
ON THE PREDICTABILITY OF EXCEPTIONAL ERROR EVENTS IN
WIND POWER FORECASTING
—AN ULTRA LARGE ENSEMBLE APPROACH—

I N A U G U R A L – D I S S E R T A T I O N
ZUR
ERLANGUNG DES DOKTORGRADES
DER MATHEMATISCH–NATURWISSENSCHAFTLICHEN FAKULTÄT
DER UNIVERSITÄT ZU KÖLN

vorgelegt von

Jonas Berndt

aus Köln

JÜLICH, 2018

Berichterstatter: PD Dr. H. Elbern
Prof. Dr. Y. Shao

Tag der mündlichen Prüfung: 02. März 2018

Absence of evidence is not evidence of absence.
Carl Sagan

Abstract

Exceptional error events in wind power forecasting impose a major obstacle to today's economic and reliable power supply. As installed capacities grow, the impact of associated forecast errors becomes increasingly critical for the electrical grid stability and requires the attendance of growing reserve capacities. The predictability of such error events is fundamentally restricted by the underlying weather forecast, resting on limitations of state-of-the-art numerical weather prediction systems. These forecasts must be furnished with likelihood, implying the operation of model ensembles. Ensembles of numerical weather predictions provide estimates of forecast uncertainties and allow users such as grid operators and energy market participants to prepare for potential forecast errors. However, present computational resources restrict meteorological ensembles to a moderate number of members, which reduces the likelihood to capture exceptional error events.

This work aims to identify imminent forecast errors affecting the energy sector. To this end, the standard sizes of meteorological ensembles are increased from $O(10)$ to an ultra large ensemble size of $O(1000)$ members to accomplish an improved approximation of the probability density function. For this purpose, a novel approach of an ensemble control system has been developed on a 5-dimensional interconnected Petaflop architecture. Within this system, the Weather Research and Forecasting (WRF) model has been modified towards a stand-alone ensemble version. The developed software constitutes the meteorological part of Ensembles for Stochastic Integration of Atmospheric Systems (ESIAS-met). Further, an increased ensemble size favors the application of nonlinear data assimilation techniques based on the particle filter, while imposing the challenge of growing computational expenses of a resampling step within the particle filter algorithm. ESIAS-met presents a computationally efficient solution to the problem, by realizing a parallel execution of the ensemble within a single executable. Performance measurements within this work demonstrate strong scalability of the system with up to 4096 ensemble members utilizing 262,144 cores. Moreover, for a fixed problem size, the computational expenses of a particle filter resampling step are shown to be independent of the ensemble size. The ESIAS-met system is further applied to investigate the benefit of an increased ensemble size on the predictability of recent exceptional error events. The analysis reveals, that despite the large ensemble size, the forecast error is only represented by single outliers. As an approach to identify imminent forecast errors, higher order moments prove to provide a robust measure of the proper direction of forecast error and to assess their likelihood of appearance. It is shown, that at least $O(100)$ of ensemble members are needed to resolve the higher order moments sufficiently well. Hence, the results achieved in this work yield important potential for future warning capabilities of exceptional error events.

Kurzzusammenfassung

Außergewöhnlich hohe Fehler in den Windleistungsprognosen stellen ein relevantes Problem für eine wirtschaftliche und gesicherte Energieversorgung dar. Mit wachsender Anzahl installierter Windkraftanlagen bleiben die Magnitude dieser Fehlvorhersagen für die Netzstabilität kritisch, und immer mehr Regelleistung muss von Übertragungsnetzbetreibern beschafft werden, um einen entsprechenden Ausgleich gewährleisten zu können. Die Vorhersagbarkeit dieser außergewöhnlich hohen Fehler ist durch die zugrunde liegende Wettervorhersage limitiert, da selbst modernste numerische Wettervorhersagemodelle inhärente Restriktionen unterliegen. Somit ist eine Abschätzung der Vorhersageunsicherheit erstrebenswert, die in der Praxis von meteorologischen Ensemblesystemen realisiert wird. Prinzipiell erlauben diese den Netzbetreibern, sich auf mögliche Vorhersagefehler vorzubereiten. Heutige operationelle Ensemblesysteme sind jedoch aufgrund beschränkter Rechenkapazität auf eine kleine Anzahl von Ensemblemitgliedern beschränkt, was die Wahrscheinlichkeit reduziert, vor außergewöhnlichen Fehlvorhersagen ausreichend warnen zu können.

Diese Arbeit hat das Ziel, potenziell auftretende Fehler in der Energievorhersage zu identifizieren. Für diesen Zweck werden die Standardgrößen meteorologischer Ensemble von $O(10)$ auf ein ultragroßes Ensemble mit $O(1000)$ Mitgliedern erweitert, um eine verbesserte Approximation der Wahrscheinlichkeitsdichtefunktion zu erlangen. Hierfür wird ein neuartiger Ansatz eines Ensemble-Kontrollsystems auf einem Petaflop-Rechner entworfen. Innerhalb dieses Systems wird das Weather Research and Forecasting (WRF) Modell zu einer eigenständigen Ensembleversion weiterentwickelt. Diese Umgebung stellt den meteorologischen Teil von Ensembles for Stochastic Integrations of Atmospheric Systems (ESIAS-met) dar. Weiterhin begünstigt die hohe Mitgliederanzahl nichtlineare Verfahren der Datenassimilation, die auf dem Partikelfilter beruhen. Jedoch wächst mit steigender Mitgliederanzahl der Rechenaufwand eines Partikelfilter Resampling Schrittes. ESIAS-met stellt eine effiziente Lösung dieses Problems dar, indem der Ensemblelauf innerhalb eines Programms realisiert wird. Laufzeitanalysen zeigen ausgeprägte parallele Skalierbarkeit bei bis zu 4096 Ensemblemitgliedern auf 262.144 Prozessoren. Weiterhin wird gezeigt, dass für eine feste Problemgröße der Rechenaufwand eines Resampling Schrittes unabhängig von der Anzahl der Ensemblemitglieder ist. Darüber hinaus wird das ESIAS-met System genutzt, um das Potenzial einer vergrößerten Ensembleanzahl im Rahmen von Vorhersagen außergewöhnlicher Fehler zu untersuchen. Die Analyse zeigt, dass trotz der hohen Ensembleanzahl nur einzelne Ausreißer den Vorhersagefehler adäquat darstellen. Dabei haben sich höhere statistische Momente als ein möglicher Ansatz herausgestellt, potenziell auftretende Fehlvorhersagen zu identi-

fizieren. Diese geben Auskunft über die Richtung des Vorhersagefehlers und bewerten dessen Wahrscheinlichkeit. Dabei wird weiterhin gezeigt, dass mindestens einige Hundert Ensemblemitglieder nötig sind, um die statistischen Momente verlässlich abzubilden. Die Ergebnisse dieser Arbeit implizieren wichtige Folgerungen für zukünftige Warnsysteme für außergewöhnliche Fehlvorhersagen.

Contents

1	Introduction	1
2	Ensemble forecasting and nonlinear data assimilation	7
2.1	Uncertainties in Numerical Weather Predictions	7
2.2	Representation of initial condition error	8
2.3	Representation of model error	12
2.3.1	Stochastically Perturbed Parameterization Tendency scheme	13
2.3.2	Stochastic Kinetic Energy Backscatter scheme	15
2.4	The Sequential Importance Resampling Filter	18
3	ESIAS – Ensembles for Stochastic Integration of Atmospheric Systems	21
3.1	The Weather Research and Forecasting (WRF) model	22
3.2	The supercomputer JUQUEEN	24
3.3	Computational performance of the WRF model on JUQUEEN	24
3.3.1	Porting and Tuning	24
3.3.2	Performance analysis of a single WRF run	25
3.3.3	Computational improvement of SPPT and SKEB	27
3.4	ESIAS-met	32
3.4.1	A parallel ensemble version of the WRF model	32
3.4.2	Proof of concept: parallel scalability of ESIAS-met	35
3.4.3	Flowchart of ESIAS-met	39
4	Wind power forecasting with the WRF model	41
4.1	WRF model setup	41
4.1.1	Domain configuration	41
4.1.2	Initial conditions and lateral boundary values	44
4.2	Meteorological observations	45
4.2.1	Comparison of model data and observations	45
4.2.2	Conventional in situ observations	46
4.2.3	Measurement towers	47
4.3	IWES physical wind power model	49

4.4	Sensitivity of the WRF model to physical parameterizations	50
4.4.1	Experimental setup	50
4.4.2	Evaluation Methodology	53
4.4.3	Overview of the sensitivity study	53
4.4.4	Results	54
4.4.5	Discussion on WRF's wind bias	65
4.5	Optimization of the WRF model for wind power forecasting	65
4.5.1	Roughness length	66
4.5.2	The YSU PBL scheme	68
4.5.3	Experimental setup	70
4.5.4	Results	71
5	Evaluation of an ultra large wind power ensemble	75
5.1	Experimental setup	75
5.2	Case study selection	76
5.3	Meteorological situation	79
5.4	Results	79
5.4.1	Meteorological evaluation	79
5.4.2	Wind power evaluation	85
5.5	Outlook: Evaluation of an ultra large solar power ensemble	90
6	Summary	95
7	Conclusion and outlook	99
A	Verification statistics at measurement towers	101

List of Figures

3.1	Flowchart of the WRF modeling system	23
3.2	Parallel scaling behavior of the WRF model on JUQUEEN	26
3.3	Necessary symmetry conditions of complex Gaussian white noise in the SPPT and SKEB scheme	28
3.4	RMSE of increments in ensemble mean of 100 m wind speed obtained with the initial STOCH code and modified versions	31
3.5	RMSE of increments in ensemble spread of 100 m wind speed obtained with the initial STOCH code and modified versions	31
3.6	Visualization of the MPI concept to modify the WRF modeling system to a stand-alone ensemble control system	33
3.7	CPU times and amount of transferred data for a complete particle filter resampling step with ESIAS-met	39
3.8	Flowchart of the ESIAS-met modeling system	40
4.1	WRF domain configuration with its nesting procedure on a terrain height map	43
4.2	Spatial distribution of onshore wind power capacity over Germany, as of 2014, and location of measurement towers	49
4.3	RMSE of hub height wind speeds at measurement towers for eight different WRF model configurations, 1 – 31 August 2014	59
4.4	RMSE of hub height wind speeds at measurement towers for eight different WRF model configurations, 1 – 30 November 2014	60
4.5	Diurnal cycle of verification statistics for the day-ahead wind power forecast with eight different WRF configurations, 1 – 31 August 2014	62
4.6	Diurnal cycle of verification statistics for the day-ahead wind power forecast with eight different WRF configurations, 1 – 30 November 2014	63
4.7	Roughness length over Germany according to the WRF model	67
4.8	Diurnal cycle of verification statistics for day-ahead wind power forecast with optimized WRF configurations, 1 – 31 August 2014	72

4.9	Diurnal cycle of verification statistics for day-ahead wind power forecast with optimized WRF configurations, 1 – 30 November 2014 . . .	73
5.1	Instantaneous SPPT perturbation pattern	77
5.2	Instantaneous SKEB perturbation patterns	77
5.3	TSO day-ahead wind power forecast and real power feed-in on 9 August 2014	78
5.4	NCEP GFS analysis at 9 August 2014	80
5.5	Meteosat Second Generation SEVIRI satellite imagery, channel 9, 10.8 μm	80
5.6	Forecast-minus-analysis increment of mean sea level pressure at 12 UTC 9 August 2014	81
5.7	Ensemble spread of 100 m wind speed for the 1024-member ensemble .	82
5.8	Spaghetti plot of geopotential height for the 1024-member ensemble .	83
5.9	Nested quantiles of hub height wind speeds at measurement towers . .	84
5.10	Day-ahead forecast of the ultra large wind power ensemble for 9 August 2014	86
5.11	Nested quantiles of day-ahead wind power forecast for 9 August 2014 .	86
5.12	Day-ahead forecast of the ultra large wind power ensemble for 9 August 2014 (SPPT)	87
5.13	Day-ahead wind power forecast of ensemble member 305 and 558 for 9 August 2014	88
5.14	Root moments of wind power distribution from $n = 2$ to $n = 5$ for 9 August 2014	89
5.15	Nested percentiles of standard deviation for bootstrapped samples of different ensemble sizes (wind power case study)	91
5.16	Nested percentiles of skewness for bootstrapped samples of different ensemble sizes (wind power case study)	91
5.17	Day-ahead forecast of the ultra large solar power ensemble for 28 November 2014	92
5.18	Root moments of solar power distribution from $n = 2$ to $n = 5$ for 28 November 2014	93
5.19	Nested percentiles of skewness for bootstrapped samples of different ensemble sizes (solar power case study)	94
A.1	Bias of hub height wind speeds at measurement towers for eight different WRF model configurations for 1 – 31 August 2014	102
A.2	Correlation between hub height wind speeds and model data at measurement towers for eight different WRF model configurations for 1 – 31 August 2014	103

A.3	Bias of hub height wind speeds at measurement towers for eight different WRF model configurations for 1 – 30 November 2014	104
A.4	Correlation between hub height wind speeds and model data at measurement towers for eight different WRF model configurations for 1 – 30 November 2014	105

List of Tables

2.1	Sources of uncertainties in NWP models	9
3.1	CPU times of the WRF model on JUQUEEN with initial and improved compiler instructions	26
3.2	CPU times of the WRF model with the SKEB scheme and varying processor number for different implementations of the STOCH code . . .	29
3.3	CPU times of ESIAS-met with increasing ensemble size	36
3.4	Number of MPI messages, transferred Mbytes and CPU times associated with a resampling procedure within ESIAS-met	38
4.1	Configuration of the WRF model domain	43
4.2	Site characteristics of onshore measurement towers	48
4.3	Summary of eight WRF configurations with different physical parameterizations	51
4.4	Verification statistics for day-ahead forecasts with eight different WRF configurations and GFS model fields for conventional in situ observations, 1 – 31 August 2014	56
4.5	Verification statistics for day-ahead forecasts with eight different WRF configurations and GFS model fields for conventional in situ observations, 1 – 30 November 2014	57
4.6	Verification statistics for the day-ahead wind power forecast with eight different WRF configurations, 1 – 31 August 2014	64
4.7	Verification statistics for the day-ahead wind power forecast with eight different WRF configurations, 1 – 30 November 2014	64
4.8	Experimental setup of WRF model optimizations	70
4.9	Verification statistics for the day-ahead wind power forecast with optimized WRF model configurations, 1 – 31 August 2014	71
4.10	Verification statistics for the day-ahead wind power forecast with optimized WRF model configurations, 1 – 30 November 2014	71
5.1	Stochastic model parameter for the SPPT and SKEB scheme	76

List of Acronyms

ARW	Advanced Research WRF
BV	Breeding Vector
CONUS	Contiguous United States
ECMWF	European Centre for Medium-Range Weather Forecasts
EM	Ensemble Mean
EPS	Ensemble Prediction System
ES	Ensemble Spread
ET	Ensemble Transformation
ESIAS	Ensembles for Stochastic Integration of Atmospheric Systems
GFS	Global Forecast System
GEFS	Global Ensemble Forecast System
IFS	Integrated Forecast System
IWES	(Fraunhofer)-Institut für Windenergie und Energiesystemtechnik
LSM	Land Surface Model
MPI	Message Passing Interface
NetCDF	Network Common Data Format
NCEP	National Centre for Environmental Prediction
NOAA	National Oceanic and Atmospheric Administration
OpenMP	Open Multi-Processing
PDF	Probability Density Function
PBL	Planetary Boundary Layer
RMSE	Root Mean Square Error
SIMD	Single Instruction, Multiple Data
SIR	Sequential Importance Resampling
SIRF	Sequential Importance Resampling Filter
SMT	Simultaneous Multithreading
SPPT	Stochastically Perturbed Parameterization Tendency (scheme)

SKEB	Stochastic Kinetic Energy Backscatter (scheme)
SV	Singular Vector
TSO	Transmission System Operator
WRF	Weather Research and Forecasting (model)

Chapter 1

Introduction

A limited predictability of the atmospheric state imposes wind power as an energy source of inherent uncertainty in space and time. With the increasing amount of installed wind power capacity, accurate wind forecasts have become invaluable for the energy sector (*Giebel et al.*, 2011; *Bremen and Wessel*, 2015). In meteorology, the realm of predictability is a research area of its own (*Kalnay*, 2003; *Palmer and Hagedorn*, 2006), which aims to quantify the flow-dependent uncertainties of forecasts.

Yet, predictability issues on weather dependent energy production are subdued by the technical and economical constraints of the energy system. In Germany, the infeed from renewable energy sources has priority over conventional electricity generators by governmental legislation, while transmission system operators (TSOs) have to ensure that supply matches demand to facilitate a secure electrical grid. In the day-ahead notice, balance responsible parties report expected load schedules for the following day (*Hirth and Ziegenhagen*, 2013). TSOs derive aggregated forecasts for the control area of Germany to anticipate regulations of power plants or negotiate supply, and accordingly, pricings are formed on the energy stock exchange. On the following day, if schedules deviate from the actual supply, balance responsible parties have the chance to adjust their portfolio on the intraday energy market (*Schroedter-Homscheidt et al.*, 2015). Here, prices can even be negative, in case companies need to pay for unloading excess power. The TSOs take care of any remaining imbalances by activating physical control capacities to safeguard the electrical grid stability. The allocation of the associated costs, the balancing price, is distributed among the balance responsible parties. The accuracy of day-ahead wind power forecasts is therefore not only crucial for electrical grid stability, but also for an economically viable integration of wind power into the electrical grid (*Hirth and Ziegenhagen*, 2013).

In general, numerical weather predictions (NWP) are the basis for wind power forecast systems for the day-ahead horizon (*Focken et al.*, 2001; *Möhrten and*

Jørgensen, 2008; Vogt et al., 2016). Power curves convert model wind speed and direction at turbine hub heights to wind power, and are either provided by the manufacturer or derived for a whole region of wind farms upon historical data (*Lydia et al., 2014*). Forecast systems utilized by the TSOs are based on multiple NWP systems and are optimized by model output statistics to remove systematic errors. Thereby, Germany can be treated approximately as one control area (*Zolotarev and Gökeler, 2011*) and local forecast errors are in general reduced by spatial smoothing effects (*Focken et al., 2002*). Nevertheless, the principle forecast skill is mainly determined by the underlying NWP system (*Giebel et al., 2005*).

It is generally agreed that current wind power forecast systems show on average a satisfactory forecast accuracy. However, there still exist exceptional error events in the day-ahead forecast which are a major obstacle to a stable and safe grid operation. Such events are caused by an erroneous representation of the atmospheric state by all available NWP systems at that time (*Dobschinski et al., 2017*). Even area aggregation across multiple countries may not be sufficient to balance the forecast errors (*Möhrlen and Jørgensen, 2017*). Exceptional error events are rare by definition, corresponding to 0.05% of all times where the demand for control power exceeds the actual capacity (*Consentec, 2010*). According to *Stark (2015)*, balancing prices, which are in general of the order of $\pm 10 \text{ €/MWh}$, may go up to $\pm 6000 \text{ €/MWh}$, and a trader's one month's profit may be ruined at the energy stock exchange (*Good, 2017*). Since 2012, the German *Bundesnetzagentur* introduced an additional penalty of 100 €/MWh to the balance responsible parties, if at least 80% of the control power capacity have to be utilized by the TSOs (*Consentec, 2012*). The revenue is shared among the consumers, however, the maintenance of control power capacity is incorporated in the electricity price. Therefore, exceptional error events in the day-ahead forecast are not only critical and costly to participants of the energy sector, but also put a constant and disproportionate cost on the electricity price and are a major obstacle to the integration of wind power into the electrical grid.

Lundgren (2015) summarizes critical weather conditions which led to exceptional error events in the past. For wind power, these conditions include strong cyclogenesis, i.e. the intensity and spatiotemporal position of low pressure systems and associated frontal movements. Further, winterly stable conditions and a pronounced summerly diurnal cycle may cause severe systematic errors. For solar power, these conditions include cloud coverage after the passage of a cold front, the spatiotemporal evolution of convective systems, formation or clearing of low stratus and large-scale dust events. Exceptional error events in wind power forecasts are hence linked to meteorological events of low predictability (*Steiner et al., 2017*), which rest on inevitable shortcomings of today's NWP systems that all major

weather centres around the world face.

In this sense, a single deterministic forecast from a NWP system has a fundamentally limited usefulness. Since the pioneering work of *Lorenz (1963)*, the atmospheric evolution is well known to be highly sensitive to its initial conditions, characterizing its dynamics as a chaotic system. Any NWP forecast must therefore be furnished with likelihood. Hence, exceptional error events can only be anticipated by a forecast of forecast skill, leading to a stochastic extension of the model's equations. Formally, the stochastic evolution of the atmospheric state with uncertain initial conditions is described by the Liouville equation, and as a stochastic-dynamic extension by the Fokker-Planck equation. However, the only feasible approach leads to the integration of model ensembles, which approximate the probability density function (PDF) of the model state by a finite sample of different realizations of the NWP model (*Epstein, 1969; Leith, 1974*).

Ensemble forecasting has become a common procedure at weather centres in the last decades (*Toth and Kalnay, 1997; Molteni et al., 1996; Li et al., 2008; Denhard et al., 2016*). Their undisputed success has been based on the identification of flow-dependent perturbations of initial conditions that exhibit the fastest growing modes (*Toth and Kalnay, 1993; Buizza et al., 1999*). Model uncertainty representation has proven to reduce the lack of dispersion among the ensemble members and thereby correcting the estimation of predictability (*Palmer et al., 2009; Berner et al., 2011*). Most notably, convective scale instabilities, boundary layer dynamics, cloud induced modulation of insolation, and the various mechanisms to trigger or influence these processes, must be accounted for in the parameterization of physical processes by various perturbations. Furthermore, ensemble forecasting is meanwhile also applied to mesoscale convection-permitting NWP models (*Wang et al., 2008; Bouttier et al., 2012; McCabe et al., 2016; Hagelin et al., 2017*). The usefulness of mesoscale ensemble modeling is based on the notion, that although higher resolution and enhanced representation of sub-grid scale processes produce more realistic forecasts of severe weather phenomena, yet it is challenged by the proper simulation of small scale processes such as convection and cloud formation (*Mass et al., 2002; Eckel and Mass, 2005; Theis et al., 2005*). This notion is supported by the work of (*Lorenz, 1969*), who showed that errors tend to grow more quickly in time with increasing resolution.

Ensemble forecasting samples the PDF of the model state, subject to all available information of uncertainty in the initial conditions and model formulation. This procedure is essentially restricted by the sample size due to limited computational resources. Operational ensembles of global circulation and convection-permitting models are all of the order of $O(10)$ ensemble members. Examples are the Global Ensemble Forecasting System (GEFS) at the National Centre for Environmental

Prediction (NCEP): 21 member; the Ensemble Prediction System (EPS) at the European Centre for Medium-Range Weather Forecasts (ECMWF): 51 members; the ICOSahedral Nonhydrostatic ensemble at the German Weather Service (DWD): 40 members; the Met Office ensemble system (MOGREPS-UK): 20 members; the high-resolution rapid refresh (HRRR) ensemble of the Weather Research and Forecasting (WRF) model at the National Oceanic and Atmospheric Administration (NOAA): 20 members. Numerous studies have proven, that a limited ensemble size restricts the ensemble performance (*Richardson, 2001; Mullen and Buizza, 2002; Hagelin et al., 2017*), driven by the certainty, that the sampling error of the sample mean is proportional to \sqrt{N} , with N the ensemble size.

Besides a restriction in the ensemble size, approximations and linearizations in NWP data assimilation techniques prohibit the capture of exceptional error events. Despite their sophistication and acknowledged performance in NWP systems (*Rabier et al., 2000; Fisher et al., 2011; Bishop and Hodyss, 2011; Wang and Lei, 2014*), operational methods all rest on Gaussian assumptions. Gradient methods minimize a cost function, and in case of multimodal prior distributions, it is not clear whether this minimization converges towards a global minimum. Four-dimensional variational data assimilation (*Courtier et al., 1994*) adds spatiotemporal consistency. However, an adjoint model is involved which relies on linearizations. The Ensemble Kalman Filter (*Evensen, 1994*) assumes normal distributions for observations as well as the model state. Although the forecast model itself is nonlinear, the Ensemble Kalman Filter forms a Gaussian posterior, characterized by the ensemble mean and associated covariances. Consequently, all operational data assimilation methods tend to prohibit the capture of exceptional error events, as the evolution of low probability tails in the prior and posterior PDF is suppressed.

Particle filters are increasingly gaining attention in the survey of a nonlinear data assimilation technique (*van Leeuwen, 2009*). The particle filter solves the full data assimilation problem, equivalent to Bayes's theorem for probability densities, without any assumption on the prior and posterior model state PDF. Each ensemble member, also referred to as a particle, is related to a posterior weight according to the likelihood upon observations. Resampling reduces the variance among the ensemble members by rejection of members with low weights and duplication of members with high weights (*Douc et al., 2005*). Thereupon, low probability events have the potential to survive the assimilation procedure of the particle filter, though the likelihood remains small by resampling.

