectly diagnosing weather and climate model deficiency and improving prediction accuracy.

Provided assumptions 1) and 2) below are somewhat satisfied, the method of Section 3 holds for any model based on a set of coupled hyperbolic partial differential equations. That is, the practical implications of mathematical relations for prognostic and diagnostic variables can be explored based on the methodology framed in sections 3 and 4. In particular, inference on systematic model variable coupling to an observed quantity of interest is facilitated, and quantification of explanatory value for individual predictor candidates is discussed. However, the applied univariate modelling framework is subject to the following assumptions: 1) A physically reasonable general linear model relation between an observed response and simulated predictor candidates can be established, 2) approximate Gaussianity, homoscedasticity, and vanishing autocorrelation can be assumed for the statistical model error. A careful analysis of the extent to which 2) holds constitutes an essential component in such studies. For nonnormal response characteristics alternative statistical modelling techniques may be more suitable.

With access to the source code practically all weather and climate model aspects can be investigated in the quest for model consistency with atmospheric measurements of interest. Conducting the study on NWP model forecasts rather than reanalysis data exposes leadtime-dependent features – such as the here uncovered departure between explanatory value of lifted index vs. temperature for the inland site as a function of leadtime – that would otherwise not be disclosed, hence adding a temporal dimension to the model deficiency analysis. Thus, one gets a peek inside the numerics engine of a complex system of discretised partial differential equations, for which less-obvious numerical scheme and formulation deficiencies can be hard to detect.

The study suggests that statistical modelling of forecasted lifted index and wind direction, by the approach described in Section 3 and executed in Section 4, contributes additional accuracy to GFS predictions of wind speed in all terrain types. For the offshore and coastal sites excluding the lifted index predictor achieves equivalent performance cf. Figure 6, though it is important to note that lifted index inclusion does not deteriorate the out-of-sample prediction at any leadtime. This is, however, not the case for the coastal site when excluding the wind direction predictor terms, leading to inferior performance of including lifted index and temperature relative to the pure LLS wind speed model. That wind direction in NWP models has systematic influence on wind speed forecast error is perhaps not surprising. In contrast, the fact that the NWP modelled lifted index contributes $\sim 0.1\,\mathrm{m\,s^{-1}}$ better MAE and RMSE than the pure wind direction model for complex terrain data is more interesting, namely when considering the equivalent performance of surface temperature for short leadtimes deteriorating for forecast hours beyond $\sim$12 hours ahead. However, the wind power forecast results of Table III substantiate the explanatory value of the lifted index for all sites and hence hint a strong seasonal model coefficient dependency, namely for the offshore and coastal sites, both of which are in proximity to the North Sea.

The lifted index, first proposed by Galway (1956) as a predictor of latent instability (severe weather), is traditionally defined as the temperature difference between the NWP modelled temperature at the 500 mb level minus the ditto simulated surface layer midpoint temperature adiabatically raised to this pressure level. The relatively significant performance impact from including this predictor in Eq. (1) could hint moisture-dependent bias in the GFS simulation of surface layer winds, cf. e.g. Bénard *et al.* (2000), which in turn points toward inadequate surface and planetary boundary layer parameterisations and/or land surface scheme.

The high PCC and similar RMSE performance offshore, relative to the coastal and inland sites, is via the theoretical link between these accuracy metrics *approximately* shown to derive from the more volatile weather conditions in the North Sea. However, the discrepancy may in part be attributed inferior representation of the influence of the ocean surface on atmospheric flow. I.e. if the simulation of large scale weather systems is fairly well temporally correlated with observed dynamics (as the high PCC suggests), the severe wind speed error magnitude penalty expressed by the MAE/RMSE relative to the PCC, compared to the other sites, could adhere to NWP model lower boundary condition issues. Hence, the authors speculate that efforts toward improving the simulation of offshore surface layer dynamics in the GFS would be feasible for efficient offshore wind energy integration in the power grid, as well as for other applications depending critically on wind speed forecast accuracy at sea.

Acknowledgements

The study was funded by the Danish transmission system operator *Energinet.dk*, within the framework of public service obligation project *PSO 10464*, to promote environmentally friendly energy. The analysis and visualisation of results are performed using R (2013), Figure 1 is created using NCL (2014). Access to wind farm data was kindly provided by the Swedish power company *Vattenfall*. The authors thank Stig Mortensen and Sven Thomsen of ENFOR A/S for discussions on statistical modelling challenges, as well as Mark Kelly and Jake Badger of DTU Wind Energy for input on the interpretation of results.

References

Akaike H. 1974. *A new look at the statistical model identification.* IEEE Transactions on Automatic Control, 19, pp. 716-723.

Alfaro M, Huelsenbeck J. 2006. *Comparative performance of Bayesian and AIC-based measures of phylogenetic model uncertainty.* Systematic Biology, 55, pp. 89-96.

Bahadur RR. 1958. *Examples of inconsistency of maximum likelihood estimates.* Sankhyā Ser A, 20, pp. 207-210.

Bénard P, Marki A, Neytchev PN, Prtenjak MT. 2000. *Stabilization of nonlinear vertical diffusion schemes in the context of NWP models.* Monthly Weather Review, 128, pp. 1937-1948.

Bénard P. 2003. *Stability of semi-implicit and iterative centered-implicit time discretizations for various equation systems used in NWP.* Monthly Weather Review, 131, pp. 2479-2491.

Birgé L. 2006. *Model selection via testing: an alternative to (penalized) maximum likelihood estimators.* Annales de l'Institute Henri Poincaré, Probabilités et Statistiques, 42, pp. 273-325.

Bulteel K, Wilderjans TF, Tuerlinckx F, Ceulemans E. 2013. *CHull as an alternative to AIC and BIC in the context of mixtures of factor analyzers.* Behavior Research Methods, 45, pp. 782-791.

Cassola F, Burlando Massimiliano. 2012. *Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output.* Applied Energy, 99, pp. 154-166.

Charney JG. 1947. *The dynamics of long waves in a barotropic westerly current.* Journal of Meteorology, 4, pp. 135-162.

Charney JG. 1948. *On the scale of atmospheric motions.* Geofysiske Publikasjoner, 17, pp. 3-17.

Charney JG. 1949. *On a physical basis for numerical prediction of large-scale motions in the atmosphere.* Journal of Meteorology, 6, pp. 371-385.

Cheng WYY, Steenburgh WJ. 2007. *Strengths and weaknesses of MOS, running-mean bias removal, and Kalman filter techniques for improving model forecasts over the western United States.* Weather and Forecasting, 22, pp. 1304-1318.

Courant R, Friedrichs K, Lewy H. 1928. *Über die partiellen differenzengleichungen der mathematischen physik.* Mathematische Annalen, 100, pp. 32-74.

Courant R, Friedrichs K, Lewy H. 1967. *On the partial difference equations of mathematical physics.* IBM Journal of Research and Development, 11, pp. 215-234.

Dee DP. 2005. *Bias and data assimilation.* Quarterly Journal of the Royal Meteorological Society, 131, pp. 3323-3343.

Dee DP, Da Silva AM. 1998. *Data assimilation in the presence of forecast bias.* Quarterly Journal of the Royal Meteorological Society, 124, pp. 269-295.

Efron B, Gong G. 1983. *A leisurely look at the bootstrap, the jackknife, and cross-validation.* American Statistician, 37, pp. 36-48.

Engel C, Ebert EE. 2007. *Performance of hourly operational consensus forecasts (OCFs) in the Australian region.* Weather and Forecasting, 22, pp. 1345-1359.

Engel C, Ebert EE. 2012. *Gridded Operational Consensus Forecasts of 2-m Temperature over Australia.* Weather and Forecasting, 27, pp. 301-322.

Fritsch JM, Hilliker J, Ross J, Vislocky RL. 2000. *Model consensus.* Weather and Forecasting, 15, pp. 571-582.

Galway JG. 1956. *The lifted index as an indicator of latent instability.* Bulletin of the American Meteorological Society, 37, pp. 528-529.

Glahn HR, Lowry DA. 1972. *The use of Model Output Statistics (MOS) in objective weather forecasting.* Journal of Applied Meteorology, 11, pp. 1203-1211.

Gneiting T, Raftery AE, Westveld III AH, Goldman T. 2005. *Calibrated probabilistic forecasting using ensemble Model Output Statistics and minimum CRPS estimation.* Monthly Weather Review, 133, pp. 1098-1118.

Hacker JP, Rife DL. 2007. *A practical approach to sequential estimation of systematic error on near-surface mesoscale grids.* Weather and Forecasting, 22, pp. 1257-1273.

Hart KA, Steenburgh WJ, Onton DJ, Siffert AJ. 2004. *An evaluation of mesoscale-model-based Model Output Statistics (MOS) during the 2002 Olympic and Paralympic Winter Games.* Weather and Forecasting, 19, pp. 200-218.

Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction.* (2nd edition). Springer (2009).

Hayashi F. *Econometrics*. Princeton University Press (2000).

Holton JR, Hakim GJ. *Dynamic Meteorology.* (5th edition). Elsevier (2013). p. 369.

Jones RH. 2011. *Bayesian information criterion for longitudinal and clustered data.* Statistics in Medicine, 30, pp. 3050-3056.

Jónsson T, Pinson P, Nielsen HAa, Madsen H. 2014. *Exponential smoothing approaches for prediction in real-time electricity markets.* Energies, 7, pp. 3710-3732.

Judd K, Reynolds CA, Rosmond TE, Smith LA. 2008. *The geometry of model error.* Journal of the Atmospheric Sciences, 65, pp. 1749-1772.

Kalman RE. 1960. *A new approach to linear filtering and prediction problems.* Journal of Basic Engineering, 82, pp. 35-45.

Kanamitsu M. 1989. *Description of the NMC global data assimilation and forecast system.* Weather and Forecasting, 4, pp. 335-342.

Kuha J. 2004. *AIC and BIC: Comparison of assumptions and performance.* Sociological Methods and Research, 33, pp. 188-229.

Leith CE. 1974. *Theoretical skill of Monte-Carlo forecasts.* Monthly Weather Review, 102, pp. 409-418.

Lydia M, Kumar SS, Selvakumar AI, Kumar GEP. 2014. *A comprehensive review on wind turbine power curve modeling techniques.* Renewable and Sustainable Energy Reviews, 30, pp. 452-460.

Madsen H, Pinson P, Kariniotakis G, Nielsen HAa, Nielsen TS. 2005. *Standardizing the performance evaluation of short-term wind power prediction models.* Wind Engineering, 29, pp. 475-489.

Madsen H, Thyregod P. *Introduction to General and Generalized Linear Models*. CRC Press (2010).

Messner JW, Zeileis A, Broecker J, Mayr GJ. 2013. *Probabilistic wind power forecasts with an inverse power curve transformation and censored regression.* Wind Energy, 17, pp. 1753-1766.

Monin AS. 1970. *The atmosperic boundary layer.* Annual Review of Fluid Mechanics, 2, pp. 225-250.

Murphy AH. 1988. *Skill scores based on the mean square error and their relationships to the correlation coefficient.* Monthly Weather Review, 116, pp. 2417-2424.

Müller MD. 2011. *Effects of model resolution and statistical postprocessing on shelter temperature and wind forecasts.* Journal of Applied Meteorology and Climatology, 50, pp. 1627-1636.

Nielsen TS, Nielsen HAa, Madsen H. 2001. *Prediction of wind power using time-varying coefficient-functions.* Proceedings of the 15th IFAC World Congress on Automatic Control, Barcelona, Spain, 21-26 July 2002.

Nielsen TS, Nielsen HAa, Madsen H. 2012. *An overview of wind power forecast types and their use in large-scale integration of wind power.* Proceedings of the 10th International Workshop on Large-Scale Integration of Wind Power into Power Systems, Aarhus, Denmark, 25-26 October 2011.

Olsson U. *Generalized Linear Models: An Applied Approach*. Studentlitteratur (2001).

Orths AG, Eriksen PB. *Wind Power in Power Systems.* Wiley (2012). pp. 531-532.

Pawitan Y. *In All Likelihood: Statistical Modelling and Inference Using Likelihood.* Oxford Science Publications, 2008 corrected edition (2001).

Pinson P. 2012. *Adaptive calibration of (u,v)-wind ensemble forecasts.* Quarterly Journal of the Royal Meteorological Society, 138, pp. 1273-1284.

Ranaboldo M, Giebel G, Codina B. 2013. *Implementation of a Model Output Statistics based on meteorological variable screening for short-term wind power forecast.* Wind Energy, 16, pp. 811-826.

Schwarz G. 1978. *Estimating the dimension of a model.* The Annals of Statistics, 6, pp. 461-464.

Stensrud DJ, Yussouf N. 2005. *Bias-corrected short-range ensemble forecasts of near surface variables.* Meteorological Applications, 12, pp. 217-230.

Sweeney C, Lynch P. 2011. *Adaptive post-processing of short-term wind forecasts for energy applications.* Wind Energy, 14, pp. 317-325.

Sweeney CP, Lynch P, Nolan P. 2013. *Reducing errors of wind speed forecasts by an optimal combination of post-processing methods.* Meteorological Applications, 20, pp. 32-40.

Termonia P, Deckmyn A. 2007. *Model-inspired predictors for model output statistics (MOS).* Monthly Weather Review, 135, pp. 3496-3505.

Thorarinsdottir TL, Gneiting T. 2010. *Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression.* Journal of the Royal Statistical Society, A173, pp. 371-388.

Valentine TJ. 1969. *A note on multicollinearity.* Australian Economic Papers, 8, pp. 99-105.

Vrieze SR. 2012. *Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion.* Psychological Methods, 17, pp. 228-243.

Warner TT. *Numerical Weather and Climate Prediction*. Cambridge University Press (2011).

Wilks DS. *Statistical Methods in the Atmospheric Sciences.* (3rd edition). Academic Press (2011).

Wilks DS. 2014. *Multivariate ensemble Model Output Statistics using empirical copulas.* Quarterly Journal of the Royal Meteorological Society, Volume ??, pp. ????.

Williams MN, Grajales CAG, Kurkiewicz D. 2013. *Assumptions of multiple regression: Correcting two misconceptions.* Practical Assessment, Research and Evaluation, 18, pp. 1-14.

Woodcock F, Engel C. 2005. *Operational consensus forecasts.* Weather and Forecasting, 20, pp. 101-111.

Yuan K-H, Bentler PM. 1999. *On normal theory and associated test statistics in covariance structure analysis under two classes of nonnormal distributions.* Statistica Sinica, 9, pp. 831-853.

Yuan Z, Wu J, Cheng X, Jian M. 2008. *The derivation of a numerical diagnostic model for the forcing of the geopotential.* Quarterly Journal of the Royal Meteorological Society, 134, pp. 2067-2078.

Zucchini W. 2000. *An introduction to model selection.* Journal of Mathematical Psychology, 44, pp. 41-61.

R Core Team. *R: A language and environment for statistical computing* [Software]. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/ (2013).

The NCAR Command Language (Version 6.2.0) [Software]. Boulder, Colorado: UCAR/NCAR/CISL/VETS. http://dx.doi.org/10.5065/D6WD3XH5 (2014).

Bibliography

[1] Noah Francis. Predicting Sudden Changes In Wind Power Generation. *North American Windpower*, 5:58–60, 2008.

[2] A. Sfetsos. A Novel Approach for the Forecasting of Mean Hourly Wind Speed Time Series. *Renewable Energy*, 27:163–174, 2002.

[3] Gregor Giebel, Richard Brownsword, George Kariniotakis, Michael Denhard, and Caroline Draxl. *The State of the Art in Short-Term Prediction of Wind Power: A Literature Overview*, chapter Upscaling and spatio-temporal correlations, page 60. In Giebel et al. [8], 2 edition, 2011.

[4] A. M. Foley, P. G. Leahy, and E. J. McKeogh. Wind Power Forecasting & Prediction Methods. *9th International Conference on Environment and Electrical Engineering, Prague, Czech Republic, May 16 19*, pages 61–64, 2010.

[5] Aoife M. Foley, Paul G. Leahy, Antonino Marvuglia, and Eamon J. McKeogh. Current methods and advances in forecasting of wind power generation. *Renewable Energy*, 37:1–8, 2012.

[6] Gregor Giebel, Richard Brownsword, George Kariniotakis, Michael Denhard, and Caroline Draxl. *The State of the Art in Short-Term Prediction of Wind Power: A Literature Overview*, chapter Introduction, page 11. In Giebel et al. [8], 2 edition, 2011.

[7] Gregor Giebel, Richard Brownsword, George Kariniotakis, Michael Denhard, and Caroline Draxl. *The State of the Art in Short-Term Prediction of Wind Power: A Literature Overview*, chapter Time series models, page 26. In Giebel et al. [8], 2 edition, 2011.

[8] Gregor Giebel, Richard Brownsword, George Kariniotakis, Michael Denhard, and Caroline Draxl. *The State of the Art in Short-Term Prediction of Wind Power: A Literature Overview, 2nd Edition.* Deliverable of the ANEMOS.plus and SafeWind research and development projects funded by the European Commision, 2011.