The application of particle filters in high-dimensional systems faces a major obstacle. The only approximation to the nonlinear solution of the data assimilation problem is in the ensemble size. Yet, ensemble sizes of NWP systems are comparably small and the particle filter in its basic form tends to degenerate. In practise,

a single member is assigned with all the weight and a posterior PDF becomes meaningless (*Snyder et al.*, 2008). Upon this notion, many different variants of the particle filter have been proposed. All have in common, that they rely on improving the ensemble members' likelihood to achieve posterior weights of similar sizes. The notion is to draw the ensemble members towards observations to form a proposal density which is subject of sampling. One might simply think of a Gaussian of a local Ensemble Kalman Filter step as proposal density, or to nudge the ensemble members towards observations (*van Leeuwen*, 2009). More sophisticated methods target similar weights by construction, e.g. the implicit particle filter (*Chorin et al.*, 2010), which introduces an optimal proposal density by mapping the implicit sampling space to the original state space, or the equal-weight particle filter (*van Leeuwen*, 2010), which obtains equal weights by a proposal density which depends on all members at the previous time steps. The localized particle filter (*Poterjoy*, 2016) adopts the idea of localization used in the Ensemble Kalman Filter to formulate weights with a limited radius of influence. The class of particle smoothers weights the ensemble at the current time step upon information of the likelihood at a future time step. One can think of many variants, with the auxiliary particle filter being the most prominent (*Pitt and Shephard*, 1999). However, despite the sophistication involved in all the approaches listed above, the ensemble size still remains crucial for their performance.

Resampling requires inter-member communication and autonomous ensemble execution appears therefore suboptimal. Further, the variance among ensemble members should ideally be monitored frequently prior to the occurrence of filter degeneracy. Both requirements may lead to substantial computational times for ensemble sizes beyond state-of-the-art. It is not necessarily clear how a software environment for this purpose may be designed, as yet no convenient approach has been proposed in the literature.

Studies on the benefit of ultra large ensemble sizes are a rarity. The terminology of ultra large refers here to ensemble sizes beyond 1000 members. Already *Buizza and Palmer* (1998) point out, that an increase in ensemble size beyond 100 members is expected to have a beneficial impact on the outlier statistic, which is the key score for any prediction of exceptional error events. *Miyoshi et al.* (2014) realize a 10,240-member ensemble of an intermediate atmospheric model and investigate the Gaussianity of the forecast PDF. For less than 10% of the forecast time, they find pronounced non-Gaussianity PDFs, and for less than 1% a pronounced bimodality. However, they conclude, a minimum of $O(1000)$ ensemble members is needed to resolve this non-Gaussianity. It can be assumed, that these results amplify in the case of a full atmospheric model, especially in the mesoscale, as higher resolution increases nonlinearity. Therefore, non-Gaussian model PDFs will most likely be

suppressed by the sampling error of operational ensemble systems, leading to insufficient representation of exceptional error events.

To summarize, the predictability of exceptional error events in wind power forecasting is likely limited by a restricted ensemble size and Gaussian data assimilation techniques in NWP modeling. Thereupon, this work increases the standard ensemble sizes in NWP modeling from the order of $O(10)$ to $O(1000)$ members in the frame of a demonstrator system. For this purpose, the Weather Research and Forecasting (WRF) model (*Skamarock et al.*, 2008) is applied, a state-of-the-art mesoscale NWP model. Each ensemble member realizes a convection-permitting, high-resolution forecast over the target area of Germany. The ultra large ensemble size appears particularly favorable to apply particle filtering or smoothing techniques as a nonlinear data assimilation technique. A software environment shall be developed which realizes particle filtering computational efficient, independent of the ensemble size.

The work's objectives may be summarized by the following questions:

- How may a software environment be designed which efficiently executes particle filtering with an ultra large ensemble size?
- What is the benefit of increasing the ensemble size of NWP models with respect to the predictability of exceptional error events in wind power forecasting?

This thesis is structured as follows: Chapter 2 formally introduces ensemble forecasting, with a particular attention to the methods utilized to generate the ultra large ensemble. Further, the basic particle filter and resampling methods are described as a basis for any software developments to follow. Chapter 3 presents a strategy to execute the particle filter in the realm of an ultra large ensemble size in a convenient and computational efficient way. A proof of concept is demonstrated on a Blue Gene system in the frame of a feasibility study. Chapter 4 is designed to set the stage for the ultra large wind power ensemble. The WRF model setup is described in detail and the wind power model is introduced. A further aim is to derive the best possible deterministic forecast to serve as the model setup for the ultra large ensemble. Thereby, Chapter 4 touches upon model optimization of the surface boundary conditions and planetary boundary layer parameterization. Chapter 5 presents an evaluation of the ultra large ensemble on one of the major error events in wind power forecasting of the last years. Results are evaluated against meteorological observations and the real wind power feed-in. Additionally, an analysis of a solar power case study supports the main findings. A summary of the work is given in Chapter 6. Concluding remarks appear in Chapter 7 and directions for future work are presented.

Chapter 2

Ensemble forecasting and nonlinear data assimilation

Ensemble forecasting aims to quantify the flow-dependent uncertainty of the atmospheric state derived by a NWP model. A finite ensemble size of different model formulations initialized from perturbed initial conditions shall represent an indistinguishable sample of the probability density function (PDF). Ensemble-based data assimilation techniques incorporate this uncertainty information to formulate the prior likelihood of the model state. Here, the class of Sequential Importance Resampling Filters and Smoothers is of special interest, as the only approximation in their formulation is the ensemble size. This chapter serves the purpose to introduce both, ensemble forecasting and nonlinear ensemble-based data assimilation.

2.1 Uncertainties in Numerical Weather Predictions

Uncertainties in Numerical Weather Prediction (NWP) models are in general divided into two types: initial condition error and model error. This distinction has been proven to be practical, as different approaches to address uncertainty target either one of them. However, it should be stressed that initial condition and model error¹ are in principle inseparable, as model errors project on the analysis uncertainty and vice versa.

Formally, a numerical integration of the NWP model may be written in its simplest form as (*Berner et al.*, 2015)

$$\mathbf{x}^f = \mathbf{x}^a + \int_{t=0}^T \left(\frac{\partial[\mathbf{x}]_{\text{dyn}}}{\partial t} + \frac{\partial[\mathbf{x}]_{\text{param}}}{\partial t} \right) dt, \quad (2.1)$$

¹The distinction between forecast error and model error as defined in *Daley* (1991) is not followed here.

where $\mathbf{x}^f = \mathbf{x}(T)$ denotes the model state forecast, $\mathbf{x}^a = \mathbf{x}(0)$ the model state analysis, t the time and T the forecast length. The contributions to the tendencies from the dynamical core $[\mathbf{x}]_{\text{dyn}}$ and the parameterizations $[\mathbf{x}]_{\text{param}}$ are written separately. In the context of ensemble forecasting, this states the control model and \mathbf{x}^f is denoted as the control, or unperturbed ensemble member forecast, subject to all available knowledge of initial conditions and model formulation. Initial condition error is inherent in \mathbf{x}^a , and model error in $\int_{t=0}^T \left(\frac{\partial[\mathbf{x}]_{\text{dyn}}}{\partial t} + \frac{\partial[\mathbf{x}]_{\text{param}}}{\partial t} \right) dt$.

Accordingly, an ensemble forecast is formulated as a set of model integrations given by

$$\mathbf{x}_i^f = (\mathbf{x}^a + \mathbf{x}_i^{\prime a}) + \int_{t=0}^T \left(\frac{\partial[\mathbf{x}_i]_{\text{dyn}}}{\partial t} + \frac{\partial[\mathbf{x}_i]_{\text{param(stoch)}}}{\partial t} + \frac{\partial[\mathbf{x}_i]_{\text{stoch}}}{\partial t} \right) dt, \quad (2.2)$$

where \mathbf{x}_i^f denotes the model state of the i th ensemble member forecast, with $i \in \{1, \dots, N\}$ and N the ensemble size. Each ensemble member is integrated forward in time starting from a different analysis, subject to perturbation $\mathbf{x}_i^{\prime a}$, with $\mathbf{x}_i^a = \mathbf{x}^a + \mathbf{x}_i^{\prime a}$, where the analysis \mathbf{x}^a of the control member is typically derived by an independent data assimilation system. The contribution from parameterizations are reformulated here as $[\mathbf{x}_i]_{\text{param(stoch)}}$, subject to a possible stochastic perturbation. Possible contributions from additional processes that stem from model uncertainty are denoted as $[\mathbf{x}_i]_{\text{stoch}}$. It is assumed that the dynamical core stays unperturbed in the ensemble formulation.

Errors in the initial conditions stem from a limited observability of the atmospheric state as well as measurement and representativity errors. This uncertainty evolves in the data assimilation process, subject to approximations and restricted knowledge of error covariances. Errors in the model formulation stem from unresolved process appearing on the sub-grid scale, lower boundary forcings and ultimately the numerical formulation of model equation. Table 2.1 summarizes sources of uncertainties and lists further examples.

2.2 Representation of initial condition error

The chaotic behavior of the atmosphere is known to be formulated in its sensitivity to the initial conditions (*Lorenz, 1963*). Ideally, a finite ensemble size selectively samples the probability density function of the initial conditions (or analysis), such that the possible ranges of model outcomes are covered to the best possible extend. Different selective sampling strategies differ in how they estimate the initial condition PDF and how it is sampled.

Table 2.1: Sources of uncertainties in NWP models divided in the categories of initial condition error and model error.

Error category	Sources
Initial conditions	Observations, e.g. <ul style="list-style-type: none"> • restricted observational coverage • measurement error • representativity Data Assimilation, e.g. <ul style="list-style-type: none"> • linearization, Gaussian errors statistic • mapping between model state and observations state • static or limited knowledge of flow dependent background covariance
Model	Unresolved processes, e.g. <ul style="list-style-type: none"> • parameterizations of physical processes • closure assumptions • feedback of energy from unresolved to resolved scales Lower boundary forcings, e.g. <ul style="list-style-type: none"> • land surface parameter: roughness length, albedo, moisture availability, vegetation fraction • sea surface temperature Numerical processing, e.g. <ul style="list-style-type: none"> • numerical diffusion • truncation error and precision

The first operational implementation of ensembles² emerged in the same years at the European Centre for Medium-Range Weather Forecasts (ECMWF, *Buizza et al.* (1999)) and the National Centre for Environmental Prediction (NCEP, *Toth and Kalnay* (1997)). Their undisputed success arose by the notion, that perturbations in the initial conditions exhibit different growth rates, favored by the underlying atmospheric flow. In other words, the sampling of the analysis PDF shall be confined to the subspace of the fastest growing perturbations. Different selective sample strategies have been proposed to identify such perturbations which are most relevant for the model dynamic.

The singular vector (SV) method (*Buizza et al.*, 1999) samples the analysis PDF by perturbations which possess the largest linear growth rates over a fixed optimization interval. The directions of such perturbations are given by the singular vectors of the tangent linear model, which are computed by applying the tangent linear model forward and backward. In practice, this equals an iteratively solution of an eigenproblem given by the tangent linear model and its adjoint, with respect to a predefined

²At the same time, the Meteorological Service of Canada (MSC) implemented an 16-member ensemble based on the Perturbed Observation (PO) method (*Houtekamer et al.*, 1996). Due to its limited success, it is not further discussed.

energy norm. The SV method owes its success to a rapid error growth, which maximizes ensemble spread and various probabilistic skill scores. Analysis perturbations \mathbf{x}_i^a are given by a linear combination of singular vectors for the norther hemisphere, southern hemisphere and the tropics, each scaled by analysis error estimates (*Molteni et al.*, 1996). The ensembles generated in this work have been finally initialized by a different ensemble system (see the discussion of Section 4.1.2), and therefore, the SV method is not discussed in further detail.

The breeding vector (BV) method (*Toth and Kalnay*, 1993, 1997) has been established at NCEP and is based on the notion, that uncertainties in the analysis are dominated by short-range forecast errors. Initial random perturbations added to the analysis \mathbf{x}^a will, after a sufficient number of assimilation cycles, grow into the direction of the leading local Lyapunov vectors of the dynamical system³. If perturbations are rescaled at the end of each assimilation cycle, growing perturbations will amplify with respect to nongrowing perturbations. The BV method comes with no further expense and may be divided in the following steps: (i) a random perturbation is added initially to the analysis, (ii) the model is integrated forward in time from the perturbed and control analysis, (iii) the difference of both is formed and rescaled, (iv) the negative and positive (paired) of this difference define the new perturbations added to the control analysis. This procedure is cycled, restarting from (ii). It is stressed again, that any initial random perturbation introduced at (i) will evolve according to the local stability properties of the underlying atmospheric flow. The breed vectors evolve according to the full nonlinear model, unlike the singular vectors which are computed under the assumption of linearity. The BV method samples the analysis PDF by the breeding vectors, and the analysis perturbations can be written as

$$\mathbf{x}_i^a = \mathbf{R}(\lambda, \phi, t)\mathbf{x}_i^{\text{BV}}, \quad (2.3)$$

where \mathbf{x}_i^{BV} denotes the i th breeding vector and \mathbf{R} defines a regional rescaling factor, with λ the latitude and ϕ the longitude, which modifies the perturbation amplitude according to a spatial climatological difference in the analysis error variance derived by two different perturbed data assimilation cycles (*Toth and Kalnay*, 1997).

This investigation makes use of initial condition perturbations constructed by the Ensemble Transform method with rescaling (ETR, (*Wei et al.*, 2008)), which is thereupon introduced more formally. The ETR method is based on the notion, that breeding is suboptimal in that it is firstly, not consistent with the data assimilation system, as perturbations do not project on the analysis error variance and the regional

³Local Lyapunov vectors define the directions in which random perturbations will grow. Their associated Lyapunov exponents determines convergence or divergence of neighboring points in the phase space and hence, characterize the system's stability (within the linear approximation). The BV method is a "nonlinear generalization of the method used to construct Lyapunov vectors" (*Kalnay*, 2003). For a review, the reader is referred to *Kalnay* (2003) and *Strogatz* (2014).

rescaling mask is constructed based on a climatology, and secondly, perturbations do not effectively cover the possible degree of freedom as they are not necessarily orthogonal, but paired. The ETR method can be understood as an extension of the BV method, which addresses both shortcomings.

In the Ensemble Transform (ET), analysis perturbations are obtained by

$$\mathbf{X}'^a = \mathbf{X}'^f \mathbf{T}, \quad (2.4)$$

where $\mathbf{X}'^f = \{\mathbf{x}'_1^f, \dots, \mathbf{x}'_N^f\}$ and $\mathbf{X}'^a = \{\mathbf{x}'_1^a, \dots, \mathbf{x}'_N^a\}$ are $K \times N$ matrices with the i th column given by $\mathbf{x}'_i^f = \mathbf{x}'_i - \bar{\mathbf{x}}^f$ and $\mathbf{x}'_i^a = \mathbf{x}'_i - \mathbf{x}^a$, respectively, and K the model state dimension. The ensemble mean forecast is defined as $\bar{\mathbf{x}}^f = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i^f$. The transformation matrix \mathbf{T} is chosen such that initial conditions in the ensemble shall be restrained explicitly by the analysis error covariance

$$\mathbf{T}^T (\mathbf{X}'^{fT} \mathbf{P}_a^{-1} \mathbf{X}'^f) \mathbf{T} = \mathbf{I}. \quad (2.5)$$

The matrix \mathbf{P}_a is assumed to be diagonal and contains analysis error variances, estimated by an assimilation system. Equation (2.5) ensures, that the variance of the analysis perturbations \mathbf{X}'^a equals that of \mathbf{P}_a , under the assumption that \mathbf{X}'^f reflects the real forecast variances. The transformation matrix \mathbf{T} is obtained by solving the eigenvalue problem

$$\mathbf{X}'^{fT} \mathbf{P}_a^{-1} \mathbf{X}'^f = \mathbf{C} \mathbf{\Gamma} \mathbf{C}^{-1}, \quad (2.6)$$

where the orthogonal eigenvectors \mathbf{c}_i are listed column-wise in matrix \mathbf{C} and the eigenvalues λ_i are listed in descending order in the diagonal matrix $\mathbf{\Gamma}$, with $i \in \{1, \dots, N\}$. *Wei et al.* (2008) shows that only the first $N - 1$ eigenvalues are non zero. Thereupon, a matrix \mathbf{G} may be defined, which sets the N th eigenvalue to a non zero constant, and the transformation matrix \mathbf{T} is derived as

$$\mathbf{T} = \mathbf{C} \mathbf{G}^{-1/2}. \quad (2.7)$$

Analysis perturbations of the ET are orthogonal under a norm defined by the inverse of analysis error variance \mathbf{P}_a , however, they are not centered, which might degrade the ensemble mean forecast. To ensure that $\sum_{i=1}^N \mathbf{x}'_i^a = 0$ holds, a simplex transformation is applied and the final transformation matrix reads

$$\mathbf{T}_p = \mathbf{C} \mathbf{G}^{-1/2} \mathbf{C}^T. \quad (2.8)$$

The simplex transformation preserves the analysis covariance, but a finite number of ensemble member becomes quasi-orthogonal. The ETR method has proven to outperform the simple breeding method in various probabilistic skill scores (*Wei et al.* (2008)).

2.3 Representation of model error

Meteorological ensembles are well known to overestimate the atmospheric predictability (*Buizza et al.*, 1999), which appears most prominently in convection-permitting ensembles of the short-range (*Romine et al.*, 2014). Even if one may assume perfect knowledge of the analysis distribution, the resulting ensemble still exhibits underdispersiveness (*Palmer et al.*, 2005; *Wilks*, 2005). The inherent uncertainty in model formulation cannot be neglected in general, which may limit atmospheric predictability to the same extent as initial condition error. This is especially true for severe weather events, e.g. explosive cyclogenesis (*Mullen and Baumhefner*, 1988).

Model errors are complex in their nature and far from being fully understood. They originate mainly from processes which appear on the sub-grid scale. Different stochastic parameterizations have been proposed and proven to increase the ensemble spread, while maximizing the ensemble reliability (*Berner et al.*, 2011; *Christensen et al.*, 2015). Their beneficial impact has also been reported for ensemble data assimilation (*Ha et al.*, 2015), as model errors affect the uncertainty introduced by initial conditions and vice versa. Multiple types of approaches exist, since model errors stem from many different sources. Most of the operational schemes are formulated as stochastic reinterpretations of deterministic parameterizations, and are somewhat *ad hoc*, as they are based on empirical assumptions which are formulated *a priori*. However, the merit of existing schemes goes beyond simply increasing ensemble spread, which is in general feasible to be addressed by postprocessing, but to trigger possible instabilities in the underlying flow to represent the range of possible outcomes. Surely, the final goal to be accomplished are true stochastic parameterizations which account for uncertainty where it actually appears.

Here, the description is restricted to the Stochastically Perturbed Parameterization Tendency (SPPT) scheme and the Stochastically Kinetic Energy Backscatter (SKEB) scheme, as both have been implemented in the WRF model following closely their formulation in the ECMWF EPS ensemble. SPPT represents sub-grid scale variability of parameterizations by sampling the net parameterized physics tendencies around their deterministic value. In contrast, SKEB accounts for spurious model dissipation by stochastically injecting energy across various spatial scales. Despite their numerous shortcomings (*Shutts*, 2015), both schemes have become broadly established in ensembles of global circulation models (*Berner et al.*, 2009; *Palmer et al.*, 2009; *Charron et al.*, 2010; *Sanchez et al.*, 2016) as well as mesoscale models (*Berner et al.*, 2011; *Tennant et al.*, 2011; *Bouttier et al.*, 2012; *Berner et al.*, 2015; *Shutts*, 2015). Their complementarity has been proven in numerous studies (*Romine et al.*, 2014; *Jankov et al.*, 2017), as the forcing introduced by SPPT is the largest in the tropics and in the planetary boundary layer, and by SKEB in the free atmosphere

and in case of strong cyclogenesis.

2.3.1 Stochastically Perturbed Parameterization Tendency scheme

By definition, physical parameterizations are estimations of unresolved atmospheric processes appearing on the sub-grid and/or evolving by processes which are insufficiently understood. Even parameters within parameterizations, arising e.g. from closure assumptions or estimations of probabilistic mean values, possess a significant uncertainty. Their feedback on the resolved flow is sometimes understood as an ensemble mean of plausible ranges consistent with the resolved-scale forcing, e.g. for cumulus parameterization (*Arakawa and Schubert, 1974*). This assumption is condemned to fail with the ever increasing resolution of atmospheric models, and in one way or the other, a transition to a stochastic reformulation of parameterizations is inevitable.

The Stochastically Perturbed Parameterization Tendencies (SPPT) scheme does so in a straight forward manner by introducing univariate multiplicative noise to the net parameterized physics tendencies and reformulating their contribution to the total local tendency at each grid point as a sample of a probability density function, which is centered on their deterministic value. With the nomenclature introduced for (2.1), the tendency equation for the prognostic variables $\mathbf{x} \in \{u, v, q, T\}$, with u and v the wind vector components, q the humidity and T the temperature, may be written as

$$\frac{\partial[\mathbf{x}]_{\text{dyn}}}{\partial t} + \frac{\partial[\mathbf{x}]_{\text{param}}}{\partial t}, \quad (2.9)$$

and a perturbed parameterization tendency is then defined as

$$\frac{\partial[\mathbf{x}]_{\text{dyn}}}{\partial t} + \frac{\partial[\mathbf{x}]_{\text{param(stoch)}}}{\partial t} = \frac{\partial[\mathbf{x}]_{\text{dyn}}}{\partial t} + (1 + r(x, y, z, t)) \frac{\partial[\mathbf{x}]_{\text{param}}}{\partial t}, \quad (2.10)$$

where r is a pattern which imposes spatial and temporal correlation on perturbations. Formulations of SPPT in various ensemble systems differ essentially by the choice of r . Here, it is restricted to the formulation as implemented in WRF, following *Berner et al. (2015)*.

Perturbations are drawn from a truncated Gaussian noise process in the range $r \in [-1, 1]$, with prescribes standard deviation η . Variations in the vertical are omitted to retain conservation laws imposed by the parameterizations, yet no instabilities in the planetary boundary layer have been reported for the WRF model. For the $X \times Y$ horizontal grid, the stochastic pattern r is defined in spectral space as

$$r(x, y, t) = \sum_{k=-K/2}^{K/2} \sum_{l=-L/2}^{L/2} r_{k,l}(t) e^{2\pi i(kx/X + ly/Y)}, \quad (2.11)$$

with Fourier modes $r_{k,l}$ depending on wavenumbers k, l in zonal x and meridional y direction of WRF's rectangular grid, respectively. A first-order autoregressive process imposes temporal correlation

$$r_{k,l}(t + \Delta t) = (1 - \alpha)r_{k,l}(t) + t_{k,l}\sqrt{\alpha}\epsilon_{k,l}(t), \quad (2.12)$$

such that small perturbations are associated with a short temporal range and vice versa. This describes a Markov process, $(1 - \alpha)$ is the autoregressive parameter with $\alpha = \Delta t/\tau \in (0, 1]$ and τ the decorrelation time. $\epsilon_{k,l}$ is a complex Gaussian white noise process with zero mean and covariance $\langle \epsilon(s)_{k,l}\epsilon^*(t)_{m,n} \rangle = \sigma^2\delta_{k,m}\delta_{l,n}\delta_{s,t}$, with σ^2 set to one. The variance spectrum of the first-order autoregressive process is determined by the amplitudes $t_{k,l}$, which are defined such that they yield an autocorrelation according to a Gaussian on a plane:

$$t_{k,l} = F_0\exp(-4\pi\kappa\rho_{k,l}^2), \quad (2.13)$$

with $\rho = \sqrt{k^2/X^2 + l^2/Y^2}$ the radial wavenumber and κ the spatial correlation length, as derived by *Weaver and Courtier* (2001). F_0 denotes a normalization factor that depends linearly on the spectral variance $\eta_{k,l}^2$:

$$F_0 = \left(\frac{\eta_{k,l}(2\alpha - \alpha^2)}{2 \sum_k \sum_l \exp(-8\pi\kappa\rho_{k,l}^2)} \right). \quad (2.14)$$

The stochastic pattern's quantities determine directly the forcing of the scheme, with the tuning parameters of standard deviation η as well as temporal and spatial correlation τ and κ , respectively. The default values in the WRF model are $\eta = 0.5$, with cutoff tails above 2.0, $\tau = 21600$ s and $\kappa = 150$ km.