[9] Torben Skov Nielsen, Henrik Madsen, and Henrik Aalborg Nielsen. Prediction of Wind Power Using Time-varying Coefficient-functions. *Proceedings of the 15th IFAC World Congress on Automatic Control*, pages 1–6, 2002.

[10] ENFOR A/S web page. Accessed 26 June 2014. URL http://www.enfor.dk/products/wppt.html.

[11] Matthias. Lange and Ulrich Focken. *Physical Approach to Short-term Wind Power Prediction.* Springer-Verlag Berlin Heidelberg, 2006.

[12] Energy & Meteo Systems web page. Accessed 26 June 2014. URL http://www.energymeteo.com/en/services/previento.php.

[13] Vilhelm Bjerknes. Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. *Meteorologische Zeitschrift*, 21:1–7, 1904.

[14] [13] translated to English and edited by Ester Volken and Stefan Brönnimann. The problem of weather prediction, considered from the viewpoints of mechanics and physics. *Meteorologische Zeitschrift*, 18:663–667, 2009.

[15] Gabriele Gramelsberger. Conceiving Meteorology as the exact science of the atmosphere: Vilhelm Bjerknes's paper of 1904 as a milestone. *Meteorologische Zeitschrift*, 18:669–673, 2009.

[16] Leonhard Euler. Principes généraux du mouvement des fluides. *Mémoires de l'académie des sciences de Berlin*, 11:274–315, 1757.

[17] O. Darrigol and D. Frisch. From Newton's mechanics to Euler's equations. *Physica D – Nonlinear Phenomena*, 237:1855–1869, 2008.

[18] Warren M. Washington. Computer modeling the twentieth- and twenty-first-century climate. *Proceedings of the American Philosophical Society*, 150:414–427, 2006.

[19] Lewis Fry Richardson. *Weather Prediction by Numerical Process*. Cambridge University Press, 1922.

[20] R. Courant, K. Friedrichs, and H. Lewy. Über die partiellen Differenzengleichungen der mathematischen Physik. *Mathematische Annalen*, 100:32–74, 1928.

[21] R. Courant, K. Friedrichs, and H. Lewy. On the Partial Difference Equations of Mathematical Physics. *IBM Journal of Research and Development*, 11:215–234, 1967.

[22] J. G. Charney. The dynamics of long waves in a baroclinic westerly current. *Journal of Meteorology*, 4:135–162, 1947.

[23] E. T. Eady. Long waves and cyclone waves. *Tellus*, 1:33–52, 1949.

[24] J. G. Charney. On the scale of atmospheric motions. *Geofysiske Publikasjoner*, 17:3–17, 1948.

[25] J. G. Charney, R. Fjörtoft, and J. von Neumann. Numerical Integration of the Barotropic Vorticity Equation. *Tellus*, 2:237–254, 1950.

[26] J. G. Charney and N. A. Phillips. Numerical Integration of the Quasi-Geostrophic Equations for Barotropic and Simple Baroclinic Flows. *Journal of Meteorology*, 10:71–99, 1953.

[27] J. Charney. The Use of the Primitive Equations of Motion in Numerical Prediction. *Tellus*, 7:22–26, 1955.

[28] J. Charney. Integration of the primitive and balance equations. *Proceedings of the International symposium on numerical weather prediction, Japan Meteorological Agency, Tokyo, Japan*, pages 131–152, 1962.

[29] A. Staniforth. Developing Efficient Unified Nonhydrostatic Models. In *Proceedings of the 50th Anniversary of Numerical Weather Prediction Commemorative Symposium, Potsdam 9-10 March, Germany, Deutsche Meteorologische Gesellschaft e.v.*, pages 185–200, 2001.

[30] L. Xu, S. Raman, and R. V. Madala. A Review of Non-Hydrostatic Numerical Models for the Atmosphere. In *World Congress of Nonlinear Analysis '92*, volume 1-4, pages 3595–3609, 1995.

[31] Z. I. Janjic and S. Gerrity Jr., J. P.and Nickovic. An Alternative Approach to Nonhydrostatic Modeling. *Monthly Weather Review*, 129:1164–1178, 2001.

[32] Kazuo Saito, Tsukasa Fujita, Yoshinori Yamada, Jun-ichi Ishida, Yukihiro Kumagai, Kohei Aranami, Shiro Ohmori, Ryoji Nagasawa, Sagai Kumagai, Chiashi Muroi, Teruyuki Kato, and Hisaki Eito. The Operational JMA Nonhydrostatic Mesoscale Model. *Monthly Weather Review*, 134:1266–1298, 2006.

[33] Andrew Staniforth and Nigel Wood. Aspects of the Dynamical Core of a Nonhydrostatic, Deep-Atmosphere, Unified Weather and Climate-Prediction Model. *Journal of Computational Physics*, 227:3445–3464, 2008.

[34] Saito, Kazuo. Nonhydrostatic Atmospheric Models and Operational Development at JMA. *Journal of the Meteorological Society of Japan*, 85B:271–304, 2007.

[35] Isidoro Orlanski. Rational Subdivision of Scales for Atmospheric Processes. *Bulletin of the American Meteorological Society*, 56:527–530, 1975.

[36] Gary Lackmann. *Midlatitude Synoptic Meteorology: Dynamics, Analysis and Forecasting*, chapter Introduction, Background, and Basics, page 3. In Lackmann [122], august 2012 reprinted with corrections edition, 2012.

[37] Da-Lin Zhang, Shunli Zhang, and Scott J. Weaver. Low-Level Jets over the Mid-Atlantic States: Warm-Season Climatology and a Case Study. *Journal of Applied Meteorology and Climatology*, 45:194–209, 2006.

[38] Adam J. Deppe, William A. Gallus Jr., and Eugene S. Takle. A WRF Ensemble for Improved Wind Speed Forecasts at Turbine Height. *Weather and Forecasting*, 28:212–228, 2013.

[39] A. Kolmogorov. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Comptes rendus de l'Académie des sciences de l'URSS*, 30:301–305, 1941.

[40] J. G. Charney. Geostrophic Turbulence. *Journal of the Atmospheric Sciences*, 28: 1087–1095, 1971.

[41] K. S. Gage and G. D. Nastrom. On the spectrum of atmospheric velocity fluctuations seen by MST/ST radar and their interpretation. *Radio Science*, 20:1339–1347, 1985.

[42] T. E. VanZandt. A Universal Spectrum of Buyancy Waves in the Atmosphere. *Geophysical Research Letters*, 9:575–578, 1982.

[43] E. M. Dewan. Stratospheric Wave Spectra Resembling Turbulence. *Science*, 204: 832–835, 1979.

[44] K. S. Gage. Evidence for a $K^{-5/3}$ Law Inertial Range in Mesoscale 2-Dimensional Turbulence. *Journal of the Atmospheric Sciences*, 36:1950–1954, 1979.

[45] D. K. Lilly. Stratified Turbulence and the Mesoscale Variability of the Atmosphere. *Journal of the Atmospheric Sciences*, 40:749–761, 1983.

[46] J. Y. N. Cho, R. E. Newell, and J. D. Barrick. Horizontal wavenumber spectra of winds, temperature, and trace gases during the Pacific Exploratory Missions: 2. Gravity waves, quasi-two-dimensional turbulence, and vortical modes. *Journal of Geophysical Research - Atmospheres*, 104:16297–16308, 1999.

[47] E. N. Lorenz. The Predictability of a Flow which Possesses Many Scales of Motion. *Tellus*, 21:289–307, 1969.

[48] Richard A. Anthes, Ying-Hwa Kuo, David P. Baumhefner, Ronald M. Errico, and Thomas W. Bettge. Predictability of Mesoscale Atmospheric Motions. *Advances in Geophysics*, 28:159–202, 1985.

[49] Richard A. Anthes. *Mesoscale Meteorology and Forecasting*, chapter The General Question of Predictability, pages 636–656. American Meteorological Society, 1986.

[50] T. Vukicevic and R. M. Errico. The Influence of Artificial and Physical Factors upon Predictability Estimates using a Complex Limited-Area Model. *Monthly Weather Review*, 118:1460–1482, 1990.

[51] J. Paegle, Q. Yang, and M. Wang. Predictability in Limited Area and Global Models. *Meteorology and Atmospheric Physics*, 63:53–69, 1997.

[52] Thomas T. Warner, Ralph A. Peterson, and Russel E. Treadon. A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bulletin of the American Meteorological Society*, 78(11):2599–2617, 1997. URL http://dx.doi.org/10.1175/1520-0477(1997)078<2599:ATOLBC>2.0.CO;2.

[53] John C. Wyngaard. Toward numerical modeling in the "terra incognita". *Journal of the Atmospheric Sciences*, 61:1816–1826, 2004.

[54] William C. Skamarock. Evaluating mesoscale NWP models using kinetic energy spectra. *Monthly Weather Review*, 132:3019–3032, 2004.

[55] Erik Lindborg. Can the atmospheric kinetic energy spectrum be explained by two-dimensional turbulence? *Journal of Fluid Mechanics*, 388:259–288, 1999.

[56] J. N. Koshyk and K. Hamilton. The horizontal kinetic energy spectrum and spectral budget simulated by a high-resolution troposphere-stratosphere-mesosphere GCM. *Journal of the Atmospheric Sciences*, 58:329–348, 2001.

[57] E. Lindborg and J. Y. N. Cho. Horizontal velocity structure functions in the upper troposphere and lower stratosphere 2. Theoretical considerations. *Journal of Geophysical Research - Atmospheres*, 106:10233–10241, 2001.

[58] William C. Skamarock, Joseph B. Klemp, Jimy Dudhia, David O. Gill, Dale M. Barker, Wei Wang, and Jordan G. Powers. *A Description of the Advanced Research WRF Version 2*. NCAR/TN468+STR, *NCAR Technical Note*, National Center for Atmospheric Research, Boulder CO, USA, 2005.

[59] The Weather Research and Forecasting model. URL http://www.wrf-model.org/index.php. [Accessed online 17 March 2015].

[60] William C. Skamarock and Joseph B. Klemp. A Time-Split Nonhydrostatic Atmospheric Model for Weather Research and Forecasting Applications. *Journal of Computational Physics*, 227:3465–3485, 2008.

[61] William C. Skamarock, Joseph B. Klemp, Jimy Dudhia, David O. Gill, Dale M. Barker, Michael G. Duda, Xiang-Yu Huang, Wei Wang, and Jordan G. Powers. *A Description of the Advanced Research WRF Version 3*. NCAR/TN475+STR, *NCAR Technical Note*, National Center for Atmospheric Research, Boulder CO, USA, 2008.

[62] A. J. Simmons, D. M. Burridge, M. Jarraud, C. Girard, and W. Wergen. The ECMWF medium-range prediction models: development of the numerical formulations and the impact of increased resolution. *Meteorology and Atmospheric Physics*, 40:28–60, 1989.

[63] M. Kanamitsu. Description of the NMC global data assimilation and forecast system. *Weather and Forecasting*, 4:335–342, 1989.

[64] M. J. P. Cullen. The Unified Forecast/Climate Model. *Meteorological Magazine*, 122:81–94, 1993.

[65] P. A. Jiménez, J. Dudhia, J. F. Gonzáles-Rouco, J. P. Montávez, E. García-Bustamante, J. Navarro, J. Vilà-Guerau de Arellano, and A. Muñoz Roldan. An evaluation of WRF's ability to reproduce the surface wind over complex terrain based on typical circulation patterns. *Journal of Geophysical Research - Atmospheres*, 118:7651–7669, 2013.

[66] Pedro A. Jiménez, J. Fidel Montávez, Gonzáles-Rouco, Juan P., Elena García-Bustamante, and Jorge Navarro. Climatology of wind patterns in the northeast of the Iberian Peninsula. *International Journal of Climatology*, 29:501–525, 2009.

[67] Kristian Horvath, Darko Koracin, Ramesh Vellore, Jinhua Jiang, and Radian Belu. Sub-Kilometer Dynamical Downscaling of Near-Surface Winds in Complex Terrain Using WRF and MM5 Mesoscale Models. *Journal of Geophysical Research*, 117, 2012.

[68] Andrea N. Hahmann, Claire L. Vincent, Alfredo Peña, Julia Lange, and Charlotte B. Hasager. Wind climate estimation using WRF model output: method and model sensitivities over the sea. *International Journal of Climatology*, 2014.

[69] René Laprise, Mundakkara Ravi Varma, Bertrand Denis, Daniel Caya, and Isztar Zawadzki. Predictability of a Nested Limited-Area Model. *Monthly Weather Review*, 128:4149–4154, 2000.

[70] D. Caya and R. Laprise. A semi-Lagrangian semi-implicit regional climate model: The Canadian RCM. *Monthly Weather Review*, 127:341–362, 1999.

[71] C. Chouinard, J. Mailhot, H. L. Mitchell, A. Staniforth, and R. Hogue. The Canadian Regional Data Assimilation System: Operational and research applications. *Monthly Weather Review*, 122:1306–1324, 1994.

[72] Lazar Lazić, Goran Pejanović, and Živković Momčilo. Wind forecasts for wind power generation using the Eta model. *Renewable Energy*, 35:1236–1243, 2010.

[73] F. Mesinger, Z. I. Janjić, S. Nicković, D. Gavrilov, and D. G. Deaven. The Step-Mountain Coordinate: Model Description and Performance for Cases of Alpine Lee Cyclogenesis and for a Case of an Appalachian Redevelopment. *Monthly Weather Review*, 116:1493–1518, 1988.

[74] Z. I. Janjić. The Step-Mountain Coordinate: Physical Package. *Monthly Weather Review*, 118:1429–1443, 1990.

[75] F. Mesinger, Z. I. Janjić, S. Nicković, D. Gavrilov, and D. G. Deaven. The Step-Mountain Coordinate: Further Developments of the Convection, Viscous Sublayers and Turbulence Closure Schemes. *Monthly Weather Review*, 122:927–945, 1994.

[76] Fedor Mesinger, Sin Chan Chou, Jorge L. Gomes, Dusan Jovic, Paulo Bastos, Josiane F. Bustamante, Lazar Lazic, André A. Lyra, Sandra Morelli, Ivan Ristic, and Katarina Veljovic. An upgraded version of the Eta model. *Meteorology and Atmospheric Physics*, 116:63–79, 2012.

[77] Daran L. Rife, Christopher A. Davis, Yubao Liu, and Thomas T. Warner. Predictability of Low-Level Winds by Mesoscale Meteorological Models. *Monthly Weather Review*, 132:2553–2569, 2004.

[78] Daran L. Rife and Christopher A. Davis. Verification of Temporal Variations in Mesoscale Numerical Wind Forecasts. *Monthly Weather Review*, 133:3368–3381, 2005.

[79] Brian A. Colle and Clifford F. Mass. Windstorms along the Western Side of the Washington Cascade Mountains. Part I: A High-Resolution Observational and Modeling Study of the 12 February 1995 Event. *Monthly Weather Review*, 126: 28–52, 1998.

[80] Clifford F. Mass, David Ovens, Ken Westrick, and Brian A. Colle. Does Increasing Horizontal Resolution Produce More Skillful Forecasts? *Bulletin of the American Meteorological Society*, 83:407–430, 2002.

[81] Tomonori Sato, Hiroaki Miura, Masaki Satoh, Yukari N. Takayabu, and Yuqing Wang. Diurnal Cycle of Precipitation in the Tropics Simulated in a Global Cloud-Resolving Model. *Journal of Climate*, 22:4809–4826, 2009.

[82] John H. Marsham, Nick S. Dixon, Luis Garcia-Carreras, Grenville M. S. Lister, Douglas J. Parker, Peter Knippertz, and Cathryn E. Birch. The Role of Moist Convection in the West African Monsoon System: Insights from Continental-Scale Convection-Permitting Simulations. *Geophysical Research Letters*, 40:1843–1849, 2013.

[83] C. E. Birch, D. J. Parker, J. H. Marsham, D. Copsey, and L. Garcia-Carreras. A Seamless Assessment of the Role of Convection in the Water Cycle of the West African Monsoon. *Journal of Geophysical Research*, 119:2890–2912, 2014.

[84] Jeffrey T. McQueen, Roland R. Draxler, and Glenn D. Rolph. Influence of Grid Size and Terrain Resolution on Wind-Field Predictions from an Operational Mesoscale Model. *Journal of Applied Meteorology*, 34:2166–2181, 1995.

[85] C. Davis, T. Warner, E. Astling, and J. Bowers. Development and Application of an Operational, Relocatable, Mesogamma-Scale Weather Analysis and Forecasting System. *Tellus Series A*, 51A:710–727, 1999.

[86] B. W. Buckley and L. M. Leslie. The Australian Boxing Day Storm of 1998 – Synoptic Description and Numerical Simulations. *Weather and Forecasting*, 15: 543–558, 2000.

[87] H. Madsen, P. Pinson, G. Kariniotakis, H. Aa. Nielsen, and T. S. Nielsen. Standardizing the performance evaluation of short-term wind power prediction models. *Wind Engineering*, 29:475–489, 2005.