Instead of a univariate formulation, one may think of perturbing each physics tendency independently, which was the initial ansatz proposed by *Buizza et al.* (1999) in the ECMWF EPS ensemble. However, multivariate perturbations are less consistent with the model physics, decrease the skill of tail distribution statistics and generate gravity waves (*Palmer et al.*, 2009). The system tends to be pushed from its preferred attractor too often, which may exceed the range of model uncertainty.

The forcing imposed by SPPT is flow-dependent, as the perturbation magnitude scales with the accumulated tendencies. However, the probability distribution of certain parameterizations varies strongly depending on the geographical location or height. Further, uncertainties evolving in the vertical, e.g. the shape of a momentum profile in the planetary boundary layer, can not be represented.

2.3.2 Stochastic Kinetic Energy Backscatter scheme

The Stochastic Kinetic Energy Backscatter (SKEB) scheme was originally developed in the context of large-eddy simulations (*Mason and Thomson, 1992*) to account for unresolved energy transfer from the sub-grid scales to the resolved scales, known as an inverse energy cascade. Parts of the dissipated energy are reinjected stochastically as kinetic energy to mimic an energy spectra derived by the counterpart of a direct numerical simulation (*Frederiksen and Davies, 2004*). This idea was exported to NWP ensembles (*Shutts, 2005; Berner et al., 2009*), motivated by the excessive energy dissipation over various scales (*Nastrom and Gage, 1985; Palmer et al., 2009*). Net energy sinks arise from parameterizations as well as the dynamical core, in particular from unbalanced motions of orographic wave drag, convection and gravity waves as well as numerical dissipation. Hence, in the context of NWP models, the SKEB scheme is not confined to the vicinity of the truncation scale, but energy is even injected upscale to the subsynoptic and synoptic scale. Coarse-graining experiments support this assumption by identifying energy sinks across the entire wavenumber spectrum (*Palmer et al., 2009; Shutts, 2013*). Their high-resolution counterparts are used to tune the SKEB in a way to correct the energy spectrum accordingly, yet in a heuristic fashion.

The SKEB scheme introduces additive noise to the tendency equation (2.1), such that

$$\frac{\partial[\mathbf{x}]_{\text{dyn}}}{\partial t} + \frac{\partial[\mathbf{x}]_{\text{param}}}{\partial t} \quad (2.15)$$

may be rewritten as

$$\frac{\partial[\mathbf{x}]_{\text{dyn}}}{\partial t} + \frac{\partial[\mathbf{x}]_{\text{param}}}{\partial t} + \frac{\partial[\mathbf{x}]_{\text{SKEB}}}{\partial t}. \quad (2.16)$$

Formulations of SKEB in various ensemble systems differ by the choice of prognostic variables subject to perturbation, i.e. $[\mathbf{x}]_{\text{SKEB}} \subseteq \{u, v, T\}$, the dissipative sources which are associated with such perturbations and the stochastic pattern r . Here, it is yet again restricted to the formulation as in the WRF model, following *Berner et al. (2011)*.

A streamfunction tendency forcing $\psi'(x, y, t) := \partial\psi(x, y, t)/\partial t$ and a potential temperature tendency forcing $\theta'(x, y, t) := \partial\theta(x, y, t)/\partial t$ are introduced to inject a domain averaged perturbation kinetic energy $E'_{\text{kin}} = \Delta E_{\text{kin}}/\Delta t$ and perturbation potential energy $E'_{\text{pot}} = \Delta E_{\text{pot}}/\Delta t$ at each time step Δt . The associated forcings are defined in two-dimensional spectral space:

$$\psi'(x, y, t) = \sum_{k=-K/2}^{K/2} \sum_{l=-L/2}^{L/2} \psi'_{k,l}(t) e^{2\pi i(kx/X + ly/Y)}, \quad (2.17)$$

$$\theta'(x, y, t) = \sum_{k=-K/2}^{K/2} \sum_{l=-L/2}^{L/2} \theta'_{k,l}(t) e^{2\pi i(kx/X + ly/Y)}, \quad (2.18)$$

with the nomenclature of the Fourier expansion already introduced in the previous section. Perturbations of the wind vector are confined to the rotational part to preserve the dynamical balance between pressure and wind, with $u'(x, y, t) = -\frac{\partial \psi'(x, y, t)}{\partial y}$ and $v'(x, y, t) = \frac{\partial \psi'(x, y, t)}{\partial x}$ the zonal and meridional wind tendency forcings, respectively.

Finite temporal correlations are imposed by evolving each spectral coefficient according to a first-order autoregressive process:

$$\psi'_{k,l}(t + \Delta t) = (1 - \alpha_\psi) \psi'_{k,l}(t) + g_{k,l} \sqrt{\alpha_\psi} \epsilon_{k,l}(t), \quad (2.19)$$

$$\theta'_{k,l}(t + \Delta t) = (1 - \alpha_\theta) \theta'_{k,l}(t) + h_{k,l} \sqrt{\alpha_\theta} \epsilon_{k,l}(t). \quad (2.20)$$

This process has already been described in-depth in the previous section by (2.12). Here, $\alpha_{\psi/\theta} = \Delta t / \tau_{\psi/\theta} \in (0, 1]$ again denotes the autoregressive parameter, with $\tau_{\psi/\theta}$ the decorrelation time. In the WRF model, temporal and spatial constant dissipation rates are assumed and hence, quantities of the first-order autoregressive process will directly transfer to the effective perturbations. This appears as a somewhat drastic simplification. However, *Berner et al.* (2011) shows that results are quite similar compared to a flow-dependent dissipation rate, as estimations of the true dissipation rate remain a challenge. One aims to prescribe the injected energies with a given power spectrum, yielding noise amplitudes of the form

$$g_{k,l} = b \rho_{k,l}^\beta, \quad (2.21)$$

$$h_{k,l} = f \rho_{k,l}^\gamma. \quad (2.22)$$

The amount of backscattered dissipated energy determines the amplitudes b and f of the forcings. For the streamfunction, an energy backscatter rate B_ψ injects a total kinetic energy ΔE_{kin} per unit mass into the full flow during a numerical time step Δt :

$$B_\psi = \frac{\Delta E_{\text{kin}}}{\Delta t} = \frac{2\pi^2}{\Delta t} \sum_{k=-K/2}^{K/2} \sum_{l=-L/2}^{L/2} \rho_{k,l}^2 \langle |\psi_{k,l}(t + \Delta t)|^2 - |\psi_{k,l}(t)|^2 \rangle \quad (2.23)$$

$$= \frac{2\pi^2 \sigma_\psi^2 \Delta t}{\alpha_\psi} \sum_{k=-K/2}^{K/2} \sum_{l=-L/2}^{L/2} \rho_{k,l}^{2\beta+2} b^2. \quad (2.24)$$

Solving for b yields:

$$b = \left(\frac{B_\psi \alpha_\psi}{2\pi \sigma_\psi^2 \Gamma_\psi \Delta t} \right)^{\frac{1}{2}}, \quad \text{with } \Gamma_\psi = \sum_{k=-K/2}^{K/2} \sum_{l=-L/2}^{L/2} \rho_{k,l}^{2\beta+2}. \quad (2.25)$$

This implies, that a streamfunction forcing with power law $\rho_{k,l}^\beta$ will result in a kinetic energy spectrum with power law $\rho_{k,l}^{2\beta+2}$ (in radial wavenumber). Advancing in the same manner for the potential temperature, an energy backscatter rate B_θ injects during a numerical time step Δt a total potential energy ΔE_{pot} per unit mass into the full flow:

$$B_\theta = \frac{\Delta E_{\text{pot}}}{\Delta t} = \frac{c_p}{2\theta_0 \Delta t} \theta^2 = \frac{c_p}{\theta_0 \Delta t} \sum_{k=-K/2}^{K/2} \sum_{l=-L/2}^{L/2} \langle |\theta_{k,l}(t + \Delta t)|^2 - |\theta_{k,l}(t)|^2 \rangle \quad (2.26)$$

$$= \frac{c_p \sigma_\theta^2 \Delta t}{\theta_0 \alpha_\theta} \sum_{k=-K/2}^{K/2} \sum_{l=-L/2}^{L/2} \rho_{k,l}^{2\gamma} f^2, \quad (2.27)$$

with $\theta_0 = 300$ K a reference potential temperature used in the WRF model and $c_p = 1004$ J/K the specific heat capacity. Solving for f yields

$$f = \left(\frac{B_\theta \alpha_\theta \theta_0}{c_p \sigma_\theta^2 \Gamma_\theta \Delta t} \right)^{\frac{1}{2}}, \quad \text{with } \Gamma_\theta = \sum_{k=-K/2}^{K/2} \sum_{l=-L/2}^{L/2} \rho_{k,l}^{2\gamma}. \quad (2.28)$$

This implies, that a potential temperature forcing with power law $\rho_{k,l}^\gamma$ will result in a kinetic energy spectrum with power law $\rho_{k,l}^{2\gamma}$ (in radial wavenumber). The derivation of (2.24) is given in the appendix of *Berner et al.* (2009) (in spherical coordinates for a global circulation model), and (2.27) is derived analogously.

In the WRF model, the SKEB scheme is fully controlled by the tuning parameters $\beta, \gamma, \sigma_\psi^2, \sigma_\theta^2, \tau_\psi, \tau_\theta, B_\psi, B_\theta$. Default values are $\beta = \gamma = -1.83$, which results in a perturbation kinetic energy spectrum of $-5/3$ and in a perturbation potential energy spectrum of $-10/3$. The noise variance is set to $\sigma_\psi^2 = (1/12)\alpha_\psi$, $\sigma_\theta^2 = (1/12)\alpha_\theta$ and decorrelation times are $\tau_\psi = \tau_\theta = 10800$ s. The backscattered energy rates are chosen as $B_\psi = 10^{-5} \text{ m}^2/\text{s}^3$ and $B_\theta = 10^{-6} \text{ m}^2/\text{s}^3$. For comparison, *Berner et al.* (2009) reports an annual global mean dissipation rate in the ECMWF model of $1.99 \text{ W}/\text{m}^2$ for deep convection. With an air density of $1 \text{ kg}/\text{m}^3$ and an air column of 1 km height, this yields $1.99 \times 10^{-3} \text{ m}^2/\text{s}^3$ per unit mass. With the remaining sources approximately of the same order, and assuming $1/10$ of the dissipated energy being backscattered (according to *Palmer et al.* (2009)), a value of $10^{-3} \text{ m}^2/\text{s}^3$ may be estimated. The choice of backscatter rates in the WRF model is therefore rather conservative.

Current implementations of the SKEB schemes in operational ensemble systems

do not come without criticism. *Thuburn et al. (2013)* and *Shutts (2013)* conclude, that at least parts of the energy sinks are likely to be of systematic nature and should rather be fixed in a deterministic manner. Further, with increasing model resolution, perturbations across the whole wavenumber range used by SKEB scheme will most likely not be justifiable any more. However, as such issues remain unresolved, the SKEB scheme states an attractive approach to account for the excessive energy dissipation problem by a probabilistic approach, while, at the same time, increasing ensemble spread and improving probabilistic skill scores.

2.4 The Sequential Importance Resampling Filter

The survey of fully nonlinear data assimilation in atmospheric science has drawn large attention to Sequential Monte Carlo techniques, or simply the particle filter. Particle filters belong to the class of recursive Bayesian estimators and solve the full data assimilation problem, with the only approximation being the limited ensemble size. For a comprehensive review, the reader is referred to *van Leeuwen (2009)*.

The particle filter relates a posterior weights to ensemble members, which estimate the likelihood of each model state given the observations. The prior density of the model state \mathbf{x} may be formulated as the sum of δ -functions, centered around each ensemble member model state:

$$p(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \delta(\mathbf{x} - \mathbf{x}_i), \quad (2.29)$$

with N the ensemble size. By applying Bayes' theorem, we find that the a posteriori density of the model state, given observations \mathbf{y} , reads

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \quad (2.30)$$

Using the particle filter expression for the prior density, it follows

$$p(\mathbf{x}|\mathbf{y}) = \sum_{i=1}^N w_i \delta(\mathbf{x} - \mathbf{x}_i), \quad (2.31)$$

where the weights w_i are given by

$$w_i = \frac{p(\mathbf{y}|\mathbf{x}_i)}{\sum_{j=1}^N p(\mathbf{y}|\mathbf{x}_j)}. \quad (2.32)$$

This formulation is known as Sequential Importance Sampling (SIS), or simply the basic particle filter. However, especially for high dimensional systems like the atmosphere, the ensemble size is very much limited and the basic particle filter tends to

degenerate. A single member is assigned with all the importance weight, with the statistical information of the system being lost. *Snyder et al.* (2008) shows, that the ensemble size has to grow exponentially with the variance of the observation log likelihood $\tau^2 = \text{var}[\log(p(\mathbf{y}|\mathbf{x}))]$ to prevent filter degeneracy. Since τ^2 is proportional to the dimension of the state vector, the basic Particle Filter is unfeasible in the realm of high dimensional systems.

Sequential Importance Resampling provides a way to reduce the variance among the ensemble member. Resampling proceeds from the weighted approximate posterior density by (2.31), and draws N samples to form an unweighted posterior density

$$\hat{p}(\mathbf{x}|\mathbf{y}) = \sum_{i=1}^N w_i \frac{n_i}{N} (\mathbf{x} - \mathbf{x}_i), \quad (2.33)$$

where n_i denotes the duplication count of particle \mathbf{x}_i , with $N = \sum_{i=1}^N n_i$. Different resampling algorithms exist, a general overview is given in *Douc et al.* (2005)). Here, the description is restricted to residual resampling (*Lui and Chen*, 1998) and stratified resampling (*Kitagawa*, 1998), as both are implemented consecutively in ESIAS-met.

Residual resampling draws

$$n'_i = \max\{n'_i \in \mathbb{Z}_+ \mid n'_i < Nw_i\} \quad (2.34)$$

copies of each particle \mathbf{x}_i . The remaining $m = N - \sum_{i=1}^N n'_i$ particles are each assigned with a probability $w'_i = Nw_i - n'_i$. It remains to draw n''_i copies of each particle \mathbf{x}_i , such that $N = \sum_{i=1}^N (n'_i + n''_i)$, using a stochastic resampling algorithm. This is realised by stratified resampling, in which an interval $I = [0, 1]$ is partitioned into m disjoint sets $I := \dot{\cup}_{i \in m} I_i$, with the length of I_i according to weight w'_i :

$$I_i = \left(\sum_{j=1}^{i-1} w'_j, \sum_{j=1}^{i-1} w'_j + w'_i \right], \quad \text{for } i = 1, \dots, m. \quad (2.35)$$

A random number is drawn from the uniform distribution on $[0, 1]$, and starting from this random number, m equally sized disjoint sets $L = \dot{\cup}_k L_k$ of length $1/m$, with $k = 1, \dots, m$, are overlayed on interval I . If the end of L_k lies in I_i , n''_i is incremented by one. If $n'_i + n''_i > 1$, multiple copies of the same model state are subject to different model error representation in upcoming assimilation cycle.

Resampling may become computational costly, especially in the vicinity of large ensemble sizes, as an increasing number of comprehensive model states has to be duplicated. It is a common procedure in geophysical applications to set a fixed resampling interval. However, this appears suboptimal, since the variance among the ensemble members depends on the predictability of the underlying flow, and not on

the forecast time. Adaptive resampling provides a means to monitor flow dependent variance to perform a resampling step only when it is needed. For this purpose, an effective sample size is taken as

$$N_{eff} = \frac{1}{\sum_{i=1}^N w_i^2}. \quad (2.36)$$

Sequential Importance Resampling is essential for the performance of particle filter methods. However, filter degeneracy is not cured. The ensemble is integrated forward in time from a posterior that has been disposed from ensemble members with low importance weight, but the quality of the posterior itself remains unaffected.

Chapter 3

ESIAS – Ensembles for Stochastic Integration of Atmospheric Systems

This chapter describes the appropriate software developments for generating ultra large meteorological ensembles on a high-performance computer. These developments are part of the software Ensembles for Stochastic Integration of Atmospheric Systems (ESIAS). ESIAS is designed to control an ensemble implementation of the Weather Research and Forecasting (WRF) model, denoted as ESIAS-met, with the dynamical correspondance to ESIAS-chem (*Franke, 2018*), an ensemble version of the European Air Pollution and Dispersion Inverse Model EURAD-IM (*Elbern et al., 2007*). The system aims to investigate short to medium range probabilistic limited area forecasts as well as emission and parameter estimation with the integration of novel algorithms in the realm of data assimilation and ensemble generation.

The WRF model is described to the extend of providing the necessary background for the software developments to follow. ESIAS has been developed on a Blue Gene supercomputer, which is introduced, and a description about porting and tuning the WRF model is provided. A computational analysis reveals the necessity for code modifications of WRF's model uncertainty schemes. As the main objective of this chapter, a computational efficient system controlling numerous WRF ensemble members is introduced. This system states a complete novel software approach to execute ensemble forecasts concurrently and is in particular favorable for Sequential Importance Resampling with large ensemble sizes. The chapter closes with a proof of concept to demonstrate the parallel scalability of the system.

3.1 The Weather Research and Forecasting (WRF) model

The underlying NWP model of the ultra large ensemble control system ESIAStmet is the Weather Research and Forecasting (WRF) Model (*Skamarock et al.*, 2008). WRF is a limited area NWP model for nonhydrostatic modelling. Since the initial release by the end of 2000, WRF has grown to a community model and is used extensively for research and operational real-time forecasting at numerous research organizations and the private sector throughout the world.

The WRF model arose as a successor of the MM5 Model (*Grell et al.*, 1994) and has been developed jointly by the National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (NOAA, which has been represented by the National Centers for Environmental Prediction (NCEP)), the Air Force Weather Agency (AFWA), the Naval Research Laboratory, the University of Oklahoma, and the Federal Aviation Administration (FAA), all in the USA. The WRF modeling system consists of multiple software components, ranging from the WRF Preprocessing System (WPS) and the WRF model itself to the WRF Data Assimilation System (*Barker et al.*, 2004; *Huang et al.*, 2009) as well as associated postprocessing utilities. A detailed description of the complete WRF modeling system can be found in *Skamarock et al.* (2008).

At the beginning of every forecast with the WRF modeling system is the preprocessing of terrestrial and meteorological input data with WPS, which is divided in multiple steps by associated programs. Within the *ungrib* program, gridded meteorological analysis and forecast fields, typically originating from global circulation models, are encoded, uncompressed and filtered. Terrestrial data is processed, projected and horizontally interpolated on the user defined domain within the *geogrid* program. The *metgrid* program merges the output from the *ungrib* and *geogrid* program by interpolating horizontally the meteorological input data onto the model domain. Vertical interpolation is done by the *real* program, which provides the meteorological initial and boundary fields for the dynamical solver. This procedure is depicted in Figure 3.1.

Two different dynamic solvers are available in the WRF modeling system, from which the Advanced Research WRF (ARW) solver version 3.7.1 is used in this work. Within the ARW solver, sophisticated numerical methods are employed. The fully compressible nonhydrostatic Euler equations are solved, which include the conservations of momentum, mass, energy and water vapor as well as the prognostic geopotential equation along with the diagnostic relations for the full pressure (vapor and dry air) and the dry inverse density. The vertical coordinate is a terrain-following hydrostatic pressure coordinate (*Laprise*, 1992), commonly know as σ -coordinates, and

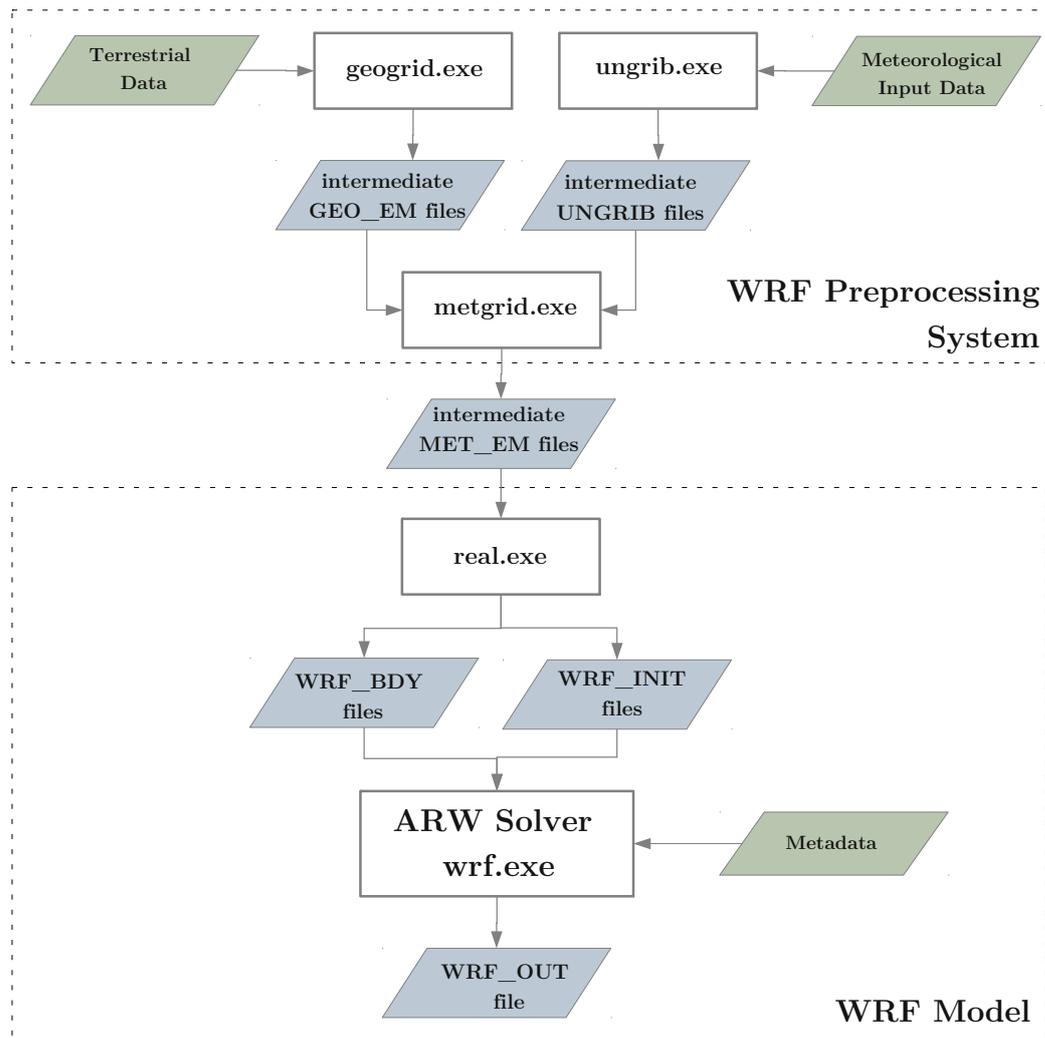


Figure 3.1: Flowchart of the WRF modeling system with the Preprocessing System (WPS) and the Advanced Research WRF (ARW) solver. Green parallelograms indicate external input data and blue parallelograms model output and intermediate files.

vertical mesh refinement is enabled. Spatial discretization is realized by the finite difference scheme using an Arakawa C-grid staggering (*Arakawa and Lamb, 1977*), where scalar values are defined at mass points and half σ -levels (except the geopotential at full σ -levels) and the wind vector components are staggered. Up to 6th-order centered and upwind-biased advection schemes are available in the horizontal and vertical direction. The time integration is realized by a 3rd-order Runge-Kutta scheme for the low-frequency modes and a time-split integration scheme is applied for the high-frequency acoustic and gravity-wave modes (*Klemp et al., 2007*). One-way and two-way nesting is possible with multiple nesting levels. Multiple projections are provided to map the earth's sphere on the Cartesian grid's plane.

3.2 The supercomputer JUQUEEN

JUQUEEN (*Jülich Supercomputing Centre*, 2015) is a supercomputer hosted by the Jülich Supercomputing Centre (JSC). With a peak performance of 5.9 Petaflops, JUQUEEN has been ranked at position 7 in the Top500 list of the most powerful supercomputer in June 2013. The IBM® Blue Gene/Q system consists of 28 racks with each 2 midplanes à 16 nodeboards. In turn, each nodeboard consists of 32 compute nodes and each compute node consists of 16 processors (with 16 GB memory per node), which gives 458,752 processors in total. The IBM PowerPC® A2 processors may execute 4 processes or threads in parallel (fourfold Simultaneous MultiThreading, SMT) and four double-precision Single Instruction Multiple Data (SIMD) can be realized. Each processor has a clock speed of 1.6 GHz, which is relatively slow in comparison to other processors on which atmospheric models run routinely, e.g. the Intel® Xeon E5-2697v4 18C processors of *Cheyenne*, NCAR, with 2.3 GHz or the Intel® Xeon E5-2690v3 12C processors of *Luna* and *Surge*, respectively, NOAA, with 2.6 GHz. Thus, application codes on JUQUEEN have to be highly parallelized to run efficiently, which is a crucial property of code development described in this chapter. The network consists of a 5D Torus with 40 Gbytes/s bandwidth between two compute nodes and a latency of 2 μ s. The maximum I/O bandwidth is determined by the number of allocated processors, with eight I/O nodes assigned to a single rack and a maximum bandwidth of 2 Gbytes/s each. A detailed description of the BLUE GENE/Q system can be found in *Chen et al.* (2012) and *Vezolle* (2012).