[88] Xinxin Zhu and Marc. G. Genton. Short-Term Wind Speed Forecasting for Power System Operations. *International Statistical Review*, 80:2–23, 2012.

[89] H. C. Davies. A Lateral Boundary Formulation for Multi-Level Prediction Models. *Quarterly Journal of the Royal Meteorological Society*, 102:405–418, 1976.

[90] Mathias Lange. Analysis of the Uncertainty of Wind Power Predictions. *Ph.d. Dissertation, University of Oldenburg, Germany*, 2003.

[91] Matthias. Lange and Ulrich Focken. *Physical Approach to Short-term Wind Power Prediction*, chapter Relating the Forecast Error to Meteorological Situations, pages 135–165. In Lange and Focken [11], 2006.

[92] Claire Louise Vincent and Andrea Noemí Hahmann. Hour-scale wind fluctuations over the north Sea. In *European Wind Energy Conference and Exhibition*, pages 190–193, 2011.

[93] T. Kohonen. *Self-Organizing Maps*. Springer series in information sciences. Springer-Verlag, Berlin, 2001.

[94] Matteo Ranaboldo, Gregor Giebel, and Codina Bernat. Implementation of a Model Output Statistics based on meteorological variable screening for shortterm wind power forecast. *Wind Energy*, 16:811–826, 2013.

[95] H. R. Glahn and D. A. Lowry. The use of Model Output Statistics (MOS) in objective weather forecasting. *Journal og Applied Meteorology*, 11:1203–1211, 1972.

[96] Daniel S. Wilks. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, chapter Statistical Forecasting. Academic Press, 3 edition, 2011.

[97] R Core Team. R: A language and environment for statistical computing. Version 3.0.2. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

[98] The NCAR Command Language. Version 6.2.0. Boulder, Colorado: UCAR/NCAR/CISL/VETS. URL http://dx.doi.org/10.5065/D6WD3XH5.

[99] Nicholas Cutler, Merlinde Kay, Kieran Jacka, and Torben Skov Nielsen. Detecting, categorizing and forecasting large ramps in wind farm power output using meteorological observations and WPPT. *Wind Energy*, 10:453–470, 2007.

[100] Naomi Oreskes, Kristin Shrader-Frechette, and Kenneth Belitz. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science*, 263:641–646, 1994.

[101] Nicholas J. Cutler, Hugh R. Outhred, and Iain F. MacGill. Using nacelle-based wind speed observations to improve power curve modeling for wind power forecasting. *Wind Energy*, 15:245–258, 2012.

[102] Horns Rev wind farm information, . URL http://en.wikipedia.org/wiki/Horns_Rev. [Accessed 22 March 2015].

[103] Google Earth. URL https://www.google.com/earth/. [Accessed 19 March 2015].

[104] Stor-Rotliden wind farm information, . URL http://www.thewindpower.net/windfarm_maps_en_10483_stor-rotliden.php. [Accessed 22 March 2015].

[105] Horns Rev wind farm information, . URL http://www.lorc.dk/offshore-wind-farms-map/horns-rev-1. [Accessed 22 March 2015].

[106] Rejsby Hede wind farm information, . URL http://corporate.vattenfall.dk/om-os/vores-virksomhed/vattenfall-i-danmark/vores-vindmoller/rejsby-hede-vindmollepark. [Accessed 22 March 2015].

[107] Stor-Rotliden wind farm information, . URL http://powerplants.vattenfall.com/node/403. [Accessed 22 March 2015].

[108] Jesper Nielsen Nissen, Vattenfall. [Private communication].

[109] Terry Davies. Lateral boundary conditions for limited area models. *Quarterly Journal of the Royal Meteorological Society*, 140:185–196, 2014.

[110] T. T. Warner. Quality assurance in atmospheric modeling. *Bulletin of the American Meteorological Society*, 92(12):1601–1610, 2011. URL http://dx.doi.org/10.1175/BAMS-D-11-00054.1.

[111] James R. Holton and Gregory J. Hakim. *An Introduction to Dynamic Meteorology (5th ed.)*, chapter Basic Conservation Laws, pages 31–66. In Holton and Hakim [262], 2012.

[112] Martin Ehrendorfer. *Spectral Numerical Weather Prediction Models*, chapter The Spectral Method, pages 169–256. Society for Industrial and Applied Mathematics (SIAM), 2012.

[113] Sarah N. Collins, Robert S. James, Pallav Ray, Katherine Chen, Angie Lassman, and James Brownlee. *Climate Change and Regional/Local Responses*, chapter Grids in Numerical and Climate Models. InTech (open access), 2013.

[114] William C. Skamarock, Joseph B. Klemp, Jimy Dudhia, David O. Gill, Dale M. Barker, Michael G. Duda, Xiang-Yu Huang, Wei Wang, and Jordan G. Powers. *A Description of the Advanced Research WRF Version 3*, chapter Governing Equations, page 9. In Skamarock et al. [61], 2008.

[115] L. J. Wicker and W. C. Skamarock. Time-splitting methods for elastic models using forward time schemes. *Monthly Weather Review*, 130:2088–2097, 2002.

[116] William C. Skamarock, Joseph B. Klemp, Jimy Dudhia, David O. Gill, Dale M. Barker, Michael G. Duda, Xiang-Yu Huang, Wei Wang, and Jordan G. Powers. *A Description of the Advanced Research WRF Version 3*, chapter Model Discretization, page 21. In Skamarock et al. [61], 2008.

[117] A. Arakawa and V. R. Lamb. *Methods of Computational Physics*. Academic Press (AP), 1977.

[118] R. Laprise. The Euler Equations of motion with hydrostatic pressure as as independent variable. *Monthly Weather Review*, 120:197–207, 1992.

[119] William C. Skamarock, Joseph B. Klemp, Jimy Dudhia, David O. Gill, Dale M. Barker, Michael G. Duda, Xiang-Yu Huang, Wei Wang, and Jordan G. Powers. *A Description of the Advanced Research WRF Version 3*, chapter Nesting, page 59. In Skamarock et al. [61], 2008.

[120] Thomas Tomkins Warner. *Numerical Weather and Climate Prediction*, chapter Numerical solutions to the equations, pages 21–22. In Warner [263], 2010.

[121] MetEd web page; 6. Boundary Conditions and Initialization. Accessed 30 March 2015. URL `http://www.meted.ucar.edu/mesoprim/models/print.htm`.

[122] Gary Lackmann. *Midlatitude Synoptic Meteorology: Dynamics, Analysis and Forecasting*. American Meteorological Society, 45 Bacon Street, Boston, Massachusetts 02108, USA, reprinted with corrections edition, 2012.

[123] Huw C. Davies. On the Lateral Boundary Conditions for the Primitive Equations. *Journal of the Atmospheric Sciences*, 30:147–150, 1972.

[124] Shuzhan Ren and Theodore G. Shepherd. Lateral boundary contributions to wave-activity invariants and nonlinear stability theorems for balanced dynamics. *Journal of Fluid Mechanics*, 345:287–305, 1997.

[125] James R. Holton and Gregory J. Hakim. *An Introduction to Dynamic Meteorology (5th ed.)*, chapter The General Circulation, page 369. In Holton and Hakim [262], 2012.

[126] European Centre for Medium-Range Weather Forecasts. URL http://www.ecmwf.int. Research → Modelling and Prediction → Atmospheric physics [Accessed online 18 August 2014].

[127] Thomas Tomkins Warner. *Numerical Weather and Climate Prediction*, chapter Physical-process parameterizations, pages 119–170. In Warner [263], 2010.

[128] Thomas Tomkins Warner. *Numerical Weather and Climate Prediction*, chapter Modeling surface processes, pages 171–196. In Warner [263], 2010.

[129] Roger A. Pielke Sr. *Mesoscale Meteorological Modeling*, chapter 7 through 9, pages 164–280. Academic Press, 2 edition, 2002.

[130] Jared A. Lee, Walter C. Kolczynski, Tyler C. McCandless, and Sue Ellen Haupt. An Objective Method for Configuring and Down-Selecting an NWP Ensemble for Low-Level Wind Prediction. *Monthly Weather Review*, 140:2270–2286, 2012.

[131] Teddy Holt and Sethu Raman. A Review and Comparative Evaluation of Multilevel Boundary Layer Parameterizations for First-Order and Turbulent Kinetic Energy Closure Schemes. *Reviews of Geophysics*, 26:761–780, 1988.

[132] Alapaty, K. and Pleim, J. E. and Raman, S. and Devdutta, S. N. and Byun, D. W. Simulation of atmospheric boundary layer processes using local- and nonlocal-closure schemes. *Journaly of Applied Meteorology*, 36:214–233, 1997.

[133] Roland B. Stull. *An Introduction to Boundary Layer Meteorology*, pages 197–250. Springer, 1988.

[134] Joseph Boussinesq. Essai sur la théorie des eaux courantes. *Memoires présentés par divers savants à lAcadémie des Sciences*, 23:46, 1877.

[135] Roland B. Stull. Transilient Turbulence Theory. Part I: The Concept of Eddy-Mixing across Finite Distances. *Journal of the Atmospheric Sciences*, 41:3351–3367, 1984.

[136] I Troen and L. Mahrt. A simple model of the atmospheric boundary layer; sensitivity to surface evaporation. *Boundary Layer Meteorology*, 37:129–148, 1986.

[137] J. E. Pleim. A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *Journal of Applied Meteorology and Climatology*, 46:1383–1395, 2007.

[138] J. E. Pleim. A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: Application and evaluation in a mesoscale meteorological model. *Journal of Applied Meteorology and Climatology*, 46:1396–1409, 2007.

[139] Xiao-Ming Hu, John W. Nielsen-Gammon, and Fuqing Zhang. Evaluation of Three Planetary Boundary Layer Schemes in the WRF Model. *Journal of Applied Meteorology and Climatology*, 49:1831–1844, 2010.

[140] S-Y. Hong, Y. Noh, and J. Dudhia. A new vertical diffusion package with explicit treatment of entrainment processes. *Monthly Weather Review*, 134:2318–2341, 2006.

[141] P. Bougeault and P. Lacarrère. Parameterization of orography-induced turbulence in a mesobeta-scale model. *Monthly Weather Review*, 117:1872–1890, 1989.

[142] B. Xie, J. C. H. Fung, A. Chan, and A. Lau. Evaluation of nonlocal and local planetary boundary layer schemes in the WRF model. *Journal of Geophysical Research*, 117, 2012.

[143] Caroline Draxl, Andrea N. Hahmann, Alfredo Peña, and Gregor Giebel. Evaluating winds and vertical wind shear from Weather Research and Forecasting model forecasts using seven planetary boundary layer schemes. *Wind Energy*, 17:39–55, 2014.

[144] M. Nakanishi. Improvement of the Mellor-Yamada turbulence closure model based on large-eddy simulation data. *Boundary-Layer Meteorology*, 99:349–378, 2001.

[145] Y. Kitamura. Modifications to the Mellor-Yamada-Nakanishi-Niino (MYNN) model for the stable stratification case. *Journal of the Meteorological Society of Japan*, 88:857–864, 2010.

[146] Andrea Noemí Hahmann, DTU Wind Energy. [Private communication].

[147] Santos-Alamillos, F. J. and Pozo-Vázquez, D. and Ruiz-Arias, J. A., Lara-Fanego, V. and Tovar-Pescador, J. Analysis of WRF Model Wind Estimate Sensitivity to Physics Parameterization Choice and Terrain Representation in Andalusia (Southern Spain). *Journal of Applied Meteorology and Climatology*, 52:1592–1609, 2013.

[148] A. G. Orths and P. B. Eriksen. *Wind Power in Power Systems*, chapter Wind Power in the Danish Power System, page 536. John Wiley & Sons, Ltd., 2 edition, 2012.

[149] G. Shevlyakov and P. Smirnov. Robust estimation of the correlation coefficient: An attempt of survey. *Austrian Journal of Statistics*, 40:147–156, 2011.

[150] B. K. Fosdick and A. E. Raftery. Estimating the correlation in bivariate normal data with known variances and small sample sizes. *American Statistician*, 66: 34–41, 2012.

[151] Stor-Rotliden wind farm data. URL http://www.thewindpower.net/windfarm_maps_en_10483_stor-rotliden.php. [Accessed 18 March 2015].

[152] Bent Acher Johansen, Vattenfall. [Private communication].

[153] Orography map generator available from. URL http://www.acemaps.com/. [Accessed 19 March 2015].

[154] Visualization & Enabling Technologies Section (VETS). *The NCAR Command Language (Version 6.2.0) [Software]*. Boulder, Colorado: UCAR/NCAR/CISL/VETS, 2014. URL http://dx.doi.org/10.5065/D6WD3XH5.

[155] Moderate resolution Imaging Spectroradiometer satellite data. URL http://modis-land.gsfc.nasa.gov/. [Accessed 19 March 2015].

[156] United States Geological Survey Global 30 Arc-Second Elevation. URL https://lta.cr.usgs.gov/GTOPO30. [Accessed 11 April 2015].

[157] E. A. Aligo, A. Gallus Jr., and M. Segal. On the impact of WRF model vertical grid resolution on midwest summer rainfall forecasts. *Weather and Forecasting*, 24:575–594, 2009.

[158] Natacha B. Bernier and Stephanie Belair. High Horizontal and Vertical Resolution Limited-Area Model: Near-Surface and Wind Energy Forecast Applications. *Journal of Applied Meteorology and Climatology*, 51:1061–1078, 2012.

[159] Information on upgrades of the Global Forecast System (GFS). URL http://www.emc.ncep.noaa.gov/GFS/impl.php. [Accessed 22 March 2015].

[160] Morris L. Weisman, Joseph B. Klemp, and Richard Rotunno. Structure and Evolution of Numerically Simulated Squall Lines. *Journal of the Atmospheric Sciences*, 45:1990–2013, 1988.

[161] William C. Skamarock, Morris L. Weisman, and Joseph B. Klemp. Three-Dimensional Evolution of Simulated Long-Lived Squall Lines. *Journal of the Atmospheric Sciences*, 51:2563–2584, 1994.

[162] K. K. Droegemeier, G. Bassett, and M. Xue. Very high-resolution, uniform-grid simulations of deep convection on a massively parallel computer: Implications for small-scale predictability. In *10th Conference on Numerical Weather Prediction*, pages 376–379, 1994.

[163] Morris L. Weisman, William C. Skamarock, and Joseph B. Klemp. The Resolution Dependence of Explicitly Modeled Convective Systems. *Monthly Weather Review*, 125:527–548, 1997.

[164] L. R. Bernadet, L. D. Grasso, J. E. Nachamkin, C. A. Finley, and W. R. Cotton. Simulating Convective Events Using a High-Resolution Mesoscale Model. *Journal of Geophysical Research*, 105:14963–14982, 2000.

[165] J. S. Kain. The Kain-Fritsch convective parameterization: An update. *Journal of Applied Meteorology*, 43:170–181, 2004.

[166] G. Thompson, P. R. Field, R. M. Rasmussen, and W. D. Hall. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review*, 136: 5096–5115, 2008.

[167] Ó. Rögnvaldsson, J. W. Bao, H. Ágústsson, and H. Ólafsson. Downslope windstorm in Iceland — WRF/MM5 model comparison. *Atmospheric Chemistry and Physics*, 11:103–120, 2011.

[168] M. Rajeevan, A. Kesarkar, S. B. Thampi, T. N. Rao, B. Radhakrishna, and M. Rajasekhar. Sensitivity of WRF cloud microphysics to simulations of a severe thunderstorm event over Southeast India. *Annales Geophysicae*, 28:603–619, 2010.

[169] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*, chapter Overview of Supervised Learning, page 22. Springer Series in Statistics, 2 edition, 2009.

[170] R. E. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.

[171] H. Aa. Nielsen. *LFLM version 1.0, an S-PLUS/R library for locally weighted fitting of linear models.* Technical Report 22. Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark, 1997.

[172] Henrik Aalborg Nielsen, ENFOR A/S. [Private communication].

[173] A. G. Orths and P. B. Eriksen. *Wind Power in Power Systems*, chapter Wind Power in the Danish Power System, pages 531–532. John Wiley & Sons, Ltd., 2 edition, 2012.

[174] Jónsson, T. and Pinson, P. and Nielsen, H. Aa. and Madsen, H. Exponential smoothing approaches for prediction in real-time electricity markets. *Energies*, 7: 3710–3732, 2014.

[175] Elías Valur Hólm. Lecture notes on assimilation algorithms, 2008. URL http://old.ecmwf.int/newsevents/training/lecture_notes/pdf_files/ASSIM/Ass_algs.pdf. Teaching material from a data assimilation course offered by the European Centre for Medium-Range Weather Forecasts, Reading, UK. [Accessed 23-July-2014].