3.3 Computational performance of the WRF model on JUQUEEN

3.3.1 Porting and Tuning

WRF has been built on the Blue Gene/Q architecture of JUQUEEN with the IBM® XL compilers. A compilation with the open source GNU compilers has also been tested, yet rejected due to poorer performance. WRF has further been built with plain MPI support as well as a hybrid MPI + OpenMP configuration. Depending on the ratio of domain size to allocated processors, i.e. the size of decomposed subdomains, a hybrid configuration may perform moderately better only in some cases. A plain MPI parallelization was therefore chosen to demonstrate the code performance and a hybrid configuration is not further discussed.

Finding the most suitable compiler instructions for code optimization is a crucial task for any code application on high-performance computers and requires detailed sensitivity tests. Although the WRF software environment provides already compiler instructions for an IBM Blue Gene system with XL compilers, such instructions have

to be refused since simulation results show a dependence on the processor number. Therefore, an optimization level of `-O2` is taken as initial point in the following discussion.

An optimization level of `-O3` is in principle applicable to the WRF code, yet higher order transformation of loops has to be disabled. The option `-qsimd=noauto` disables the conversion of loop array operations into vector instructions and has a positive impact on the central processing unit (CPU) time. Further, the function level of WRF is dominated by slow intrinsic mathematical functions and linking with the IBM MASS library is not effective since the library contains different names of entry points. By adding `-qstrict=nolibrary` to the compiler instructions, the XL compiler does not change the names of floating point routines and faster alternatives can be used. The suboption `-qnohot=noarraypad:level=2:novector:fastmath` enables fast scalar versions of math functions instead of the default. Reciprocal and square root functions need special considerations on Blue Gene/Q. Scalar and SIMD estimate instructions (e.g. Newton's approximation method) with high throughput exist, and their utilization can be enabled by adding `-qdebug=recipf:forcesqrt`. In summary, the final compiler optimization instructions read:

```
-O3 -qnohot=noarraypad:level=2:novector:fastmath -qstrict=nolibrary
-qdebug=recipf:forcesqrt -qsimd=noauto -qarch=qp -qtune=qp.
```

Table 3.1 compares CPU times with initial and improved compiler optimization for varying processor numbers for a fixed problem size of $330 \times 330 \times 50$ grid points and a forecast time of three hours (the fixed problem size defines the outer domain of the nesting procedure used throughout this work and introduced in detail in Section 4.1.1). The reduction in CPU time with improved compiler optimization is up to 100%. Figure 3.2 shows the corresponding parallel scaling with improved compiler optimization. Since the problem size is kept fixed, this states the notion of *strong scaling* in the notation of high performance computing. The parallel scaling performance decreases noticeably beyond 256 processors and therefore defines the maximum processor number for each ensemble member used for production runs.

3.3.2 Performance analysis of a single WRF run

The computational performance of the WRF model on JUQUEEN has been monitored and evaluated with the Scalasca Tools version 1.3.0 (Geimer *et al.*, 2010) with trace analysis support by Score-P (Knüpfer *et al.*, 2012). The following performance metrics are computed with the improved compiler optimization described in the previous section for a single WRF run without model perturbation, referred to as a control run. The forecast time is again three hours for the fixed problem size. A typical number of 256 processors is allocated and 32 ranks have been put on a node,

Table 3.1: CPU times for the WRF model on JUQUEEN with initial and improved compiler instructions and varying processor number. The forecast time is three hours for a fixed problem size.

# processors	initial compiler instructions	improved compiler instructions
512	119 s	66 s
256	192 s	108 s
128	322 s	205 s
64	593 s	363 s
32	1125 s	686 s

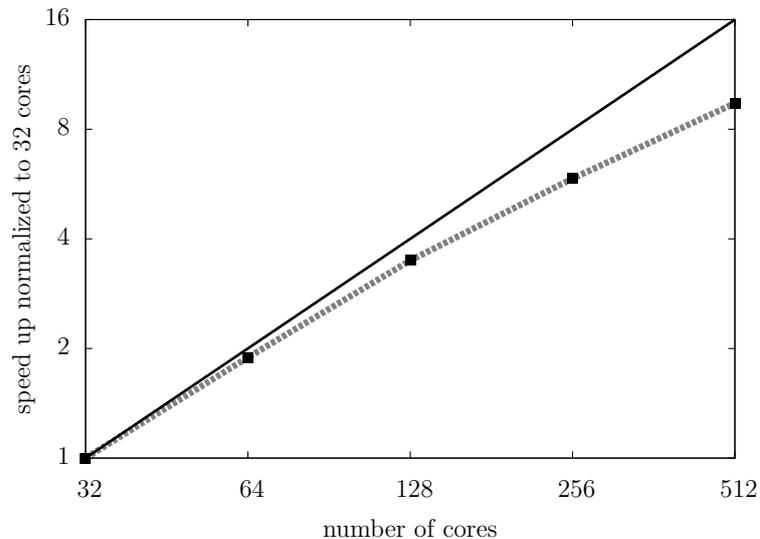


Figure 3.2: Parallel scaling behavior of the WRF model on JUQUEEN for a fixed problem size (dashed line). The speed up has been normalized to 32 processors and a perfect speed up is shown for comparison (solid line). Corresponding CPU times are given in Table 3.1, third column.

i.e. 32 processes are realized on each node. Further SMT is not limited by memory (~ 100 MB per processor), but does not necessarily lead to performance improvements due to additional expenses during parallel I/O instructions. The application is in general compute-bound with 70% and MPI tasks require up to 20%. Point-to-point MPI calls prevail over collective calls since a local discretization scheme is utilized. The average MPI message size is 16 kB and load imbalance is acceptable with up to 5% of the total CPU time. At least in the case of a WRF run without a model perturbation scheme, a computational bottleneck can not be monitored.

Numerous I/O strategies suitable for different problem sizes are provided by the WRF model. Model output may be written individually by each processor into separate files. As favorable this strategy may appear on other systems, it is exceptionally slow on JUQUEEN due to the inability of the GPFS file system to process a large

number of I/O instructions congruently. WRF also offers the possibility of quilting, that is the privation of exclusive I/O processes. However, such a partitioning appears to be disadvantageous, since the domain is only moderately large and the application CPU-bound, hence, computational resources are rather invested in compute tasks. To conclude, performing parallel I/O in a classical manner with PnetCDF support turns out to be most suitable.

The I/O behavior on JUQUEEN has been monitored and evaluated with the characterization tool Darshan version 2.3.1 (*Carns et al.*, September 2009). Aggregated I/O costs are about 10% of the total CPU time. The bulk of this time is due to collective calls of parallel NetCDF operations (i.e. MPI-IO) to build larger blocks. The costs of POSIX I/O (standard output and metadata) are negligible compared to MPI I/O (input of initial and boundary values as well as output file). An average 43 MB/s of I/O throughput can be estimated for the fixed problem size.

3.3.3 Computational improvement of SPPT and SKEB

The previous section addressed a performance analysis of the WRF model without model perturbation. Here, it is turned to the the case in which model uncertainty is either represented by the SPPT or SKEB scheme (see Section 2.3 for a detailed description of both schemes). If either one of the schemes is utilized, CPU times increase drastically and a distinct bottleneck can be identified, which is associated with the stochastic pattern generator used by both schemes. This matter is discussed in the following with different approaches of improvement.

In the SPPT and SKEB scheme, random numbers are generated in two-dimensional spectral space on the rectangular grid to either formulate a parameterization tendency perturbation or a tendency forcing, respectively. The energy associated with such forcings shall be transformed to physical space as a whole, and therefore, it has to be ensured that the Fourier back transformation is real-valued. This constraint is fulfilled as long as the symmetry condition $\Re(\vartheta(k, l)) = \Re(\vartheta(-k, -l))$ and $\Im(\vartheta(k, l)) = \Im(-\vartheta(-k, -l))$ holds, where $\vartheta \in \{r, \psi', \theta'\}$ denotes the spectral coefficients of both schemes according to (2.11), (2.17) and (2.18). This implies that the real and imaginary parts of $\vartheta_{k,l}$ have to be an even or odd function of wavenumbers k and l , respectively. The spectral coefficients evolve according to a first-order autoregressive process (see (2.12), (2.19) and (2.20)), and hence, the symmetry condition translates directly to the generation of a complex Gaussian white noise $\epsilon_{k,l}$ on the two-dimensional grid, depicted in Figure 3.3. In the officially released WRF code, the symmetry condition is simply realized by using global indexes for the calculation of $\epsilon_{k,l}$ and the necessary symmetry conditions can be implemented in a straight forward manner since random numbers exist over the whole domain and are identical on every processor. This implies for each process

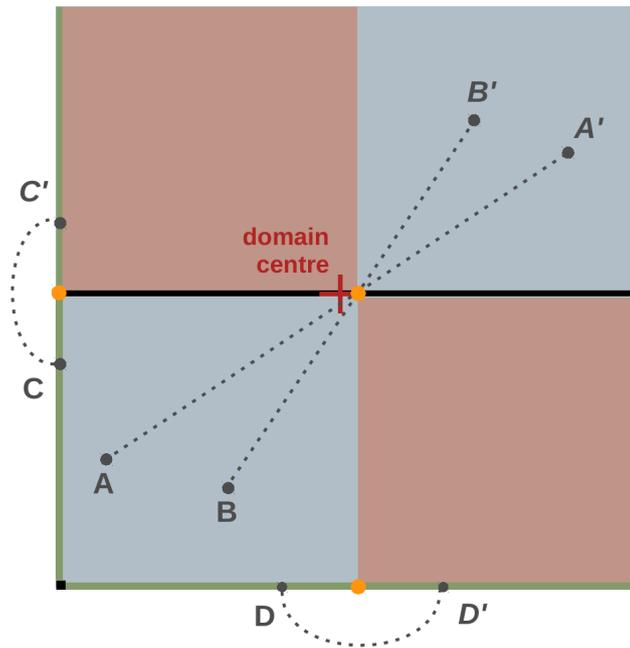


Figure 3.3: Necessary symmetry conditions of complex Gaussian white noise $\epsilon_{k,l}$ in spectral space to be ensured for the stochastic pattern generation in the SPPT and SKEB scheme. Orange points mark symmetry points associated with areas of same color, but with a change of sign. Black colored areas are left disregarded. Note that the inner symmetry point does not coincide with the domain centre, but is shifted by one grid point to the right.

the utilization of a pseudo random number generator¹(PRNG) for every grid point of the entire domain at every time step, to use a Box Muller transform to compute $\epsilon_{k,l}$. Mentionable, the procedure has to be done once in case of the SPPT scheme and twice in case of the SKEB scheme. Adopting the nomenclature of *Romine et al.* (2014), this initial implementation is denoted simply as **STOCH** code hereafter.

The **STOCH** code appears to be exceptionally slow on JUQUEEN, since the computational expenses per processor are high, while the processors themselves possess a relatively slow clock speed. The second column in Table 3.2 lists CPU times of the WRF model with SKEB for varying number of processors. The simulation setup is chosen identical as in Section 3.3.2. Starting with 32 processors, CPU times of the WRF model approximately double compared to an unperturbed run (see Table 3.1 for comparison). However, in case of 512 processors, CPU times increase up to a factor of 14, as computational expenses for each processor are independent of the processor number. Two competitive approaches have been investigated to solve this issue, with associated CPU times listed in the remaining

¹The XL compilers use per default a classic multiplicative congruential method (*Park and Miller*, 1988) for pseudo random number generation.

Table 3.2: CPU times of the WRF model with the SKEB scheme and varying processor number for different implementations of the `STOCH` code. `STOCH` denotes the initial implementation. The simulation setup is chosen identical as in Table 3.1.

#processors	<code>STOCH</code>	<code>STOCH_PARALLEL</code> (collective MPI)	<code>STOCH_PARALLEL</code> (point-to-point MPI)	<code>STOCH_SEED</code>
512	925 s	96 s	87 s	106 s
256	973 s	159 s	145 s	164 s
128	1094 s	273 s	251 s	283 s
64	1302 s	505 s	470 s	500 s
32	1780 s	968 s	921 s	962 s

columns of Table 3.2. The different approaches will be introduced in the following.

The most straightforward approach to improve the computational performance of the `STOCH` code is to parallelize the computation of $\epsilon_{k,l}$ over the model grid. Each process calculates unique random numbers exclusively over the decomposed subdomain, with subsequent MPI communication thereafter. This implementation is denoted as `STOCH_PARALLEL` code. To ensure different random numbers on different processes, a unique tag defines itself by each process' rank within the `MPI_COMM_WORLD` and is added to the initial seed of the PRNG. The necessary MPI communication has been implemented via collective MPI, i.e. random numbers from each subdomain are broadcasted to all processes, as well as point-to-point MPI, i.e. communication exclusively at matching symmetry points. The implementation with point-to-point MPI communication appears nontrivial since the centre of symmetry does not coincide with the domain centre (see Figure 3.3), and a partitioning of the domain among the processes is not necessarily even. This instance excludes block-wise code implementation to ensure universality. The correctness of this implementation is simply verified by collective MPI communication. Note that simulation results with this implementation depend on the process number and therefore differ from results obtained with the `STOCH` code.

In a second approach, random numbers are again exclusively computed over the processes' decomposed subdomain. Yet, it is ensured that identical random numbers exist at symmetry points. In practice, a unique seed for the PRNG has to be defined for each grid point and time step of the decomposed subdomain, which is identical at symmetry points but different elsewhere. This tag is added to the initial seed of the PRNG during model initialization and modified during model integration by an additional time step dependent scalar. As PRNGs tend to exhibit artifacts in the statistical distribution of random numbers, an additional large prime number ensures a more uniform spread of random numbers over the plane. The necessity of MPI communication is therefore excluded and it is referred to the `STOCH_SEED`

code hereafter. Note that simulation results obtained with this implementation are independent of the process number, but again differ from the results obtained with the `STOCH_SEED` code.

Returning to the efficiency results given in Table 3.2, all modified versions of the `STOCH` code reduce CPU times and increase the parallel scalability to an extent comparable with an unperturbed WRF model run. Additional computational expenses of approximately 20% have to be invested for the `SKEB` scheme, whereas this value reduces to about 10% for the `SPPT` scheme (not shown here). Surprisingly, both implementations of the `STOCH_PARALLEL` code perform slightly superior to the `STOCH_SEED` code, although no MPI communication is involved in the latter. Hence, initialization of a pseudo-random number sequence with the FORTRAN procedure `RANDOM_SEED` appears to be slower on `JUQUEEN` than an associated MPI communication. The `STOCH_PARALLEL` code with point-to-point communication performs best for all given processor numbers. Collective communication appears to be disadvantageous, since a latency of 2 μ s ensures little computational expenses as the number of pending MPI requests increases with decreasing processor number.

It remains to be shown, that for $N \rightarrow \infty$ and vanishing sampling error, with N the number of ensemble members, the ensemble's statistical properties obtained with either the `STOCH_PARALLEL` or `STOCH_SEED` code are identical with those obtained with the initial `STOCH` code. For this purpose, several test runs under different atmospheric conditions have been carried out to investigate the spatial and temporal evolution of the ensemble mean $\overline{\mathbf{x}}^f = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i^f$, and the ensemble spread, defined as the standard deviation of ensemble members around the ensemble mean:

$$\text{ES}(\mathbf{X}^f) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i^f - \overline{\mathbf{x}}^f)^2}, \quad (3.1)$$

where $\mathbf{X}^f = \{\mathbf{x}_1^f, \dots, \mathbf{x}_N^f\}$, with the nomenclature of Chapter 2. It is suggested, that a point-wise root mean square comparison over the domain is a robust norm to compare simulations results obtained with the initial `STOCH` code and modified versions, which reads for $\overline{\mathbf{x}}^f$ and in case of a two-dimensional variable:

$$\text{RMSE}(\overline{\mathbf{x}}^f - \overline{\mathbf{x}}_{\text{STOCH}}^f) = \sqrt{\frac{1}{XY} \sum_x \sum_y \left(\overline{x}^{f(x,y)} - \overline{x}_{\text{STOCH}}^{f(x,y)} \right)^2}, \quad (3.2)$$

with X, Y the number of horizontal grid points in each direction and $\overline{\mathbf{x}}_{\text{STOCH}}^f$ the ensemble mean obtained with the `STOCH` code. $\text{RMSE}(\text{ES}(\mathbf{x}^f) - \text{ES}(\mathbf{x}_{\text{STOCH}}^f))$ is defined accordingly.

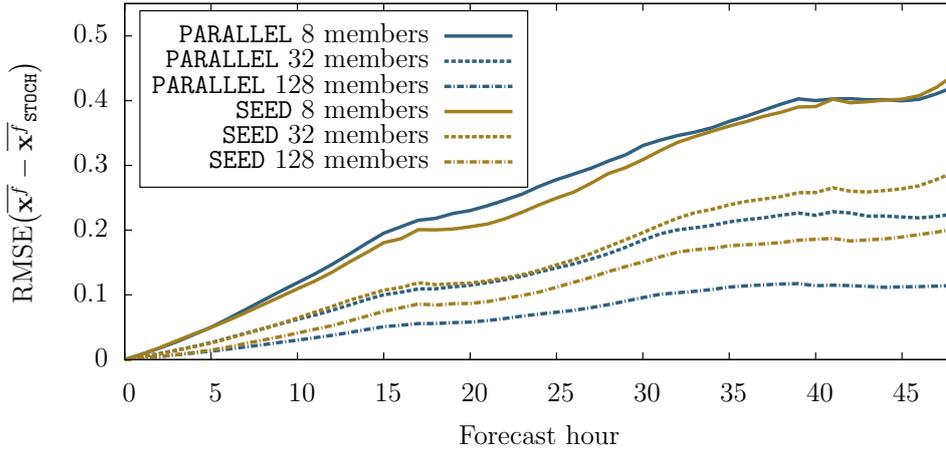


Figure 3.4: RMSE of increments in ensemble mean of 100 m wind speed obtained with the initial `STOCH` code and the modified versions `STOCHPARALLEL` (blue lines) as well as `STOCHSEED` (orange lines). The SKEB ensemble shares identical initial conditions. Solid lines denote 8 members, dashed lines 32 and dashed-dotted 128. The forecast is initialized at 00 UTC 23 June 2014.

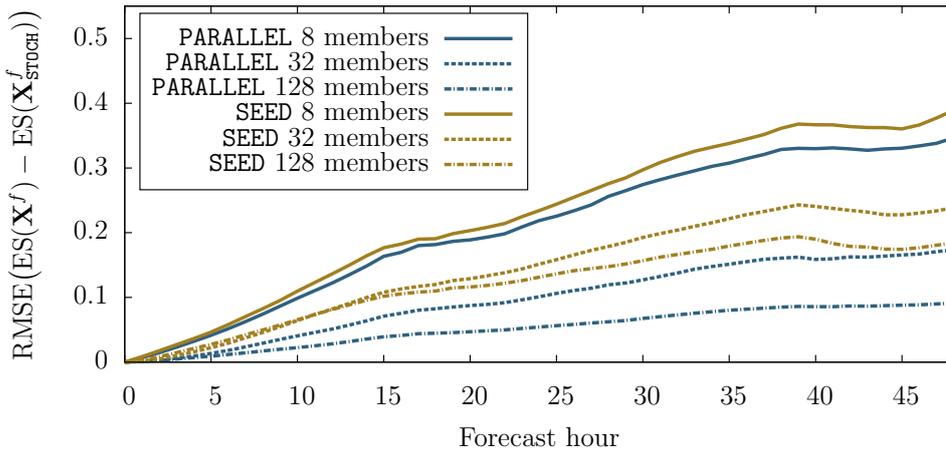


Figure 3.5: Same as Figure 3.4, but for ensemble spread.

Here, the discussion is restricted to a single model run, initialized at 00 UTC 23 June 2014, and a forecast time of 48 hours. Results obtained for other periods show analogous results. Ensemble members share identical initial conditions and boundary values to isolate the effect of different `STOCH` code realizations on model results. The ensemble members have been generated solely by SKEB, while the limit $N \rightarrow \infty$ is approximated by $N = 128$.

Figure 3.4 and Figure 3.5 show the temporal evolution of $\text{RMSE}(\overline{\mathbf{x}}^f - \overline{\mathbf{x}}^f_{\text{STOCH}})$ and $\text{RMSE}(\text{ES}(\mathbf{X}^f) - \text{ES}(\mathbf{X}^f_{\text{STOCH}}))$, respectively, for 100 m wind speed and varying ensemble size. $\text{RMSE}(\overline{\mathbf{x}}^f - \overline{\mathbf{x}}^f_{\text{STOCH}})$ and $\text{RMSE}(\text{ES}(\mathbf{X}^f) - \text{ES}(\mathbf{X}^f_{\text{STOCH}}))$ decreases continuously with increasing

ensemble size for both, the `STOCH_PARALLEL` and `STOCH_SEED` code. However, the decrease in case of the `STOCH_SEED` code shows a slower convergence rate. Since a seed is defined manually each time, artifacts arise which degrade the randomness and translate to structural deviations in ensemble spread. A similar behavior is observed for point-wise comparison of the ensemble mean (not shown here). To conclude, solely for the `STOCH_PARALLEL` code it can be assumed, that $\lim_{N \rightarrow \infty} \text{RMSE}(\text{ES-ES}_{\text{STOCH}}) = 0$ holds. Thereupon, the `STOCH_PARALLEL` code has been implemented into ESIAS-met.

3.4 ESIAS-met

Sequential Importance Resampling links data assimilation to continuous updates of the model ensemble. Resampling involves rejection of ensemble members with low importance weight and duplication of ensemble members with high importance weight to restore the ensemble size and represent the posterior density. Thus, in one way or the other, resampling requires communication among the ensemble members. However, present approaches to execute ensemble forecasts do not provide this requirement explicitly. Each realization of the atmospheric model runs autonomously, i.e. each ensemble member is associated with a single executable, supervised by a generic script. In the notation of high performance computing, this states the notion of farming. Such a procedure is in general convenient as atmospheric models are complex and undergo continuous updates.

In the context of Sequential Importance Resampling with very large ensemble sizes, the drawback of these approaches is twofold. First, the absence of explicit communication between the ensemble members prohibits the monitoring of an effective sample size N_{eff} and resampling intervals have to be defined a priori. Second, and more importantly, the ensemble model's states and boundary values have to be duplicated during the resampling procedure. One can think of various ways to realize this, but as long as the ensemble members run autonomously, considerable execution time will result as the ensemble size increases. This is an undesirable circumstance, as certain users demand NWP products as early as possible to anticipate economic or environmental strategies accordingly.

ESIAS-met is a software system which addresses the above. In the following, the basic idea to execute the WRF model in parallel is introduced in detail. The system's strong parallel scalability is demonstrated and its computational efficiency for Sequential Importance Resampling.

3.4.1 A parallel ensemble version of the WRF model

The WRF modeling system has been modified towards a stand-alone ensemble control system. The basic concept is to gather all ensemble members within a single

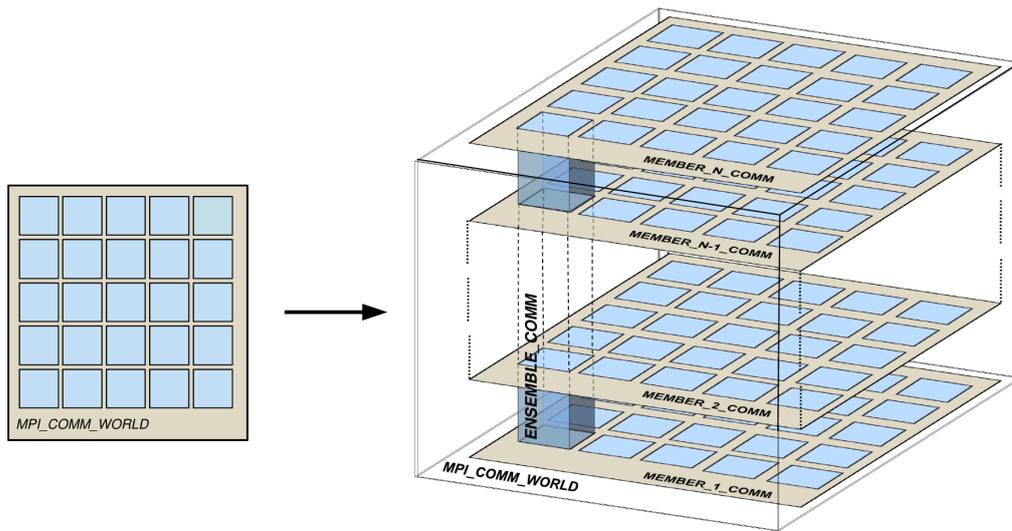


Figure 3.6: Visualization of the MPI concept to modify the WRF code to a stand-alone ensemble control system (as part of ESIAS-met). Left: Initial point where the `MPI_COMM_WORLD` communicator includes all processes distributed among a two-dimensional decomposed domain. Right: Novel concept in which N ensemble member are executed in parallel, each within its `MEMBER_COMM` communicator. The `ENSEMBLE_COMM` communicators gather the ensemble member's partitions of the decomposed model domain.

executable and establish the necessary inter-member communication with the MPI library. This concept is illustrated in Figure 3.6 and will be explained in detail in the following.