[176] F. Bouttier and P. Courtier. Data assimilation concepts and methods, 1999. URL http://old.ecmwf.int/newsevents/training/lecture_notes/pdf_files/ASSIM/Ass_cons.pdf. Teaching material from a data assimilation course offered by the European Centre for Medium-Range Weather Forecasts, Reading, UK. [Accessed 23-July-2014].

[177] Eugenia Kalnay. *Atmospheric modeling, data assimilation and predictability.* Cambridge University Press, 2003.

[178] Richard A. Anthes. Data Assimilation and Initialization of Hurricane Prediction Models. *Journal of the Atmospheric Sciences*, 31:702–719, 1974.

[179] James E. Hoke and Richard A. Anthes. The Initialization of Numerical Models by a Dynamic-Initialization Technique. *Monthly Weather Review*, 104:1551–1556, 1976.

[180] David R. Stauffer and Nelson L. Seaman. Use of Four-Dimensional Data Assimilation in a Limited-Area Mesoscale Model. Part I: Experiments with Synoptic-Scale Data. *Monthly Weather Review*, 118:1250–1277, 1990.

[181] Geir Evensen. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367, 2003.

[182] Zhaoxia Pu, Hailing Zhang, and Jeffrey Anderson. Ensemble Kalman filter assimilation of near-surface observations over complex terrain: comparison with 3DVAR for short-range forecasts. *Tellus Series A - Dynamic Meteorology and Oceanography*, 65, 2013.

[183] Eugenia Kalnay, Hong Li, Takamasa Miyoshi, Shu-Chih Yang, and Joaqim Ballabrera-Poy. 4-D-Var or ensemble Kalman filter? *Tellus Series A - Dynamic Meteorology and Oceanography*, 59:758–773, 2007.

[184] Meng Zhang, Fuqing Zhang, Xiang-Yu Huang, and Xin Zhang. Intercomparison of an Ensemble Kalman Filter with Three- and Four-Dimensional Variational Data Assimilation Methods in a Limited-Area Model over the Month of June 2003. *Monthly Weather Review*, 139:566–572, 2011.

[185] Jeffrey Anderson, Tim Hoar, Kevin Raeder, Hui Liu, Nancy Collins, Ryan Torn, and Avelino Avellano. The Data Assimilation Research Testbed: A Community Facility. *Bulletin of the American Meteorological Society*, 90:1283–1296, 2009.

[186] Jeffrey L. Anderson. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129:2884–2903, 2001.

[187] Jeffrey Anderson. A Local Least Squares Framework for Ensemble Filtering. *Monthly Weather Review*, 131:634–642, 2003.

[188] So-Young Ha and Chris Snyder. Influence of Surface Observations in Mesoscale Data Assimilation Using an Ensemble Kalman Filter. *Monthly Weather Review*, 142:1489–1508, 2014.

[189] J. P. Hacker and W. M. Angevine. Ensemble Data Assimilation to Characterize Surface-Layer Errors in Numerical Weather Prediction Models. *Monthly Weather Review*, 141:1804–1821, 2013.

[190] Nigel Wood and Andrew Staniforth. The Deep-Atmosphere Euler Equations with a Mass-Based Vertical Coordinate. *Quarterly Journal of the Royal Meteorological Society*, 129:1289–1300, 2003.

[191] Andrew Staniforth and Nigel Wood. The Deep-Atmosphere Euler Equations in a Generalized Vertical Coordinate. *Monthly Weather Review*, 131:1931–1938, 2003.

[192] Andrew Staniforth, Nigel Wood, and Claude Girard. Energy and Energy-Like Invariants for Deep Nonhydrostatic Atmospheres. *Quarterly Journal of the Royal Meteorological Society*, 129:3495–3499, 2003.

[193] Arne Sundström. Stability Theorems for the Barotropic Vorticity Equation. *Monthly Weather Review*, 97:340–345, 1969.

[194] Joseph Oliger and Arne Sundström. Theoretical and Practical Aspects of Some Initial Boundary-Value Problems in Fluid-Dynamics. *SIAM Journal of Applied Mathematics*, 35:419–446, 1978.

[195] Roger Temam and Joseph Tribbia. Open Boundary Conditions for the Primitive and Boussinesq Equations. *Journal of the Atmospheric Sciences*, 60:2647–2660, 2003.

[196] Piet Termonia. Monitoring and Improving the Temporal Interpolation of Lateral-Boundary Couplin Data for Limited-Area Models. *Monthly Weather Review*, 131: 2450–2463, 2003.

[197] Piet Termonia, Alex Deckmyn, and Rafiq Hamdi. Study of the lateral boundary condition temporal resolution problem and a proposed solution by means of boundary error restarts. *Monthly Weather Review*, 137:3551–3566, 2009.

[198] Fedor Mesinger. Forward-backward scheme, and its use in a limited area model. *Contributions To Atmospheric Physics*, 50:200–210, 1977.

[199] Fedor Mesinger and Katarina Veljovic. Limited Area NWP and Regional Climate Modeling: A Test of the Relaxation vs. Eta Lateral Boundary Conditions. *Meteorology and Atmospheric Physics*, 119:1–16, 2013. URL 10.1007/s00703-012-0217-5.

[200] R. Lehmann. On the Choice of Relaxation Coefficients for Davies' Lateral Boundary Scheme for Regional Weather Prediction Models. *Meteorology and Atmospheric Physics*, 52:1–14, 1993.

[201] Nils Gustafsson, Erland Källen, and Sigurdur Thorsteinsson. Sensitivity of Forecast Errors to Initial and Lateral Boundary Conditions. *Tellus Series A - Dynamic Meteorology and Oceanography*, 50:167–185, 1998.

[202] Song-You Hong and Hann-Ming Henry Juang. Orography Blending in the Lateral Boundary of a Regional Model. *Monthly Weather Review*, 126:1714–1718, 1998.

[203] Andrew Brown, Sean Milton, Brian Golding, John Mitchell, and Ann Shelly. Unified Modeling and Prediction of Weather and Climate: A 25-Year Journey. *Bulletin of the American Meteorological Society*, 93:1865–1877, 2012.

[204] Ryan D. Torn, Gregory J. Hakim, and Chris Snyder. Boundary conditions for limited-area ensemble Kalman filters. *Monthly Weather Review*, 134:2490–2502, 2006.

[205] Song-You Hong and Masao Kanamitsu. Dynamical Downscaling: Fundamental Issues from an NWP Point of View and Recommendations. *Asia-Pacific Journal of Atmospheric Sciences*, 50:83–104, 2014.

[206] D. Funaro and D. Gottlieb. A New Method of Imposing Boundary Conditions in Pseudospectral Approximations of Hyperbolic Equations. *Mathematics of Computation*, 51:599–613, 1988.

[207] Mark H. Carpenter, David Gottlieb, and Saul Abarbanel. Time-Stable Boundary Conditions for Finite-Difference Schemes Solving Hyperbolic Systems: Methodology and Application to High-Order Compact Schemes. *Journal of Computational Physics*, 111:220–236, 1994.

[208] Jan Nordström and Magnus Svärd. Well-posed boundary conditions for the Navier-Stokes equations. *SIaM Journal on Numerical Analysis*, 43:1231–1255, 2005.

[209] Matthias Hieber and Sylvie Monniaux. Well-Posedness Results for the Navier-Stokes Equations in the Rotational Framework. *Discrete and Continuous Dynamical Systems*, 33:5143–5151, 2013.

[210] Qionglei Chen, Changxing Miao, and Zhifei Zhang. Global well-posedness for the 3D rotating Navier-Stokes equations with highly oscillating initial data. *Pacific Journal of Mathematics*, 262:263–283, 2013.

[211] Jan Hesthaven. A stable penalty method for the compressible Navier-Stokes equations: I. Open boundary conditions. *SIaM Journal on Scientific Computing*, 17: 579–612, 1996.

[212] Magnus Svärd, Mark H. Carpenter, and Jan Nordström. A stable high-order finite difference scheme for the compressible Navier-Stokes equations, far-field boundary conditions. *Journal of Computational Physics*, 225:1020–1038, 2007.

[213] Heinz O. Kreiss and Lixin Wu. On the Stability Definition of Difference Approximations for the Initial Boundary Value Problem. *Applied Numerical Mathematics*, 12:213–227, 1993.

[214] Huw C. Davies. Limitations of Some Common Lateral Boundary Schemes used in Regional NWP Models. *Monthly Weather Review*, 111:1002–1012, 1983.

[215] A. McDonald. A Review of Lateral Boundary Conditions for Limited Area Forecast Models. *PINSA*, 65:91–105, 1999.

[216] B. Engquist and A. Majda. Absorbing Boundary Conditions for the Numerical Simulation of Waves. *Mathematics of Computation*, 31:629–651, 1977.