In terms of distributed memory parallelism, all allocated processes are grouped together within the global MPI communicator `MPI_COMM_WORLD`. If one associates each of these processes with a single partition of the decomposed model domain, the `MPI_COMM_WORLD` communicator may then be associated with the whole model domain. However, it should instead also be possible to associate the `MPI_COMM_WORLD` communicator with the whole model ensemble. For this purpose, the `MPI_COMM_WORLD` communicator is split into N new communicators denoted as `MEMBER_COMM`, with N the number of ensemble members. Each of the `MEMBER_COMM` communicators groups the processes allocated for a single ensemble member and thus, replaces the initial role of the `MPI_COMM_WORLD` communicator. For the sake of clarification, if the model domain decomposition represents a first stage of parallelism, then a second stage is added which realizes the parallel execution of the ensemble members.

In practice, the MPI library offers an intrinsic function to split any communicator. A color has to be defined for every process within this communicator, which groups all processes with the same color into the new communicator. Here, the overall allo-

cated processes np group equally distributed into the N `MEMBER_COMM` communicators to give

$$np_{member} = np/N,$$

with np_{member} the number of processes per ensemble member. The color for this grouping can be defined in ascending order as

$$num = rank/np_{member},$$

where $rank$ denotes the process' rank in the `MPI_COMM_WORLD` and num equals the largest integer that does not exceed the range of num .

Once a second stage of parallelism is introduced, the required inter-member communication is realized in an analogous manner. The `MPI_COMM_WORLD` communicator is again split into np_{member} `ENSEMBLE_COMM` communicators. A single `ENSEMBLE_COMM` communicator groups N processes, each associated with a different ensemble member, but with the same partition of the decomposed model domain. The color for this grouping is the processes' corresponding rank within the `MEMBER_COMM` communicator. In summary, WRF is executed N times in parallel within N `MEMBER_COMM` communicators, and the `ENSEMBLE_COMM` communicator gather the ensemble member's partitions of the decomposed model domain. Now, initial and boundary values as well as the state vector can be exchanged among the ensemble members in a convenient way.

Integrating the novel MPI concept into the ARW solver can be realized in a straight forward manner. The `MPI_COMM_WORLD` communicator is split immediately after the MPI initialization and consecutively replaced by the `MEMBER_COMM` communicator throughout the code. However, as convenient this implementation appears, there are further issues to be considered. Blocking among processes during model I/O has to be avoided, which includes model input and output with PnetCDF as well as POSIX I/O of metadata and standard output/error. Since the ensemble typically possess a larger ensemble size than the global ensemble to serve as input, the additional ensemble members are excluded from the input procedure and initial and boundary values are distributed among the ensemble thereafter. Blockings during model output is simply avoided by providing a single file to every ensemble member. In case of numerous metadata files, which are also subject to perturbation, each file is duplicated N times. Standard output/error is handled task local in the WRF model, i.e. a single file is created each per processes. Utilizing thousands of processes, this strategy appears to be exceptionally slow on JUQUEEN. This pitfall is avoided by redirecting standard output/error into a single file per ensemble member, which is created prior to the model run with append access.

The novel MPI concept has also been applied to the *real* program and the entire

WPS system in the same manner as described above, such that the whole ensemble modeling system may be executed in parallel. There are no restrictions in model execution with ESIAS-met compared to an autonomous WRF model run and the following list highlights key features of the system:

- parallel execution of WRF ensemble members within a single executable
- inter-member communication realized for efficient ensemble initialization and Sequential Importance Resampling
- convenient ensemble execution on batch systems to avoid farming
- no restriction in nesting procedure
- multi-physics ensemble enabled, i.e. different parameterization among ensemble members
- adaptive time-stepping possible, also in realm of particle filter
- change in perturbation strategy after a resampling step (SKEB, SPPT, SPP, including resampling of stochastic field and random number seeds)

ESIAS-met converts the WRF model to a stand-alone ensemble control system. As the WRF model undergoes continuous updates, at least biannually, it has to be ensured that the ESIAS-met can be updated accordingly with reasonably low effort. This can be seen as fulfilled, as such an update has been carried out once in the frame of this work, with the investment of approximately two days of work. According to Powers *et al.* (2017), there are no changes to the WRF Software Framework proposed explicitly for the future, which might prohibit the update of ESIAS-MET to recent versions of the WRF model.

3.4.2 Proof of concept: parallel scalability of ESIAS-met

ESIAS-met has been designed with the scope of a completely parallel and computational efficient execution of an ultra large WRF ensemble. It remains to prove (i) the principle parallel scalability of the system and that (ii) such a parallel scalability is independent of the ensemble size. In the context of Sequential Importance Resampling, it remains to prove that also (iii) the computational efficiency of a resampling step is independent of the ensemble size. The demonstration of (i) - (iii) is done each in terms of a feasibility study and a conclusion about the universality is drawn at the end of this section.

Each feasibility study is defined as a *real case study* and the forecast date is randomly initialized at 00 UTC 24 November 2014. Each ensemble member is executed on a moderate domain size of $240 \times 240 \times 30$ grid points. A constant number of 64

Table 3.3: CPU times of ESIAS-met with increasing ensemble size for an identical ensemble without particle filtering. The number of processors per ensemble member is kept fixed.

# members	# processors	CPU time [s]	Loss in CPU time [%]
1	64	1161	–
2	128	1165	0.3
4	256	1166	0.4
8	512	1167	0.5
16	1,024	1165	0.3
32	2,048	1171	0.9
64	4,096	1170	0.8
128	8,192	1185	2.0
256	16,384	1188	2.3
512	32,768	1184	1.9
1024	65,536	1189	2.4
2048	131,072	1199	3.2
4096	262,144	1200	3.3

processors is allocated per ensemble member. CPU times are measured while successively increasing the ensemble size to 4096, utilizing 262,144 processors.

To demonstrate (i) the parallel scalability of the system and (ii) its independence of ensemble size, the forecast time has to be sufficiently large and is chosen as 24 hours. Ensemble members share identical initial and boundary values to exclude variations in CPU time due to a diverse evolution of the atmospheric state. Measured CPU times of this feasibility study are listed in Table 3.3. As it has been claimed above, without the expanse of communication between the members, the CPU time remains approximately constant with increasing ensemble size. Deviations from a perfect scalability are due to a limited bandwidth during I/O processes, albeit of the order of 1–3% and therefore negligible. Hence, the performance analysis for a single WRF model run (Section 3.3) also holds for the entire ensemble.

It is proceeded in the same manner to demonstrate (iii), but now a resampling step of the basic particle filter is performed one forecast hour after initialization at 01 UTC 24 November 2014. Each ensemble member utilizes again 64 processors on the given test domain and the ensemble shares 16 different sets of initial conditions and boundary conditions, which are sent equally distributed to all remaining ensemble members during the initialization process. The ensemble control system developed within this work states a full modeling system and thus is subject to the assimilation of data from various sources. Data assimilation is however yet restricted to conventional in situ observations, the assimilation of data derived from lidar, profiler, radar and satellites is not within the scope of this work and is left to future efforts. Reports

from the *NCEP ADP Global Upper Air and Surface Weather Observations* dataset and various measurement towers (see Section 4.2 for a detailed description) are collected according to a time stamp of 2014-11-24.01:00:00 \pm 00:15:00, which yields a sum of 15,788 observations, and assimilated thereafter. In practice, resampling within ESIAS-met involves the following steps:

R.1 Calculation of ensemble member weights

- (i) The root process reads observational data, sorts out reports depending on the time stamp, location as well as observation type, and the selected observations are broadcasted to all remaining processes
- (ii) Each ensemble member assigns each report to a corresponding subdomain defined by the nearest grid point (within `MEMBER_COMM`) and model values are interpolated for comparison
- (iii) On each subdomain, the sum of all model-observation increments is calculated and the results are in turn summed up for each ensemble member (within `MEMBER_COMM`) to yield the individual ensemble member weight
- (iv) Ensemble member weights are normalized (within `ENSEMBLE_COMM`)

R.2 If the effective sampling size N_{eff} (Equation (2.36)) drops below a threshold: resampling of the ensemble by the root process (a combination of residual and stratified resampling, Equation (2.34) and (2.35)), results are broadcasted to all remaining processes

R.3 Duplication of resampled members until the initial ensemble size is restored, i.e. resampled members send the model state, boundary values and meta data to rejected members (within `ENSEMBLE_COMM`)

In the frame of the feasibility study, it is ensured that 10% of the ensemble members are resampled with identical weights. Let n_{max} denote the maximum duplication count of a resampling step:

$$n_{\text{max}} = \max_i(n_i), \quad \text{with } i \in \{1, \dots, N\}, \quad (3.3)$$

with $n_i = n'_i + n''_i$ the duplication count of member i , then $n_{\text{max}} \leq 10$ holds $\forall N$. By imposing this constraint, the number of duplications per resampled member is independent of ensemble size. However, step R.2 is explicitly executed for the sake of performance measurement. Table 3.4 summarizes the number of MPI messages, transferred data and CPU times associated with a resampling procedure for a single member, separately for model state, boundary values and metadata. Here, a MPI message is either associated with a character, a scalar, or a one-, two- or three-dimensional array. The number of MPI messages is independent of processor number

Table 3.4: Number of MPI messages, transferred Mbytes and CPU times associated with the resampling procedure for a single member.

type of transfer	# MPI messages per process	# transferred Mbytes	CPU time [s]
model state	525	1541	3.1×10^{-1}
boundary conditions	80	187	4.59×10^{-2}
metadata	2	0.13	6.35×10^{-5}

or domain size, whereas the amount of transferred bytes depends in principle on domain size and increases with the processor number due to an increasing amount of ghost points. CPU time is not necessarily limited by the maximum bandwidth, which is 10 Gbytes/s per ensemble member utilizing 64 processors each. Rather, a large number of MPI messages has to be processed with each performing an initial handshake between the communicators.

Associated CPU times for the resampling procedure with increasing ensemble size are shown separately for steps R.1 – R.3 in Figure 3.7. The ensemble size is again doubled successively from 32 to 4096 members. CPU times for transfer of metadata are negligible and excluded. Selection of suitable reports by the root process (step R.1) takes a constant CPU time of 17 seconds. As this procedure depends on the amount of observations assimilated, it is not further discussed. Starting with 32 members, the measured CPU times for the data transfer (R.2) are in good agreement with a linear extrapolation of CPU times for a single member (Table 3.4), i.e. CPU times given in Table 3.4 are multiplied by n_{\max} . As the ensemble size is doubled consecutively to up to 4096 members, the amount of transferred data doubles, which is shown separately for the model state and boundary values. However, there is only a negligible decrease in computational performance, as the increasing number of MPI messages have to be transported across various nodeboards. This effect is clearly noticeable for 256 members, which marks the point where an additional nodeboard is allocated for the first time. In the case of an extreme setup with 4096 members, 8.6 Tbytes are send and received across 16 racks (262,144 cores) within 11 seconds during a resampling step.

To conclude, it has been shown for a fixed problem size, that the CPU times of a resampling step are independent of the ensemble size, and exclusively depend on n_{\max} . This results in minimal computational expenses that are in principle negligible compared to the model integration time. A linear scaling behavior of the CPU time with n_{\max} does not raise concern, since ESIAS-met offers the possibility to monitor the effective sample size N_{eff} , which prevents a large n_{\max} in general.

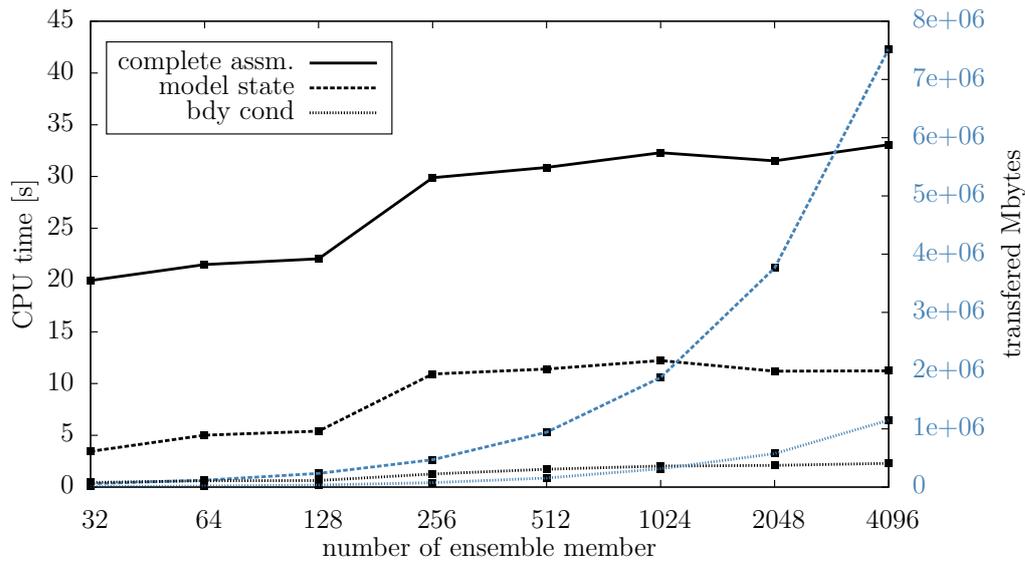


Figure 3.7: Computational analysis for a complete particle filter resampling step with ESIAS-met for varying ensemble sizes. Dashed and dotted lines denote CPU times and transferred amount of data for duplication of the model state and boundary values, respectively. The solid line denotes the overall CPU time, which includes the processing of observations and calculation of particle weights.

3.4.3 Flowchart of ESIAS-met

A summary of the ESIAS-met system is depicted in Figure 3.8, and may be compared to the initial point of the WRF modeling system depicted in Figure 3.1. M different meteorological input files from a global ensemble system and M terrestrial files are input to the ESIAS-met version of WPS. M intermediate output files are computed, which contain horizontally interpolated input data on the model's grid. The ensemble version of *real* computes in parallel the vertically interpolated fields to yield initial and boundary values. Within the ESIAS-met version of the ARW solver, input files are processed by M ensemble members and distributed equally among the remaining $N-M$ ensemble members. Stochastic perturbations of surface parameter define N different sets of metadata. M ensemble members are dynamically downscaled and the remaining $N-M$ ensemble members are generated by additional model uncertainty representation by the SKEB, SPPT or SPP scheme. The SPP scheme is a model uncertainty scheme which has been developed to perturb parameter within the parameterizations. Since results with this scheme are not shown in this work, it is not further discussed. The system enables an efficient execution of a Sequential Importance Resampling Filter or Smoother within the ESIAS-met software environment for continuous updates of the model's ensemble.

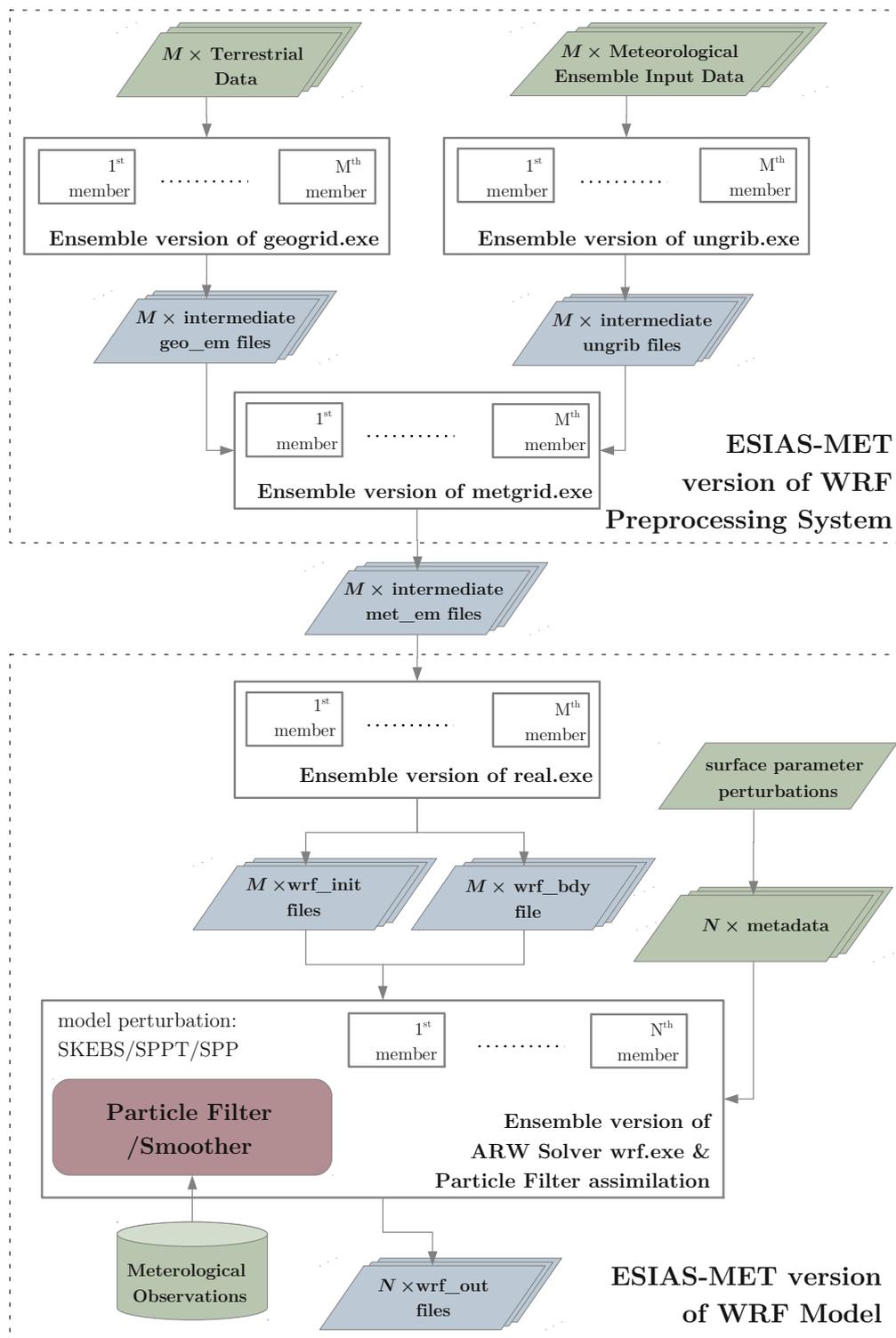


Figure 3.8: Flowchart of the ESIAS-met modeling system. Green parallelograms indicate external input data, blue parallelograms model output as well as intermediate files. Staggered files indicate the existence of multiple files.

Chapter 4

Wind power forecasting with the WRF model

The deterministic forecast skill of turbine hub height winds derived by the Weather Research and Forecasting (WRF) model is investigated. At the beginning, the WRF model setup, meteorological observations and the utilized wind power model are introduced. In a first assessment, the WRF model performance for eight different parameterization suites is evaluated against conventional in situ observations including winds at turbine hub heights. The day-ahead forecast of wind power over Germany is derived and is compared with the real power feed-in, which is estimated by the transmission system operators (TSO). Initial results show a large positive wind bias for all model configurations. Thereupon, the best WRF configuration is further optimized to match a deterministic forecast skill comparable to that of the TSO day-ahead forecast. This configuration serves as the deterministic WRF model setup for the ultra large ensemble.

4.1 WRF model setup

A detailed description of the WRF model is given in Section 3.1. This section serves the purpose to define a suitable domain configuration for to the ultra large ensemble system and determine appropriate initial conditions from a global ensemble system.

4.1.1 Domain configuration

An ensemble size of the order of $O(1000)$ member certainly puts an upper limit on the computational expenses for each ensemble member, and consequently on the model domain configuration. As it is typically the case for regional NWP modelling, a suitable domain configuration states a compromise between a sufficiently large domain size and model resolution. This is especially true for ensemble integrations of

regional NWP models, as domain size and model resolution have additional effects on the ensemble dispersion.

The benefit of an increasing domain size is multifold. A larger domain size generally reduces errors introduced by the forcing of the lateral boundary conditions (*Lee et al.*, 2008; *Lowrey and Yang*, 2008). Spurious gradients due to temporal and spatial interpolation may generate gravity-inertia waves, which propagate with different speeds to the meteorological area of interest (*Warner et al.*, 1997). This is especially true with increasing latitude, since the atmospheric condition tends to be more baroclinic, resulting in a strong flow across the boundaries. *Nutter et al.* (2004) discusses this limitation in the frame of ensemble forecasting and shows how infrequently and coarsely resolved lateral boundary conditions remove small scale features, which acts to constrain ensemble dispersion on a small domain. In general, such limitations may be reduced by limiting the grid ratio between the coarser-resolution model providing the lateral boundaries and the outer domain of the regional NWP model. This requirement leads necessarily to a lateral-buffer zone of moderate horizontal resolution and integrated nested grids which realize high-resolution forecasts. Further, an ultra large ensemble size may only be beneficial if the domain size is large enough such that ensemble perturbations evolve on the synoptic scale. Otherwise, at least for explosive cyclogenesis, ensemble dispersion will mainly be constrained to the global ensemble system. At the same time, however, a sufficiently high model resolution has to be provided, as it is well known to increase the general forecast skill of NWP models. This necessarily holds for winds at turbine hub heights (*Olsen et al.*, 2016; *Siuta et al.*, 2017). Additionally, in the context of ensemble forecasting, an increased horizontal resolution leads to stronger nonlinear growth of ensemble perturbations.

Based on the considerations discussed above, the domain configuration of the WRF model is summarized in Table 4.1 and the nesting procedure is depicted in Figure 4.1. The outer domain covers most of Europe with a horizontal resolution of 12 km, which corresponds to horizontal resolutions of operational deterministic global forecast systems, e.g. the Global Forecast System (GFS) from NCEP, 13 km, (*Caplan et al.*, 1997); the Integrated Forecast System (IFS) from ECMWF, 9 km, (*Ritchie et al.*, 1995); the Global Deterministic Prediction System (GDPS) from MSC, 15 km, (*Côté et al.*, 1998), and is notably higher than their corresponding ensemble system (the Global Ensemble Forecasting System (GEFS), 33 km, (*Toth and Kalnay*, 1997); the Ensemble Prediction System (EPS), 18 km (*Molteni et al.*, 1996); the Canadian Global Ensemble Prediction System (GEPS), 50 km, (*Li et al.*, 2008)). The inner nest covers most of Central Europe with a horizontal resolution of 4 km and is the target domain for any power predictions aggregated over Germany. The domain extension and resolution is comparable to that of other convective-scale ensembles which run operationally at weather centres, e.g. the COSMO-DE-EPS from the German Weather

Table 4.1: Configuration of the WRF model domain. The outer domain is denoted as d01 and the inner domain as d02.

Model version	3.7.1
Map projection	Lambert Conformal
Land use data	MODIS
Central point	54° N, 12.5° W
Model top	50 hPa
Horizontal resolution d01	12 km
Horizontal resolution d02	4 km
Number of horizontal gridpoints d01	330 × 330
Number of horizontal gridpoints d02	301 × 361
Vertical layers	50
Maximum time step d01	72 s
Maximum time step d02	24 s

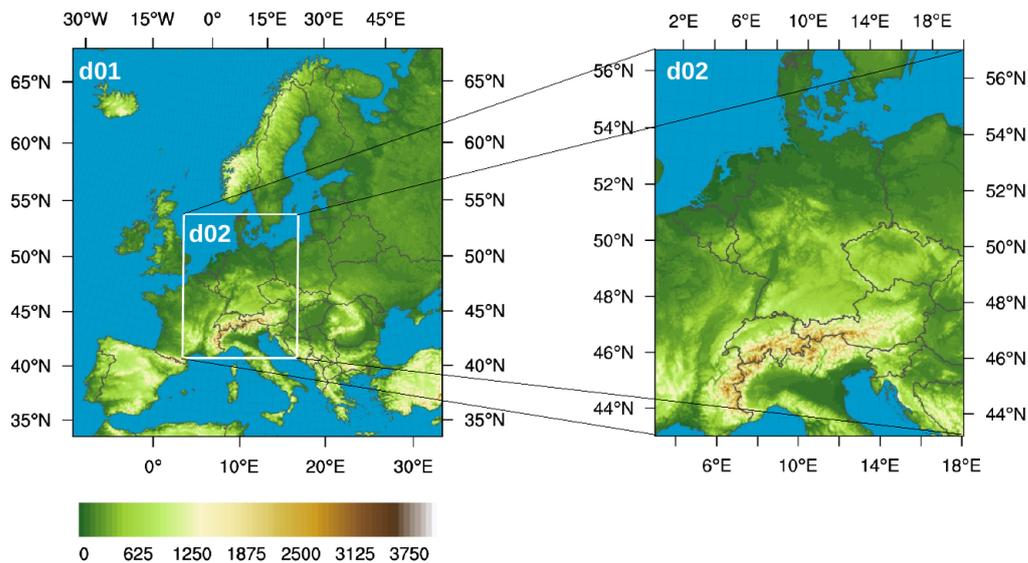


Figure 4.1: WRF domain configuration with its nesting procedure on a terrain height map.