[217] A. McDonald. Boundary conditions for semi-Lagrangian schemes: Testing some alternatives in one-dimensional models. *Monthly Weather Review*, 128:4084–4096, 2000.

[218] A. McDonald. Transparent boundary conditions for the shallow-water equations: Testing in a nested environment. *Monthly Weather Review*, 131:698–705, 2003.

[219] A. McDonald. Transparent Lateral Boundary Conditions for Baroclinic Waves: A Study of two Elementary Systems of Equations. *Tellus Series A - Dynamic Meteorology and Oceanography*, 57:171–182, 2005.

[220] A. McDonald. Transparent Lateral Boundary Conditions for Baroclinic Waves II. Introducing Potential Vorticity Waves. *Tellus Series A - Dynamic Meteorology and Oceanography*, 58:210–220, 2006.

[221] A. McDonald. Transparent Lateral Boundary Conditions for Baroclinic Waves III. Including Vertical Shear. *Tellus Series A - Dynamic Meteorology and Oceanography*, 61:227–231, 2009.

[222] Piet Termonia and Fabrice Voitus. Externalizing the lateral boundary conditions from the dynamic core in semi-implicit semi-Lagrangian models. *Tellus Series A – Dynamic Meteorology and Oceanography*, 60:632–648, 2008.

[223] F. Voitus, P. Termonia, and P. Bénard. Well-posed lateral boundary conditions for spectral semi-implicit semi-Lagrangian schemes: Tests in a one-dimensional model. *Monthly Weather Review*, 137:315–330, 2009.

[224] Edward Norton Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20:130–141, 1963.

[225] Barry Saltzman. Finite amplitude free convection as an initial value problem – 1. *Journal of the Atmospheric Sciences*, 19:329–341, 1962.

[226] Takashi Hikihara, Philip Holmes, Tsutomu Kambe, and Giuseppe Rega. Introduction to the Focus Issue: Fifty Years of Chaos: Applied and Theoretical. *Chaos*, 22, 2012.

[227] Peter G. Baines. Lorenz, EN 1963: Deterministic nonperiodic flow. Journal of the Atmospheric Sciences 20, 130-41. *Progress in Physical Geography*, 32:475–480, 2008.

[228] K. Judd, C. A. Reynolds, T. E. Rosmond, and L. A. Smith. The geometry of model error. *Journal of the Atmospheric Sciences*, 65:1749–1772, 2008.

[229] C. E. Leith. Theoretical skill of Monte-Carlo forecasts. *Monthly Weather Review*, 102:409–418, 1974.

[230] A. R. Brown, A. C. M. Beljaars, and H. Hersbach. Errors in parametrizations of convective boundary-layer turbulent momentum mixing. *Quarterly Journal of the Royal Meteorological Society*, 132, 2006.

[231] S. F. Milton and C. A. Wilson. The impact of parameterized subgrid-scale orographic forcing on systematic errors in a global NWP model. *Monthly Weather Review*, 124:2023–2045, 1996.

[232] J. G. Charney. On a physical basis for numerical prediction of large-scale motions in the atmosphere. *Journal of Meteorology*, 6:371–385, 1949.

[233] T. L. Keller. Implications of the hydrostatic assumption on atmospheric gravity waves. *Journal of the Atmospheric Sciences*, 51:1915–1929, 1994.

[234] Peter Linz. A critique of numerical analysis. *Bulletin of the American Mathematical Society*, 19:407–417, 1988.

[235] P. Bénard. Stability of semi-implicit and iterative centered-implicit time discretizations for various equation systems used in NWP. *Monthly Weather Review*, 131: 2479–2491, 2003.

[236] A. Gassmann. A global hexagonal C-grid non-hydrostatic dynamical core (ICON-IAP) designed for energetic consistency. *Quarterly Journal of the Royal Meteorological Society*, 139:152–175, 2013.

[237] J. Thuburn. Numerical wave propagation on the hexagonal C-grid. *Journal of Computational Physics*, 227:5836–5858, 2008.

[238] W. C. Skamarock, J. B. Klemp, M. G. Duda, L. D. Fowler, S-H. Park, and T. D. Ringler. A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tesselations and C-grid staggering. *Monthly Weather Review*, 140:3090–3105, 2012.

[239] Ulf Olsson. *Generalized Linear Models. An Applied Approach*, chapter General Linear Models, page 3. Studentlitteratur, 2002.

[240] B. I. Miller, E. C. Hill, and P. P. Chase. A revised technique for forecasting hurricane movement by statistical methods. *Monthly Weather Review*, 96:540–548, 1968.

[241] W. H. Klein, B. H. Lewis, and I. Enger. Objective prediction of five-day mean temperature during winter. *Journal of Meteorology*, 16:672–682, 1959.

[242] W. H. Klein and H. R. Glahn. Forecasting local weather by means of model output statistics. *Bulletin of the American Meteorological Society*, 55:1217–1227, 1974.

[243] Michael Ek, NCEP Environmental Modeling Center, Michael Ek. [Private communication].

[244] Jacks, E. and Bower, J. B. and Dagostaro, V.J. and Dallavalle, J. P. and Erickson, M. C. and Su, J. New NGM-based MOS guidance for maximum/minimum temperature, probability of precipitation, cloud amount, and surface wind. *Weather and Forecasting*, 5:128–138, 1990.

[245] Erickson, M. C. and Bower, J. B. and Dagostaro, V.J. and Dallavalle, J. P. and Jacks, E. and Jensenius Jr, J. S. and Su, J. C. Evaluating the impact of RAFS changes on the NGM-based MOS guidance. *Weather and Forecasting*, 6:142–147, 1991.

[246] Müller, M. D. Effects of model resolution and statistical postprocessing on shelter temperature and wind forecasts. *Journal of Applied Meteorology and Climatology*, 50:1627–1636, 2011.

[247] Yudi Pawitan. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, chapter Modelling relationships: regression models, page 151. Oxford University Press, 2001.

[248] T. J. Valentine. A note on multicollinearity. *Australian Economic Papers*, 8: 99–105, 1969.

[249] Henrik Madsen. *Time Series Analysis*, chapter Identification, estimation, and model checking, page 149. Chapman & Hall/CRC, 2008.

[250] Andrzej Gałecki and Tomasz Burzykowski. *Linear Mixed-Effects Models Using R. A Step-by-Step Approach*, chapter 10 through 12, pages 177–243. Springer, 2013.

[251] Henrik Madsen. *Time Series Analysis*, chapter Stochastic processes, page 119. Chapman & Hall/CRC, 2008.

[252] Henrik Madsen. *Time Series Analysis*, chapter Stochastic processes, page 108. Chapman & Hall/CRC, 2008.

[253] B. Jonsson. Prediction with a linear regression model and errors in a regressor. *International Journal of Forecasting*, 10:549–555, 1994.

[254] Monache, Luca Delle and Eckel, F. Anthony and Rife, Daran L. and Nagarajan, Badrinath and Searight, Keith. Probabilistic Weather Prediction with an Analog Ensemble. *Monthly Weather Review*, 141:34983516, 2013.

[255] Alessandrini, S. and Monache, L. Delle and Sperati, S. and Nissen, J. N. A novel application of an analog ensemble for short-term wind power forecasting. *Renewable Energy*, 76:768–781, 2015.

[256] Nils P. Wedi. Increasing horizontal resolution in numerical weather prediction and climate simulations: illusion or panacea? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372, 2014.

[257] M. Annunziato. A finite difference method for piecewise deterministic processes with memory. *Mathematical Modelling and Analysis*, 12:157–178, 2007.

[258] M. Annunziato. A finite difference method for piecewise deterministic processes with memory. II. *Mathematical Modelling and Analysis*, 14:139–158, 2009.

[259] Palmer, T. Towards the probabilistic Earth-system simulator: a vision for the future of climate and weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 138:841–861, 2012.

[260] Piotr K. Smolarkiewicz, Christian Kühnlein, and Nils P. Wedi. A consistent framework for discrete integrations of soundproof and compressible PDEs of atmospheric dynamics. *Journal of Computational Physics*, 263:185–205, 2014.

[261] Kaas, E. and Sørensen, B. and Lauritzen, P. H. and Hansen, A. B. A hybrid Eulerian-Lagrangian numerical scheme for solving prognostic equations in fluid dynamics. *Geoscientific Model Development*, 6:2023–2047, 2013.

[262] James R. Holton and Gregory J. Hakim. *An Introduction to Dynamic Meteorology (5th ed.)*. Academic Press (AP), 2012.

[263] Thomas Tomkins Warner. *Numerical Weather and Climate Prediction*. Cambridge University Press, 2010.