Service 2.8 km, (*Hagelin et al.*, 2017); the AROME-France-EPS from Meteò France 2.5 km, (*Bouittier et al.*, 2012); the MOGREPS-UK from Met Office 2.2 km, (*McCabe et al.*, 2016). A horizontal resolution between 10 to 5 km is intentionally avoided due to poor performance of cumulus parameterization within this range (commonly referred to as the "grey zone"). The vertical grid spacing is reduced in the boundary layer, with 2 layers within 100 m above the surface. The WRF model offers USGS and MODIS based land use classifications and the latter have been proven to be more suitable for the German area. The numerical time-stepping is adaptive, with the Courant-Friedrichs-Lewy condition as a dynamical constraint.

4.1.2 Initial conditions and lateral boundary values

The ultra large WRF ensemble is initialized by downscaling a coarser resolution global ensemble system. Boundary values are provided in the same manner. This approach has been a common procedure for ensembles of regional NWP models, in case of experimental studies (*Eckel and Mass, 2005; Berner et al., 2011*) as well as operational systems (*Montani et al., 2011; Tennant, 2015*). However, one has to be aware that, firstly, the atmospheric model state undergoes a "spin-up" time to adopt to the dynamics of the regional NWP model, and secondly, insufficient resolution of the analysis may degrade the forecast skill of the regional NWP model and constrain ensemble spread as small scale features are smoothed (*Stensrud et al., 2000*).

Two global ensemble systems have been tested to serve as input to the WRF model, the EPS ensemble (ECMWF) and the GEFS ensemble (NCEP). The methods to generate the associated analysis perturbations of both systems have been described in Section 2.2. During the time periods of interest, the 50-member EPS ensemble features a horizontal resolution of 32 km on 91 vertical layers and the 20-member GEFS ensemble a horizontal resolution of 50 km on 42 vertical layers. However, in practice, analysis and forecast fields from operational global ensemble prediction systems are provided solely in restricted resolution to external users. In particular, the EPS ensemble results are available 3-hourly at 55 km resolution on 12 pressure levels, and the GEFS ensemble 6-hourly at 110 km resolution on 11 pressure levels. Multiple test runs under different atmospheric conditions have been conducted by dynamically downscaling both ensemble systems with the WRF model, i.e. no model uncertainty representation has been utilized. The resulting ensemble spread appeared insufficient and little relationship to the model error could be found. Especially a low resolution in the vertical heavily smooths analysis perturbations and makes both data sets unsuitable to drive a regional NWP system.

As the resolution of meteorological driving data has proven to be crucial for the WRF model performance, the *ESRL/PSD GEFS Reforecast Version 2* (*Hamill et al., 2013*) is utilized to provide initial conditions and boundary values to the ultra large ensemble. The National Oceanic and Atmospheric Administration (NOAA) produces a reforecast with a fixed version of the 2012 GEFS model (version 10). The ensemble has a size of 11 members, one control member and 10 perturbed members. The reforecast is initialized once a day at 00 UTC, from December 1984 to present. The horizontal resolution is 50 km for the first 8 days on 42 vertical levels. Analysis perturbations are generated with the Ensemble Transform method with rescaling (*Wei et al., 2008*), described in detail in Section 2.2. The 6-hourly cycling of ETR perturbations utilized in the operational GEFS model is preserved, though only 10 members are involved. Forecast fields are given 3-hourly and the data set is available in full model resolution at <http://portal.nersc.gov/project/refcst/v2/>. A detailed description

about the data set can be found in *Hamill et al. (2013)*.

Initial conditions and boundary values of the GEFS reforecast control member are replaced by the final analysis and 3-hourly forecast fields, respectively, from the GFS model, available at <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forecast-system-gfs>. The data set has an identical horizontal resolution of 50 km, and 27 pressure levels are available. A series of test runs have been carried out by dynamically downscaling both control forecasts with the WRF model. At least for 100 m wind speed and forecast times of 48 hours, simulation results did not show any noteworthy difference.

4.2 Meteorological observations

Conventional meteorological in situ observations are taken to evaluate the performance of the WRF model. To target a forecast variable of 100 m wind speed utilized for wind power predictions, meteorological observations are obtained from synoptic measurement towers.

4.2.1 Comparison of model data and observations

Spatial comparison between model results and observations are realized in the horizontal by inverse distance weighting, as the model grid is irregular due to the map projection of the earth's sphere on WRF's rectangular grid. In general, an inverse distance weighting interpolation for the variable C to the point \mathbf{z} based on given values $C(\mathbf{z}_i)$, with $i = 1, \dots, M$, reads

$$C(\mathbf{z}) = \begin{cases} \frac{\sum_i^M a_i(\mathbf{z})C(\mathbf{z}_i)}{\sum_i^M a_i(\mathbf{z})}, & \text{for } d(\mathbf{z}, \mathbf{z}_i) \neq 0 \\ C(\mathbf{z}_i), & \text{for } d(\mathbf{z}, \mathbf{z}_i) = 0, \end{cases} \quad (4.1)$$

where $a_i(\mathbf{z}) = 1/d(\mathbf{z}, \mathbf{z}_i)^p$ defines the weight, d is a metric operator (here the distance between two points) and p the power parameter, with smaller values tending to increase the influence from more distant grid points. Distances between two points are approximated with the haversine formula yielding the great-circle distance on a sphere. The number of points involved in the interpolation is chosen as $M = 4$ and the power parameter as $p = 2$, such that the inverse distance weighting comes closest to a bilinear interpolation (and is equivalent in the case of a regular grid). For comparison of model derived winds with observations, wind components have to be de-staggered in the horizontal from WRF's staggered grid by linear interpolation to mass points. Vertical interpolation of upper-air observations is linear in pressure for wind and linear in the logarithm of pressure for temperature, humidity and geopotential. Vertical interpolation to 100 m winds and appropriate heights of various

measurement towers involves two steps. Firstly, a height information is derived at half σ -levels by de-staggering the geopotential, and secondly, a linear interpolation of winds from neighboring levels is used, which has been shown to be sufficient by *Drechsel et al.* (2012). Besides, it can be assumed that the vertical interpolation to 100 m is nearly free of errors, since the second lowest half σ -level is placed at a height which approximately coincides.

Wind components derived by the WRF model are in a model-relative coordinate system defined by the Lambert conformal map projection, in which each grid point is associated with a local rotation angle α_r between the y -axis and the meridians. The back transformation of the wind vector components \mathbf{u}_{model} to an earth-relative coordinate system with a zonal and meridional decomposition of \mathbf{u}_{earth} is realized according to

$$\mathbf{u}_{earth} = \begin{pmatrix} \cos(\alpha_r) & -\sin(\alpha_r) \\ \cos(\alpha_r) & +\sin(\alpha_r) \end{pmatrix} \mathbf{u}_{model}. \quad (4.2)$$

Finally, the wind vector \mathbf{u}_{earth} has to be rotated from the earth-relative coordinate system (in which the wind direction θ_{earth} is zero degree eastward with increasing values in counter-clockwise direction) to the observational coordinate system (in which wind direction θ_{obs} is defined as the direction where the wind comes from) according to $\theta_{obs} = \text{mod}(270^\circ - \theta_{earth})$.

4.2.2 Conventional in situ observations

Conventional in situ observations are obtained from the *NCEP ADP Global Upper Air and Surface Weather Observations*, dataset *ds337.0*, stored and made publicly available at the Research Data Archive (RDA) at NCAR (<https://rda.ucar.edu/datasets/ds337.0>). The dataset comprises the majority of conventional surface and upper air observations, which are operationally collected by NCEP via the Global Telecommunication System (GTS) and thereafter enter NCEP's various assimilation systems (among others, the Global Data Assimilation System used by the GFS model). This includes reports from land surface stations (SYNOP, METAR) and marine surface (ships, bouys and platforms), upper-air measurements (radiosondes, pilot balloons and dropsondes) as well as aircrafts. Observed variables are pressure, specific humidity, temperature, wind speed and direction. As the dataset provides the final form of observations prior to the assimilation, each observation has undergone a quality check beforehand and is assigned with a corresponding quality flag. Thereupon, no further quality check is conducted in this investigation and the selection of observations follows the same threshold as it is the standard in NCEP's assimilation systems.

4.2.3 Measurement towers

Records from the onshore measurement towers Cabauw, Falkenberg, Hamburg, Jülich and Karlsruhe as well as the three offshore research platforms FINO1, FINO2, and FINO3 have been collected. The temporal resolution of records is 10 minutes for all measurement towers, as averages of more frequent measurements. Model derived winds are thereupon averaged in the same manner for comparison. The position of all measurement towers together is shown in context of the installed onshore wind power capacity in Figure 4.2.

The measurement towers are subject to sites with different topography and surface features, resulting in varying complexity of planetary boundary layer dynamics and therefore varying predictability of wind speeds. Site characteristics of all onshore towers are summarized in Table 4.2, together with the height level used for evaluation. The terrain complexity has been grouped into classes depending on the subgrid-scale variance of topography σ_t on the outer domain. The corresponding terrestrial data is provided by the U.S. Geological Survey (USGS), as it is the default dataset used by the WRF model. Adopting the classification of *Drechsel et al. (2012)*, the terrain of the Cabauw, Hamburg and Falkenberg site can be grouped as flat with $\sigma_t < 25$ m, and the terrain of the Jülich and Karlsruhe site as moderately complex with $\sigma_t > 75$ m. Surface roughness is given for the sites of Cabauw, Hamburg, Falkenberg and Karlsruhe as derived by experimental studies. Appropriate studies are missing for the Jülich tower. Taking the Jülich site characteristics into account and using a simple approximation of the roughness length $z_0 = rH$, with H the height of the surrounding surface element and $0.08 < r < 0.15$ (*Hansen, 1962*), an estimation of $z_0 \sim 1.5$ appears reasonable. The measurement towers at Cabauw and Falkenberg can be grouped upon surface roughness into *low-roughness* sites, and the measurement towers at Hamburg, Jülich and Karlsruhe into *high-roughness sites*, respectively. Surface roughness at the Hamburg site is highly dependent on wind direction, since the surroundings are heterogeneous within a range of 200 m. The concept of roughness length is formally introduced in Section 4.5.1, together with the WRF model equivalent for all sites.

Anemometers are installed on three booms per level at the Cabauw and Falkenberg tower and selected according to the wind direction to avoid shadowing effects. The Hamburg, Jülich and Karlsruhe towers provide solely one boom per level. Here, the largest observations errors due to mast effects can be expected for the Hamburg tower, as it is a radio tower with considerable diameter. Therefore, a correction dependent on wind direction according to *Jacob (2013)* has been applied, which involves the evaluation of a 13th-order polynomial. Measurements at anemometer heights closest to 100 m are taken for model evaluation. Exceptions from this are the Cabauw and Jülich towers. Here, measurements are utilized

Table 4.2: Site characteristics of onshore measurement towers. Classification of topography is adopted from *Drechsel et al.* (2012). Land use and roughness length is given according to references.

Tower	Topography	Land use	Roughness Length [m]
Hamburg	flat	croplands, gardens, buildings and industrial area (<i>Brümmer et al.</i> , 2012)	0.13 – 0.96 (<i>Konow</i> , 2015)
Cabauw	flat	cropland, maize fields to the west (<i>van Ulden and Wieringa</i> , 1996)	0.05 – 0.15 (<i>Beljaars and Holstag</i> , 1991)
Falkenberg	flat	grassland and agricultural (<i>Beyrich and Adam</i> , 2007)	0.02 (<i>Beyrich and Adam</i> , 2007)
Karlsruhe	moderately hilly	broadleaf forest (<i>Wenzel et al.</i> , 1991)	1.5 (<i>Thomas and Vogt</i> , 1993)
Jülich	moderately hilly	clearing surrounded by trees and buildings	unknown, ~ 1.5

at 140 m and 120 m, respectively, by reason of least disturbances (upon personal communication with Fred Bosvelt, Royal Netherlands Meteorological Institute (KNMI), and Axel Knaps, Forschungszentrum Jülich (FZJ)). Wind direction is measured at the same height as wind speed for all onshore towers.

Three offshore research platforms have been installed in the frame of the *Forschungsplattformen in Nord- und Ostsee* (FINO) project of the German government. FINO1 and FINO3 are located in the North sea, FINO1 45 km north to the island of Borkum and FINO3 80 km west to the island of Sylt. FINO2 is located in the Baltic Sea 33 km north to the island of Rügen. FINO1 and FINO2 provide anemometers on two booms per level, whereas FINO3 provides three booms, with the same procedure of selection as described above for the onshore measurement towers. At FINO1 and FINO2, wind speed is measured at a height of 102 m with the corresponding wind direction at 90 m. At FINO3, wind speed is measured at 108 m and wind direction at 100 m.

A comparison of the measurement towers' locations with the spatial distribution of installed wind power capacity illustrates the limited observability of hub height winds in areas of pronounced wind power capacity (see Figure 4.2). The Hamburg measurement tower is located closest to areas of pronounced wind power capacity, whereas the site characteristics of such areas are rather comparable to the surroundings of the Cabauw and Falkenberg tower.

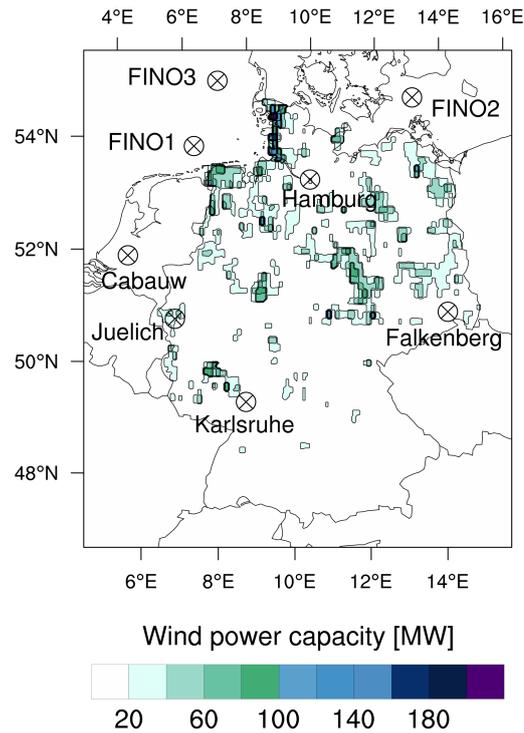


Figure 4.2: Spatial distribution of onshore wind power capacity over Germany, as of 2014, and location of measurement towers. Installed offshore capacity is not included in the IWES wind power model.

4.3 IWES physical wind power model

The winds derived by the WRF model are converted to wind power using a physical wind power model applied at the Fraunhofer IWES (Vogt *et al.*, 2016). Conversion of 100 m wind speeds to wind power is realized by an *equivalent* power curve described in detail in McLean (2008). The *equivalent* power curve provides an averaged estimate of wind power over a region of typically diverse turbines and compensates for shadowing effects within the wind farm as well as topographical and electrical losses. In this manner, wind power is estimated defining gridded regions over Germany, where each grid square defines the sum of all encompassed wind power plants. The normalized feed-in at each grid point given by the *equivalent* power curve and the 100 m wind speed is then multiplied by the capacity of all associated wind power plants to yield the absolute value. The aggregated value for Germany is finally given by the sum of all grid points. The TSO provide all required metadata for the wind power model, which includes the individual power plant’s postal code as a location, the capacity as well as the commission and decommission date. Since locations of wind power plants are not given precisely, the installed capacity has to be assigned to each grid square corresponding to the area portion of the postal code.

Figure 4.2 shows the installed wind power capacity over Germany, as it enters the

wind power model. In 2014, the overall onshore capacity has been 35 GW. Offshore wind parks are not included in the wind power model. Since this work restricts itself on periods within the year 2014, this simplification is assumed to have a negligible effect on model performance, as the installed offshore capacities has been less than 1 GW in this year.

The wind power model described above is simple in its design as it is fundamentally physical, i.e. it does not rely on any historic data on which it is trained upon. It has been shown by *Vogt et al.* (2016), that a significant error reduction can be achieved by combining the physical wind power model with a reference wind power plant model. However, a combination with a statistical model does not appear to be suitable for this work, as the resulting wind power forecasts shall reflect the ensemble's statistical properties and skills, without any additional correction due to mapped statistics from the past. Further details about the wind power model and its performance can be found in Section III.B of *Vogt et al.* (2016).

4.4 Sensitivity of the WRF model to physical parameterizations

A sensitivity study is carried out to determine the best parameterization configuration of the WRF model with regard to the day-ahead wind power forecast over Germany. The WRF model is known for the variety of parameterizations schemes that are offered to the user. Numerous studies have demonstrated the model's sensitivity to the planetary boundary layer (PBL) parameterization (*Cohen et al.*, 2015; *García-Díez et al.*, 2013), land-surface model (LSM) (*Jiming et al.*, 2010; *Ruiz et al.*, 2010), microphysics parameterization (*Min et al.*, 2015; *Pieri et al.*, 2015), radiation parameterization (*Borge et al.*, 2008; *Walaszek et al.*, 2014) and cumulus parameterization (*K. Gilliland and Rowe*, 2007; *Pennelly et al.*, 2014). Finding the optimal combination of parameterizations schemes for any particular region is crucial for the model's forecast skill. *Krieger et al.* (2009) emphasizes the importance of interactions between various parameterization schemes and concludes, that variation of one parameterization at a time relative to a control is insufficient. Rather, different configurations have to be tested to find the optimal setup. In the following, it is proceeded in the same manner. Results are evaluated against conventional in situ observations, records from measurements towers and the day-ahead forecast of associated wind power.

4.4.1 Experimental setup

The sensitivity study is confined to eight different combinations of parameterizations, which are summarized in Table 4.3. Different configurations are numerated in

Table 4.3: Summary of eight WRF configurations with different physical parameterizations.

Setup	Microphysics	Shortwave	Longwave	LSM	PBL	Cumulus
WRF1	Kessler	Dudhia	RRTM	Thermal Slab	YSU	Kain-Fritsch
WRF2	WSM6	Dudhia	RRTM	Thermal Slab	YSU	Kain-Fritsch
WRF3	WSM5	RRTMG	RRTMG	Thermal Slab	MYJ	Grell-Freitas
WRF4	WSM6	RRTMG	RRTMG	Thermal Slab	MYNN	Grell-Freitas
WRF5	Kessler	Dudhia	RRTM	Unified Noah	YSU	Kain-Fritsch
WRF6	WSM5	Dudhia	RRTM	Unified Noah	MYJ	Grell-Freitas
WRF7	WSM6	RRTMG	RRTMG	Unified Noah	YSU	Grell-Freitas
WRF8	WSM6	RRTMG	RRTMG	RUC	MYNN	Betts-Miller

ascending order from WRF1 to WRF8. One has to be aware of the potential incompatibility of certain schemes. Therefore, standard combinations are tested, which are also listed in *Skamarock et al. (2008)*. The different parameterization schemes are described in the following. For a detailed description, the reader is referred to the corresponding references.

Modeled hub height winds are expected to show the highest sensitivity to the PBL parameterization and the LSM, which is an agreement with the studies of *Yang et al. (2013)*, *Wharton et al. (2013)* and *Siuta et al. (2017)*. Thereupon, the PBL parameterization and the LSM are investigated in greater variety and described in more detail. The LSM computes the surface heat and moisture fluxes over land (and for sea ice) determined by radiative forcing, precipitation and exchanges with the surface layer scheme, to provide the lower boundary conditions for their vertical transport throughout the PBL. The PBL parameterization computes the transport of unresolved vertical fluxes (momentum, heat and moisture) due to turbulent eddies in the PBL and the free atmosphere. The following PBL parameterization and LSM model have been tested:

Planetary Boundary Layer parameterization

- Yonsei University (YSU) PBL scheme: first-order non-local closure scheme, countergradient correction term in downgradient diffusion, explicit treatment of the entrainment layer at the PBL top, PBL height diagnosed from buoyancy profile (*Hong et al., 2006*),
- Mellor-Yamada-Janjic (MYJ) PBL scheme: 1.5-order local closure scheme, prognostic equation for turbulent kinetic energy (TKE) for eddy diffusion and diagnostic of PBL height (*Janjić, 1994*),
- Mellor-Yamada-Nakanishi and Niino Level 3 (MYNN3) scheme: second-order local closure scheme, inclusion of buoyancy in the parameterization of pressure covariances (in opposition to MYJ scheme), prognostic equation for turbulent

kinetic energy for eddy diffusion and diagnostic of PBL height (*Nakanishi and Niino, 2009*).

To each of the PBL schemes, at least one complementary surface layer scheme exists, and it is referred to *Skamarock et al. (2008)* and references within for a description.

Land-surface model

- 5-layer Thermal Diffusion (Thermal Slab): simple model, temperature prediction and constant soil moisture, inclusion of snow cover flag, no explicit vegetation effects, 5 soil layers (*Dudhia, 1996*),
- Noah Land Surface Model: temperature and soil moisture (4-layers) and frozen soil (1-layer) prediction with force-restore method, accounts for vegetational and hydrological processes (*Tewari et al., 2004*),
- NOAA Rapid Update Cycle (RUC): temperature and soil moisture prediction (default 6-layers), as well as frozen soil (2-layer) with layer approach for the energy budget, accounts for vegetational and hydrological processes (*Benjamin et al., 2004*).

Feedback from convective and shallow clouds effects on the resolved scale are parameterized on the outer domain, whereas the inner domain is convection-permitting. The Kain-Fritsch scheme (*Kain, 2004*), the Betts-Miller-Janjic scheme (*Janjić, 1994*) and the Grell-Freitas Ensemble scheme (*Grell and Freitas, 2014*) have been tested for this purpose. The Purdue Lin scheme (*Lin et al., 1983*) has also been tested, but appeared to be numerically unstable and results with this scheme are not shown here. The parameterization of microphysical processes, considering the growth of cloud droplet and ice crystals as well as the associated fallout, has been tested with the Kessler scheme (*Kessler, 1969*), the WRF Single-Moment 5-Class (WSM5, *Hong et al. (2004)*) and 6-Class Microphysics scheme (WSM6, *Hong and Lim (2006)*). The Kessler scheme is a simple warm cloud scheme, i.e. solely water vapor, cloud water and rain are considered. In opposition, the WSM5 is a more sophisticated scheme which considers freezing and melting processes (cloud ice and snow), each handled separately to allow for mixed-phase clouds. By the inclusion of graupel, the WSM6 scheme is an extension of the WSM5 scheme. Atmospheric radiation, i.e. the calculation of radiative budgets in the atmosphere and the total radiative flux at the surface, is parameterized for the longwave with the Rapid Radiative Transfer Model (RRTM, *Mlawer et al. (1997)*) or its extension, the RRTMG (also suitable for general circulation models, *Iacono et al. (2008)*), which allows for random cloud overlap. Shortwave radiation is handled either by the Dudhia scheme (*Dudhia, 1989*) or by the RRTMG scheme.

4.4.2 Evaluation Methodology

Retaining the nomenclature of Chapter 2, \mathbf{x}^f denotes the forecast model state and \mathbf{y} the observation vector. Let T denote the number of forecast-observation pairs. The sample mean forecast is then defined as $\bar{x}^f = \frac{1}{T} \sum_{i=1}^T x_i^f$, and the sample mean observation as $\bar{y} = \frac{1}{T} \sum_{i=1}^T y_i$. The continuous verification statistics of root mean square error

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{i=1}^T (x_i^f - y_i)^2}, \quad (4.3)$$

overall bias or mean error

$$\text{BIAS} = \frac{1}{T} \sum_{i=1}^T (y_i - x_i^f) = \bar{y} - \bar{x}^f, \quad (4.4)$$

and the Pearson correlation coefficient

$$\text{CORR} = \frac{\sum_{i=1}^T (x_i^f - \bar{x}^f)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^T (x_i^f - \bar{x}^f)^2} \sqrt{\sum_{i=1}^T (y_i - \bar{y})^2}}, \quad (4.5)$$

are computed for conventional in situ observations, records from measurement towers and the derived wind power forecast for Germany.

In case of conventional in situ observations, observations outside Germany are sorted out by applying a point in polygon test (*Hacker, 1962*) based on the Jordan Curve Theorem. Within this scheme, an arbitrary ray is shoot from the location of interest and the number of intersections with the polygon of the German border is counted. If and only if the number is odd, the location lies inside the polygon and vice versa.

4.4.3 Overview of the sensitivity study

The sensitivity study is carried out over the periods of 1 – 31 August 2014 and 1 – 30 November 2014. Both periods were chosen in order to include extreme error events in the energy forecast, criteria as exposed in Chapter 5. The eight different WRF configurations are each initialized daily at 00 UTC by the control GFS analysis (see Section 4.1.2) and the forecast lead time is 48 hours. Verification statistics are computed according to the day-ahead wind power forecast, i.e. only lead times of 24 – 48 hours are considered. Verification statistics for the GFS forecast fields are computed for comparison. Feedback from the inner to the outer nest is disabled to investigate the potential benefit of an integrated high-resolution nest. All results are given in Coordinated Universal Times (UTC).

4.4.4 Results

Conventional in situ observations

The evaluation of numerous conventional in situ observations aims to derive a comprehensive picture of the WRF model forecast skill. Verification statistics are computed for 10 m wind speed, 2 m temperature, 2 m specific humidity and surface pressure, as well as upper air wind speed and temperature at 850 hPa, 500 hPa and 250 hPa height. Observations are collected from METAR and SYNOP stations as well as radiosondes. Marine surface observations are excluded to be consistent with the IWES wind power model, which includes solely onshore wind capacity. Between 24 – 48 hours forecast lead time and in 3-hourly intervals, matching forecast-observation pairs are collected and a single verification statistic is calculated for root mean square error, bias and correlation. Matching forecast-observation pairs for surface pressure are only included if the difference in terrain height does not exceed a threshold of 25 m, and no verification statistics for surface pressure are calculated for GFS forecast fields due a lower representativity.

Table 4.4 and 4.5 summarize the results for the eight different WRF configurations (inner domain) and GFS forecast fields, separately for forecast periods 1 – 31 August 2014 and 1 – 30 November 2014, respectively. Bold numbers indicate the lowest root mean square error and bias as well as the highest correlation. The evaluation focuses on the results obtained on the inner domain.

Starting with the forecast period 1 – 31 August 2014, it can be summarized that the forecast skill of the eight different WRF configurations is noticeably different and varies among the forecast variables. Since different configurations share partly the same parameterizations, one can clearly determine a correlation of forecast skill. As claimed above, surface winds show the largest sensitivity to the PBL parameterization. The MYNN PBL scheme leads to the lowest root mean square error (WRF4 and WRF8), mainly due to a low bias, whereas the MYJ PBL scheme shows a significant positive bias (WRF3 and WRF6). The forecast skill of the YSU PBL scheme, utilized by the remaining configurations, can be grouped in between. However, within this group, the simple Thermal Slab LSM (WRF1 and WRF2) appears to perform superior to the more complex Noah LSM (WRF5 and WRF7). This results is surprising, as the Noah LSM includes soil moisture prediction, which should have a beneficial effect on the stability estimation throughout the boundary layer. A similar picture can be drawn for surface temperature, except that the simple Kessler microphysics scheme produces a significant negative bias (WRF1 and WRF5). Verification statistics for surface pressure are quite similar for all eight configurations, with the highest root mean square error for WRF1 and WRF5, most likely caused by the simple Kessler microphysics scheme. The same holds for the surface humidity, where the configuration

WRF2 shows the best results. As can be expected, correlations between forecast skill and parameterization setup are less pronounced for upper air variables. A sensitivity of verification statistics to the radiation parameterization can not be identified.

Overall, there is only little added forecast value by dynamically downscaling the GFS forecast fields with the WRF model, at least for periods under investigation and the given aggregated verification statistics over Germany. Despite the lower resolution of GFS forecast fields, a competitive forecast skill can be noted for surface wind speeds, surface humidity and upper air temperature. Moreover, the WRF model degrades the forecast of 850 hPa and 500 hPa wind speed, mainly due to lower correlations. Solely in case of surface temperature, the WRF model adds forecast skill. This is mainly caused by a well known bias in the diurnal cycle of the GFS model. It should be stressed again, that the GFS model actually possessed a horizontal resolution of 28 km during the periods of interest, whereas analysis and forecast fields are only available to the public in a lower resolution of 55 km. Most likely, this is the main reason for the observed forecast degradation, as the coarser resolution of initial conditions smoothes smaller scale features. This effect can not be counterbalanced by an increase in orography representation of the WRF model grid. However, improvements are to be expected in case of an independent regional data assimilation system for WRF.

Among the eight WRF configurations, WRF2 appears to be most the favorable. One can note the lowest root mean square error for 850 hPa temperature, 250 hPa wind speed and temperature as well as surface humidity. Further, in case of the remaining forecast variables, no exceptional error statistic can be observed.

Results for the forecast period 1 – 30 November 2014 show a similar picture concerning the intercomparison of different WRF configurations. Noticeably, the verification statistics for the November period are higher in general. Here, the 10 m and 850 hPa wind speed show the largest bias. As this is not the case for the GFS forecast fields, this leads to the assumption that the lower resolution initial conditions of the WRF model do not sufficiently resolve the stronger wintery baroclinicity, degrading the forecast skill. As a result, the WRF model shows worse verification statistics for nearly all upper air variables. In this context, the MYNN PBL parameterization shows again the lowest errors for surface variables (WRF4 and WRF8).

Results for the outer domain are not shown explicitly. The computational effort of utilizing a high-resolution and convection-permitting nest pays off in a reduction of root mean square error by approximately 10% for all surface variables and all WRF model configurations. This effect is due both, an increase in correlations and a decrease in bias. Since this is observed for all WRF model configurations, one can not identify a favorable cumulus scheme. Although clearly noticeable, the benefit of utilizing an inner domain is less pronounced for upper air variables.

Table 4.4: Verification statistics for day-ahead forecasts (24 – 48 hours forecast lead time) with eight different WRF configurations and GFS model fields for conventional in situ observations. The forecast period is 1 – 31 August 2014. Results for the WRF model are given in case of the inner domain. Bold numbers indicate the lowest root mean square error and bias as well as the highest correlation. The number of observations involved is given at the bottom.

	GFS	WRF1	WRF2	WRF3	WRF4	WRF5	WRF6	WRF7	WRF8
RMSE									
U _{10m} [m/s]	1.60	1.60	1.59	1.78	1.54	1.64	1.86	1.64	1.53
U _{850hPa} [m/s]	2.36	2.56	2.47	2.61	2.49	2.63	2.62	2.54	2.46
U _{500hPa} [m/s]	2.55	2.87	2.80	2.74	2.66	2.89	2.75	2.66	2.81
U _{250hPa} [m/s]	4.68	5.45	4.65	4.77	4.73	5.46	4.75	4.75	4.67
T _{2m} [K]	2.24	2.42	1.65	1.71	1.61	2.60	1.81	1.86	1.54
T _{850hPa} [K]	1.04	0.93	0.73	0.83	0.81	1.00	0.86	0.83	0.78
T _{500hPa} [K]	0.91	0.94	0.84	0.84	0.85	0.92	0.84	0.86	0.85
T _{250hPa} [K]	1.29	2.09	1.55	1.55	1.53	2.15	1.60	1.59	1.46
Q _{2m} 10 ⁻³ [g/g]	1.01	1.00	0.95	1.33	1.33	1.13	0.98	1.02	1.06
P _{SFC} [hPa]	-	3.30	3.16	3.14	3.14	3.27	3.15	3.10	3.15
BIAS									
U _{10m} [m/s]	-0.16	0.02	0.13	0.71	0.06	0.25	0.88	0.36	-0.06
U _{850hPa} [m/s]	-0.29	-0.03	-0.07	0.03	0.01	0.08	0.23	-0.06	0.06
U _{500hPa} [m/s]	- 0.11	-0.29	-0.38	-0.24	-0.20	-0.26	-0.22	-0.22	-0.17
U _{250hPa} [m/s]	- 0.64	-1.45	-1.16	-1.25	-1.03	-1.45	-1.11	-1.00	-0.93
T _{2m} [K]	-0.27	-1.30	0.02	0.32	0.06	-1.54	-0.23	0.55	-0.11
T _{850hPa} [K]	0.60	-0.24	0.01	0.13	-0.05	-0.43	-0.13	0.36	-0.14
T _{500hPa} [K]	0.33	-0.13	0.12	0.17	0.24	-0.09	0.15	0.34	0.24
T _{250hPa} [K]	0.59	1.34	0.93	0.98	0.92	1.38	1.05	1.05	0.83
Q _{2m} 10 ⁻⁴ [g/g]	- 0.74	-0.94	1.44	8.94	8.63	-5.63	-1.19	-1.17	6.12
P _{SFC} [hPa]	-	0.84	0.48	0.15	0.13	0.73	0.35	-0.20	0.21
CORR									
U _{10m}	0.30	0.35	0.37	0.41	0.40	0.35	0.41	0.36	0.38
U _{850hPa}	0.66	0.56	0.61	0.58	0.61	0.55	0.61	0.59	0.60
U _{500hPa}	0.71	0.66	0.66	0.69	0.69	0.66	0.66	0.67	0.65
U _{250hPa}	0.77	0.67	0.78	0.75	0.75	0.68	0.74	0.75	0.77
T _{2m} [K]	0.50	0.56	0.68	0.65	0.68	0.55	0.64	0.66	0.70
T _{850hPa}	0.85	0.78	0.87	0.84	0.84	0.75	0.84	0.87	0.84
T _{500hPa}	0.83	0.84	0.85	0.85	0.86	0.84	0.84	0.86	0.84
T _{250hPa}	0.74	0.50	0.62	0.64	0.63	0.47	0.63	0.64	0.64
Q _{2m}	0.55	0.45	0.52	0.49	0.48	0.47	0.51	0.51	0.54
P _{SFC}	-	0.77	0.79	0.78	0.78	0.77	0.77	0.77	0.77
Number of observations									
U _{10m}	36,397								
U _{850hPa}	1,166								
U _{500hPa}	1,139								
U _{250hPa}	1,136								
T _{2m}	37,758								
T _{850hPa}	1,144								
T _{500hPa}	1,141								
T _{250hPa}	1,137								
Q _{2m}	37,512								
P _{SFC}	16,878								

Table 4.5: Verification statistics for day-ahead forecasts (24 – 48 hours forecast lead time) with eight different WRF configurations and GFS model fields for conventional in situ observations. The forecast period is 1 – 30 November 2014. Results for the WRF model are given in case of the inner domain. Bold numbers indicate the lowest root mean square error and bias as well as the highest correlation. The number of observations involved is given at the bottom.

	GFS	WRF1	WRF2	WRF3	WRF4	WRF5	WRF6	WRF7	WRF8
RMSE									
U _{10m} [m/s]	1.81	1.94	1.93	2.23	1.81	1.83	2.08	1.83	1.63
U _{850hPa} [m/s]	2.61	3.08	3.06	3.10	3.03	3.08	3.11	3.04	3.08
U _{500hPa} [m/s]	3.29	3.69	3.52	3.52	3.58	3.71	3.57	3.56	3.57
U _{250hPa} [m/s]	4.59	6.06	4.89	5.02	4.85	6.07	4.96	4.77	4.91
T _{2m} [K]	2.22	1.95	1.90	1.9	1.88	2.18	2.37	2.19	1.79
T _{850hPa} [K]	1.41	1.50	1.44	1.48	1.48	1.50	1.51	1.44	1.46
T _{500hPa} [K]	1.03	1.20	1.05	0.98	1.01	1.20	0.98	0.99	1.00
T _{250hPa} [K]	1.13	2.12	1.55	1.61	1.56	2.09	1.58	1.59	1.52
Q _{2m} 10 ⁻³ [g/g]	0.69	0.62	0.63	0.71	0.71	0.78	0.74	0.70	0.63
P _{SFC} [hPa]	-	3.38	3.47	3.53	3.55	3.37	3.48	3.39	3.51
BIAS									
U _{10m} [m/s]	-0.65	0.63	0.72	1.25	0.48	0.48	1.04	0.60	-0.03
U _{850hPa} [m/s]	-0.68	0.69	0.81	0.84	0.82	0.72	0.80	0.73	0.84
U _{500hPa} [m/s]	-0.81	0.01	0.25	0.25	0.44	0.17	0.37	0.36	0.41
U _{250hPa} [m/s]	-0.96	-0.14	-0.63	-0.62	-0.41	-0.03	-0.55	-0.49	-0.65
T _{2m} [K]	-0.27	0.01	0.12	0.36	0.44	-0.34	-0.49	0.51	0.02
T _{850hPa} [K]	-0.03	0.21	-0.11	0.16	0.01	0.29	0.29	0.12	-0.13
T _{500hPa} [K]	0.52	0.43	0.27	0.10	0.15	0.43	-0.03	0.17	0.15
T _{250hPa} [K]	0.47	0.46	0.89	1.09	1.00	0.48	1.04	1.06	0.92
Q _{2m} 10 ⁻⁴ [g/g]	-3.57	0.87	0.81	3.28	3.33	-4.51	-3.78	-2.90	1.68
P _{SFC} [hPa]	-	-0.07	0.50	0.49	0.39	-0.16	0.46	0.14	0.49
CORR									
U _{10m}	0.33	0.48	0.48	0.50	0.48	0.48	0.51	0.47	0.50
U _{850hPa}	0.66	0.59	0.63	0.62	0.63	0.58	0.63	0.62	0.62
U _{500hPa}	0.72	0.68	0.70	0.70	0.69	0.68	0.69	0.70	0.72
U _{250hPa}	0.77	0.64	0.75	0.72	0.75	0.63	0.73	0.76	0.74
T _{2m}	0.51	0.72	0.73	0.71	0.74	0.69	0.68	0.70	0.75
T _{850hPa}	0.77	0.74	0.73	0.74	0.73	0.75	0.73	0.74	0.72
T _{500hPa}	0.72	0.74	0.78	0.78	0.77	0.74	0.78	0.78	0.77
T _{250hPa}	0.69	0.50	0.56	0.59	0.59	0.51	0.59	0.59	0.60
Q _{2m}	0.54	0.54	0.55	0.54	0.54	0.52	0.52	0.53	0.57
P _{SFC}	-	0.85	0.86	0.84	0.85	0.85	0.85	0.85	0.85
Number of observations									
U _{10m}	36,539								
U _{850hPa}	1,160								
U _{500hPa}	1,160								
U _{250hPa}	1,159								
T _{2m}	37,556								
T _{850hPa}	1,161								
T _{500hPa}	1,160								
T _{250hPa}	1,159								
Q _{2m}	36,981								
P _{SFC}	16,420								

Measurement towers

Verification statistics of turbine hub height wind speeds are computed at all measurement towers, again in 3-hourly intervals. In case of the WRF model forecast fields, linear interpolation is used to derive wind speeds at anemometer heights (see Section 4.2.3 for a discussion). The GFS forecast fields include 100 m winds routinely, which are taken for comparison. Results at the Cabauw measurement tower are not shown for the GFS model, since the anemometer height used for evaluation deviates significantly from 100 m. A maximum threshold of one-third of missing records is set, which is not exceeded for any measurement towers and all times. Records from the Jülich measurement tower could not be provided for the period of August.

Figure 4.3 and 4.4 show the root mean square error of hub height wind speeds at all available measurement towers for the forecast period 1 – 31 August 2014 and 1 – 30 November 2014, respectively. The corresponding statistics of bias and correlation are shown in Appendix A, Figure A.1 and A.2, as well as Figure A.3 and A.4, respectively.

Starting with the forecast period 1 – 31 August 2014 and offshore measurement towers FINO1, FINO2 and FINO3, wind speeds show only little sensitivity to the WRF model configuration. A complete different picture appears for the onshore measurement towers. Most striking, a positive wind bias can be observed for all WRF model configurations and all measurement towers, except for Cabauw. The tendency of a positive bias in 10 m wind speed has been noted in the previous section. Yet, at turbine hub height the bias appears more pronounced. The overestimation of wind develops right after model initialization and holds throughout the day. At the Hamburg, Falkenberg and Karlsruhe measurement tower, the overestimation is of the order of 1 – 2 m/s. It is noted that the initial negative bias for all onshore sites is caused by the vertical interpolation of wind fields to the WRF model grid, which does not take any stability correction into account.

A pronounced diurnal cycle of verification statistics can only be noticed for the WRF4 and WRF8 configuration, which share the MYNN PBL scheme. The scheme produces too little mixing during stable conditions, leading to an underestimation of boundary layer height and too strong low level jets. Further, a slightly higher bias remains during day time compared to the remaining configurations. For all onshore measurement towers, the combination of the YSU PBL scheme with the Thermal Slab LSM performs best (WRF1 and WRF2).

Surprisingly, it is not straightforward to group the predictability of hub height winds according to the site's surface roughness (see Section 4.2.3). One can notice the highest correlations for the *low-roughness site* Cabauw, whereas the correlations at the *low-roughness site* Falkenberg are comparable to those at the *high-roughness*

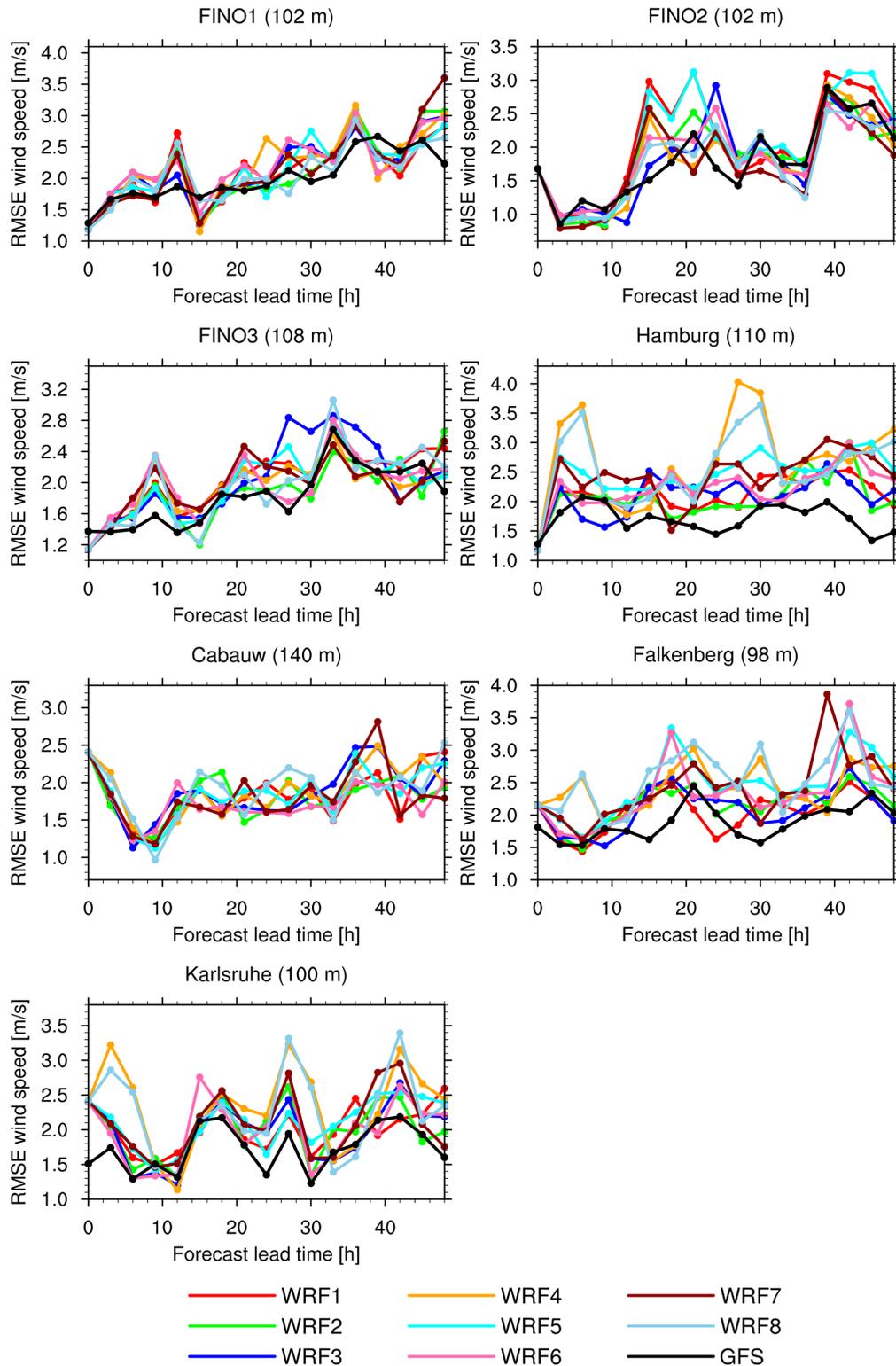


Figure 4.3: Root mean square error of hub height wind speeds at measurement towers for 1 – 31 August 2014. Different colors correspond to eight WRF model configurations. Results for GFS forecast fields are shown in black color for comparison.

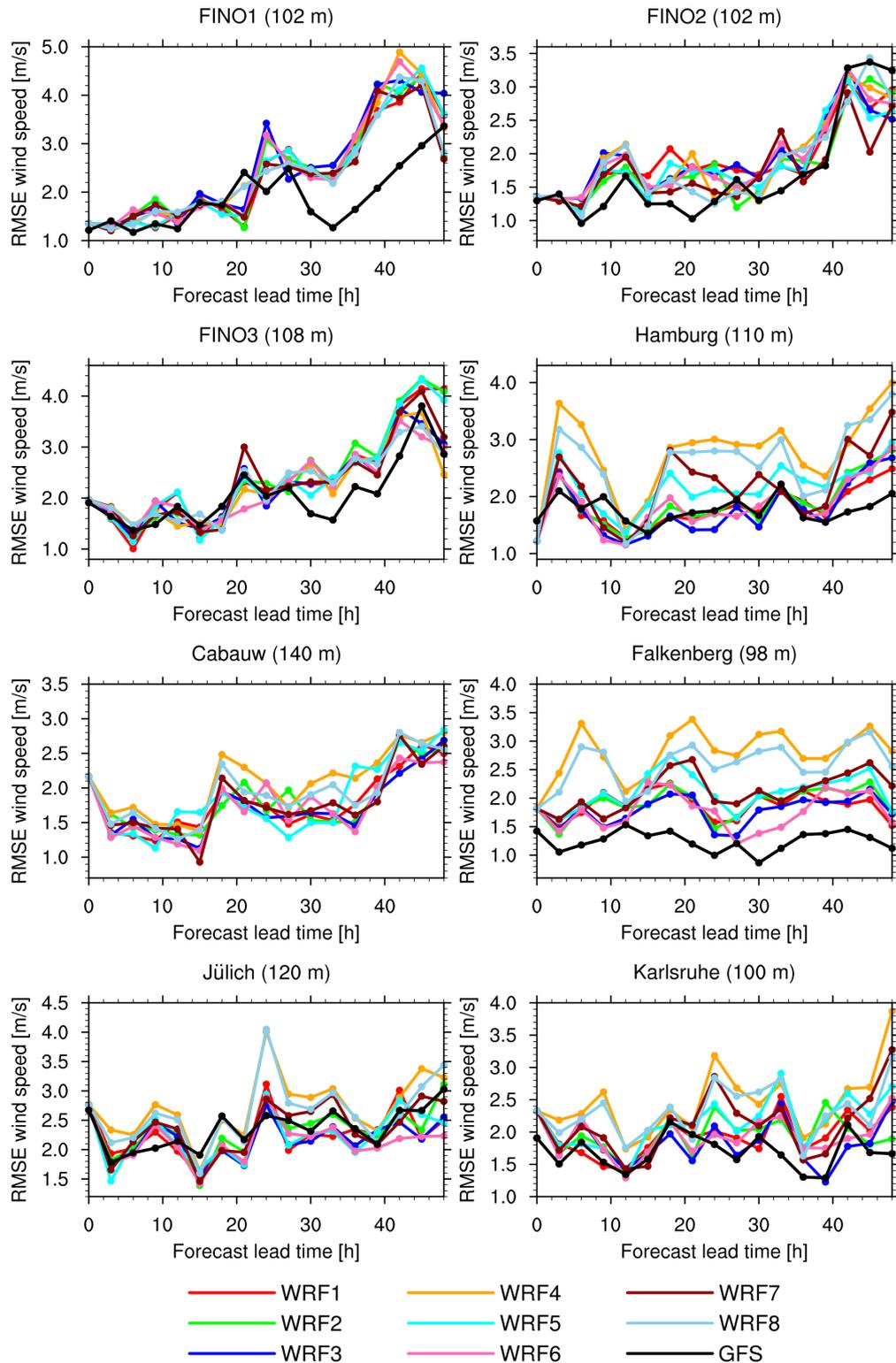


Figure 4.4: Root mean square error of hub height wind speeds at measurement towers for 1 – 30 November 2014. Different colors correspond to eight WRF model configurations. Results for GFS forecast fields are shown in black color for comparison.

site Karlsruhe. Despite the lower resolution forecast fields, wind speeds of the GFS model show the lowest root mean square error at all onshore and offshore sites. This is caused solely due the positive wind bias observed for all WRF model configurations, whereas correlations are of comparable magnitude.

Results for the forecast period 1 – 30 November 2014 show a similar picture for the wind bias. At the Hamburg, Falkenberg, Jülich and Karlsruhe measurement tower, all model configurations tend to overestimate the wind speed. As for the August period, the overestimation of wind speed is of the order of 1 – 2 m/s. Result obtained at the Cabauw measurement tower state an exception. Again, the WRF4 and WRF8 configurations show the most pronounced bias, which holds throughout the day for the November period, and the WRF1 and WRF2 configurations show the lowest bias and lowest root mean square errors.

Wind power

A physical wind power model (see Section 4.3) has been applied by Fraunhofer IWES to derive an aggregated wind power forecast over Germany. Results are compared with the real power feed-in estimated by the TSO, which can be understood as the "observation". Diurnal cycles are calculated for the sample mean, root mean square error, bias and correlation, for the periods 1 – 31 August 2014 and 1 – 30 November 2014. Results are shown in Figure 4.5 and 4.6, respectively. All values are normalized by the total installed wind power capacity over Germany, excluding offshore capacity. For comparison, the statistically postprocessed, multi-model day-ahead forecasts of the German TSOs is shown. Further, 100 m wind fields of the GFS model have been converted to wind power. It may be assumed, that the GFS model data is involved in the multi-model approach of the TSOs day-ahead forecasts. Table 4.6 and 4.7 list the associated aggregated verification statistics throughout the day.

The TSO day-ahead forecast shows a root mean square error of 0.031 for August and 0.030 for November. One may assume, that postprocessing removes a possible bias to an acceptable extend. In comparison, the wind power forecast derived with the GFS model data performs only slightly worse. However, results for the wind power derived with the WRF model show a consistent picture with the previous section. As the cube of the wind speed enters the wind power model, the observed wind speed bias is strongly amplified. This instance eventually leads to an overestimation of wind power in the order of 100 % (Figure 4.5a and 4.6a). In this context, all WRF configurations show a comparably large root mean square error for all forecast lead times and both periods under investigation. However, despite the large bias, correlations are of comparable size, indicating that the WRF model shows in principle a distinct forecast skill.

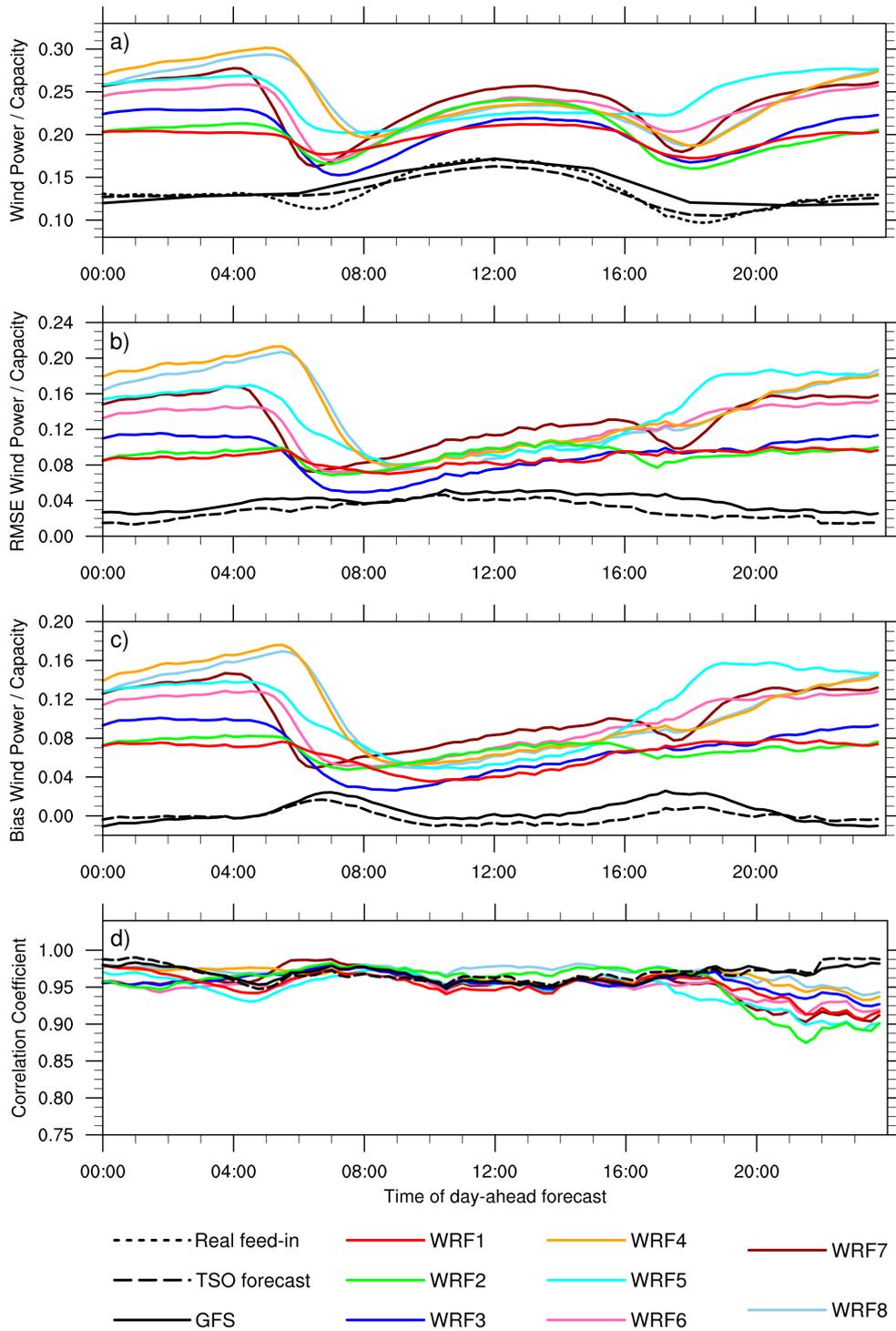


Figure 4.5: Diurnal cycle of verification statistics for the normalized day-ahead wind power forecast with eight different WRF configurations: (a) sample mean, (b) root mean square error, (c) bias and (d) correlation. The forecast period is 1 – 31 August 2014. For comparison, the real power feed-in, the TSO forecast and the derived forecast with GFS model data and the IWES power model is shown.

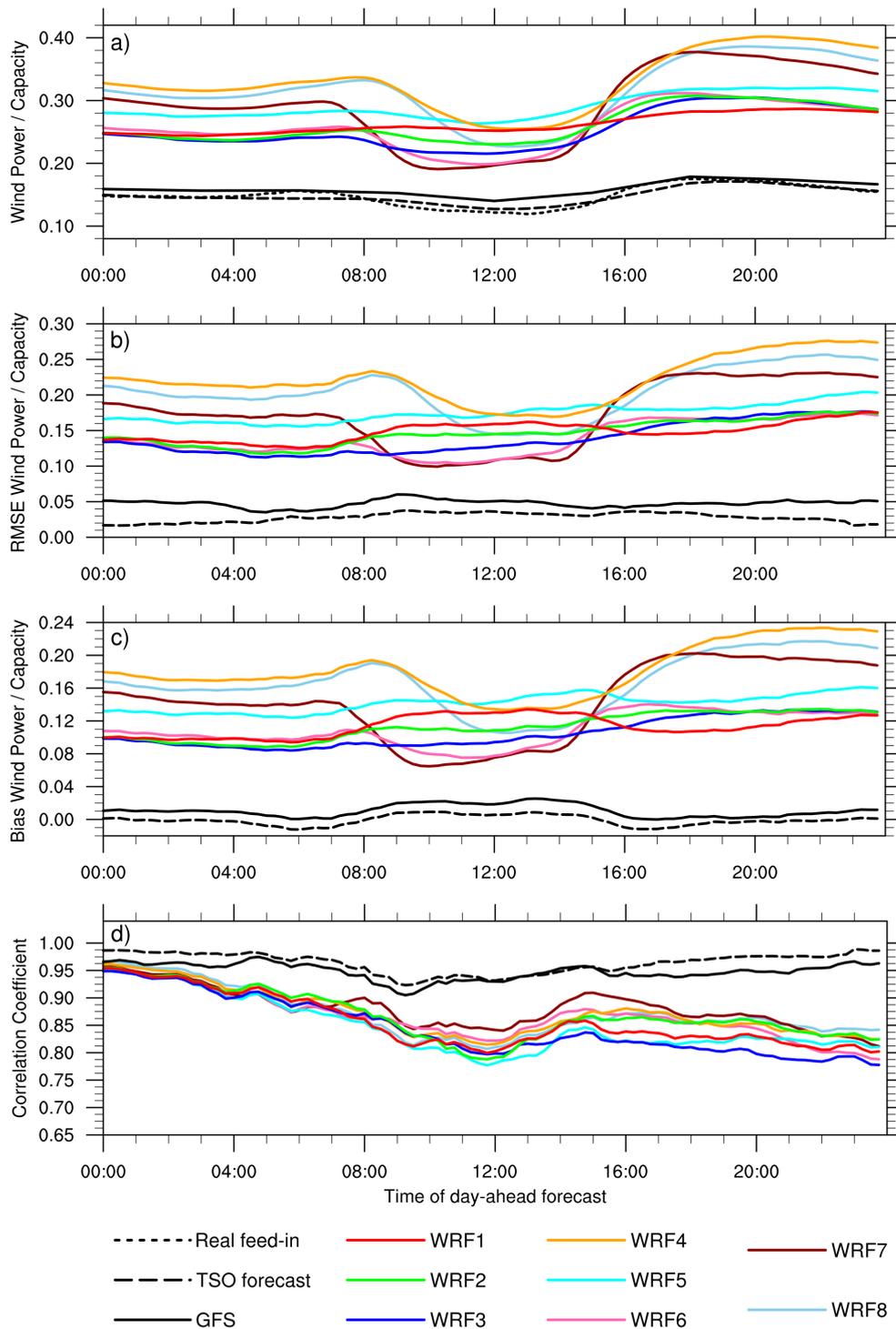


Figure 4.6: Diurnal cycle of verification statistics for the normalized day-ahead wind power forecast with eight different WRF configurations: (a) sample mean, (b) root mean square error, (c) bias and (d) correlation. The forecast period is 1 – 30 November 2014. For comparison, the real power feed-in, the TSO forecast and the derived forecast with GFS model data and the IWES power model is shown.

Table 4.6: Verification statistics for the day-ahead wind power forecast with eight different WRF configurations and GFS model fields. The TSO day-ahead forecast and the derived forecast with GFS model data is shown for comparison. The forecast period is 1 – 31 August 2014. All values are normalized with the installed wind power capacity.

	TSO	GFS	WRF1	WRF2	WRF3	WRF4	WRF5	WRF6	WRF7	WRF8
RMSE	0.031	0.040	0.088	0.092	0.937	0.150	0.142	0.124	0.132	0.146
BIAS	-0.001	0.005	0.063	0.068	0.069	0.108	0.108	0.098	0.103	0.105
CORR	0.969	0.967	0.952	0.954	0.957	0.965	0.946	0.951	0.952	0.971

Table 4.7: Verification statistics for the day-ahead wind power forecast with eight different WRF configurations and GFS model fields. The TSO day-ahead forecast and the derived forecast with GFS model data is shown for comparison. The forecast period is 1 – 30 November 2014. All values are normalized with the installed wind power capacity.

	TSO	GFS	WRF1	WRF2	WRF3	WRF4	WRF5	WRF6	WRF7	WRF8
RMSE	0.0291	0.048	0.149	0.147	0.140	0.223	0.175	0.140	0.179	0.207
BIAS	-0.001	0.011	0.113	0.112	0.105	0.181	0.142	0.109	0.142	0.167
CORR	0.964	0.948	0.856	0.867	0.845	0.872	0.847	0.865	0.88	0.87

The observed wind bias has a pronounced diurnal cycle for all configurations, due to an overestimation of nocturnal low level jets in case of both periods. This is again observed to be most distinct for WRF4 and WRF8, consistent with the previous section. Again, the combination of the YSU PBL scheme and the Thermal Slab LSM appears most favorable (WRF1 and WRF2). This results is surprising, as the simple Thermal Slab LSM does not include soil moisture prediction, and one would expect the more sophisticated LSMs to provide more realistic boundary conditions for the PBL scheme. From this group, configuration WRF2 agrees best with the diurnal cycle of the observations. The WRF2 configuration shows a root mean square error of 0.092 and 0.147 for August and November, respectively.

For the August period, one can notice local minima in the daily mean cycle around sunrise and sunset, visible in the observation as well as for all WRF model configurations. The planetary boundary layer occasionally falls below turbine hub height under stable conditions. After sunrise, wind speed decreases as the inversion passes the turbine hub height and air masses of lower momentum rise. With increasing turbulence induced by convection during day time, higher momentum from upper levels is mixed down and wind speeds increase again. In case of the minimum around sunset, an opposite effect can be assumed.

4.4.5 Discussion on WRF's wind bias

In this section it has been identified, that the WRF model shows a significant positive wind speed at the surface and nearby elevated levels. The wind bias is moderate for 10 m wind speed, but significant at turbine hub heights. The wind bias is further amplified in the wind power forecast, as the derived power is proportional to the cube of wind speed. A diurnal cycle can be observed, with a stronger overestimation of wind speed during night. The wind bias is apparent for all tested model configurations as well as all lead times, and develops right after the initialization. Evidently, this leads to the conclusion of a systematic model deficit.

A systematic overestimation of surface and elevated winds by the WRF model has been noted in numerous studies (*Cheng and Steenburgh, 2005; Bernardet et al., 2008; Roux et al., 2009*). A plausible explanation could be a missing representation of unresolved terrain effects and additional drag at the surface. Different attempts to include unresolved terrain effects in the formulation of the YSU PBL parameterizations have been proposed. *Jiménez and Dudhia (2012)* add a sink term that depends on the standard deviation of the subgrid orography as well as on the Laplacian of the topographic field. *Mass and Ovens (2011)* modulate the friction velocity according to the subgrid orography. Both approaches have been investigated but do not correct wind speeds in the area of pronounced installed wind power capacity, where the terrain is typically flat and subgrid orography vanishes.

The physics suite of configuration WRF2 has proven to be most favorable for forecasts of 100 m wind speed and associated wind power. Further, the configuration has shown the most consistent verification statistics for numerous in situ observations, including upper air temperature and wind speed. Thereupon, for any further investigation, the final physics suite reads: YSU PBL scheme, Thermal Slab LSM, WSM6 microphysics scheme, RRTM longwave and Dudhia shortwave radiation scheme as well as Kain-Fritsch cumulus scheme.

4.5 Optimization of the WRF model for wind power forecasting

In this section it is shown, that with straightforward optimizations of model parameters, the positive wind speed bias at 100 m can be reduced and day-ahead wind power forecasts with comparable skill to the TSO multi-model approach are achievable. The plausibility of such modifications is discussed with the primary argument, that the WRF model is optimized for the Contiguous United States (CONUS).

The most straightforward way to decrease wind speeds above ground is to increase the roughness of the surface. Rougher surfaces increasingly generate drag in

the opposite direction to the flow and enhance turbulent kinetic energy. This induces downward momentum flux which decreases the vertical wind speed gradient. The wind speed bias of the WRF model shows a summerly diurnal cycle with maximum values during night times. Hence, solely increasing the surface roughness is insufficient. The YSU PBL scheme is thereupon modified to enhance mixing during stable conditions.

4.5.1 Roughness length

The surface layer is the lowest layer in the atmosphere and can be distinguished from the remaining PBL by turbulent fluxes which are constant with height. Turbulent motion solely maintains the surface layer, while Coriolis as well as pressure forces can be neglected. The similarity theory of Monin and Obukhov has been a standard to describe the shape of the wind profile under these assumptions and derived the dimensionless vertical wind speed gradient as

$$\frac{kz}{u_*} \frac{\partial \bar{u}}{\partial z} = \phi_m(z/L), \quad (4.6)$$

with k the Kármán constant, z the height above ground, u_* the friction velocity, \bar{u} the mean flow according to Reynolds averaging and L the Obukhov length scale, which estimates the ratio of wind shear and buoyancy in the process of turbulence generation¹. $\phi_m(z/L)$ denotes the stability dependent non-dimensional universal function for the momentum transport which is written here in the most general form depending on the stability conditions:

$$\phi_m\left(\frac{z}{L}\right) = \begin{cases} 1 + a\frac{z}{L} & \text{stable} \\ 1 & \text{neutral} \\ (1 - b\frac{z}{L})^{-1/c} & \text{unstable,} \end{cases} \quad (4.7)$$

with values of the parameters a, b and c being under ongoing debate for the last decades (Foken, 2006). Vertical integration of the gradient in (4.7) leads to a logarithmic wind profile

$$\bar{u} = \frac{u_*}{k} \left[\ln\left(\frac{z}{z_0}\right) - \Phi_m\left(\frac{z}{L}\right) \right], \quad (4.8)$$

with Φ_m the integral of ϕ_m and z_0 the integration constant. For neutral conditions the wind profile takes the well known form $\bar{u} = \frac{u_*}{k} \ln\left(\frac{z}{z_0}\right)$. Parameter z_0 is the roughness length defined as the height where the logarithmic wind profile becomes zero. The roughness length has a direct influence on the shape of the wind profile, with increasing values decreasing $\partial\bar{u}/\partial z$.

¹For a derivation and the corresponding equation for the heat exchange, the reader is referred to Stull (1988).

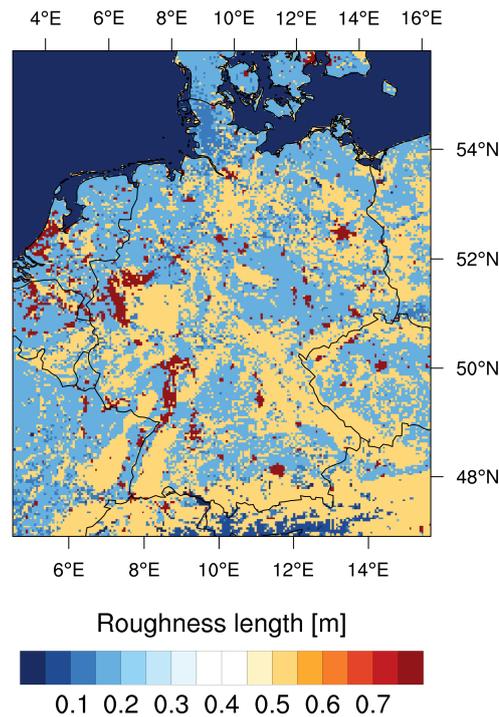


Figure 4.7: Roughness length over Germany according to the WRF model, based on the MODIS land cover type data set.

Hub height winds are strongly effected by z_0 . Although the surface layer may not reach such heights, typically the case for a non convective boundary layer, fluxes derived at the surface layer top still serve as lower boundary conditions for the calculation of their transport throughout the remaining mixed layer of the PBL. In principle, z_0 depends on the composition and protrusions of the surface. The reader is referred to *Pielke* (1984) for a comprehensive summary of theoretical and experimental studies on the derivations of roughness length. Theoretical and experimental studies are typically derived under simplified conditions or ideal locations of to a site, where surface features are less inhomogeneous, and it is therefore argued that z_0 is a plausible parameter of high uncertainty in mesoscale modeling.

The WRF model associates a constant value of z_0 to each land use category, which have to be interpreted as averaged values. Two look-up tables exist for summer and winter values. The Thermal Slab land-surface parameterization does not include interpolation between the biannual values. Figure 4.7 shows values of the roughness length z_0 according to the WRF model based on the MODIS land cover type data set over Germany. Predominant values are 0.15 m (0.05 m during winter) associated with the land use category *Croplands*, which covers the majority of installed wind capacity, and 0.5 m (0.2 m during winter) associated with *Mixed Forests*. Noticeably,

in the area of highest wind power capacity in Germany, the coastal area of the North Sea in Schleswig Holstein (see Figure 4.2 for comparison), z_0 has a value of 0.14 m (0.05 m during winter) associated with *Cropland/Natural Vegetation Mosaic*. Also mentionable at this point is the land use category *Urban/Build-Up* with a maximum value of 0.8 m (constant throughout the year).

In the following it is argued, that the surface roughness is in general assumed too small in the WRF model. At least, this holds for the domain under investigation and for the predominant land use categories mentioned above. The WRF model is by development an American model and optimized for the Contiguous United States (CONUS). Values of roughness length are thereupon not necessarily suitable for other domains, since the same classes of land use categorization may possess quite different characteristics depending on their cultivation. In particular, values of roughness length for agricultural land use have been under active debate (see e.g. *Pielke (1984)*). As the main argument, it is stated here that agricultural land use is by far more inhomogeneous in Europe compared to the CONUS, with typical agglomerations of trees, buildings and roads within a few hundred meters. This assumption is supported by *van Dop (1983)*, who derives a value of 0.25 m for European agricultural land use, which is almost twice as large as in the WRF model.

An increase in surface roughness is justified with respect to other mesoscale models which run operationally over the Germany area. Referring to the COSMO model maintained by the German Weather Service, land use categories are in principle associated with higher values of z_0 . The COSMO model uses a re-evaluation of the combined CORINE and GLC2000 data sets for surface parameter classification. Although these data sets associate agricultural land use with a value of 0.15 m as well, the category of *Mosaic Crop/Tree/Net Vegetation* with 0.25 m is quite widespread over Northern Germany. Carrying on this comparison to further land use categories, surface roughness for *Mixed Forests* is assumed twice as large and urban areas have a value of 1.0 m.

The assumption of too little surface roughness is in agreement with the results obtained in Section 4.4.4 for the measurement towers. For example, the ECOCLIMAP data set (*Masson et al., 2003*) gives a value of 1.117 m and 1.3 m for the sides of the Hamburg and Karlsruhe measurement towers, respectively, and it has been already mentioned in Section 4.2, that *Konow (2015)* derives a value of 0.96 m for the predominant wind direction at the Hamburg side and *Thomas and Vogt (1993)* 1.5 m for Karlsruhe.

4.5.2 The YSU PBL scheme

The Yonsei University (YSU) PBL scheme has been developed by *Hong et al. (2006)* and follows a first-order non-local closure approach pioneered by *Troen and*

Mahrt (1986). The scheme is a revised version of the former Medium-Range Forecast (MRF) PBL scheme, which has been beforehand a standard in NCEP's operational models, the MM5 Model and early versions of the WRF model. A major modification to the MRF PBL scheme is an additional explicit treatment of entrainment processes for momentum at the inversion layer (and not exclusively for moisture and heat), proposed originally by *Noh et al.* (2003). Currently, the YSU scheme is the most frequently used PBL parameterization within the WRF model user community.

K-theory relates the turbulent stresses to vertical gradients of the mean flow via the concept of eddy viscosity (*Stull*, 1988) and the turbulence diffusion equation reads

$$\frac{\partial C}{\partial t} = -\frac{\partial}{\partial z}(\overline{w'c'}) = \frac{\partial}{\partial z}K_c\left(\frac{\partial C}{\partial z}\right), \quad (4.9)$$

with C a prognostic variable of momentum, heat or moisture and K_c the associated eddy viscosity coefficient. The non-local K approach modifies K-theory by adding a parameterized countergradient correction term γ_c to the turbulence diffusion equation accounting for buoyancy induced large-scale eddies in a well mixed environment:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial z}K_c\left(\frac{\partial C}{\partial z} - \gamma_c\right). \quad (4.10)$$

The eddy diffusivity coefficients K_c are estimated within the PBL by a prescribed parabolic function with quantities evaluated over multiple points in the vertical (in opposition to a local approach using local spatial gradients), which is formulated for the momentum as

$$K_m = kw_s\left(1 - \frac{z}{h}\right)^p, \quad (4.11)$$

with w_s the velocity scale and h the PBL height. The exponent p determines the shape and magnitude of the profile. The stable regime is defined by the condition $\overline{(w'\theta')_0} < 1$ in the YSU PBL scheme, with $\overline{(w'\theta')_0}$ the surface flux of potential temperature. The corresponding velocity scale is given by $w_s = u_*/\phi_m$ (in the unstable regime, w_s is modified by a convective velocity scale). As has been noted in the previous section, the shape of the stability universal function ϕ_m is under ongoing debate, but takes the general form $\phi_m = 1 + r\frac{z}{L}$ for stable conditions, see (4.7). However, according to *Hong et al.* (2006), the initial version of the YSU PBL scheme assumed

$$\phi_m = 1 + 5\frac{0.1h}{L}, \quad (4.12)$$

which has been changed after consecutive tuning to

$$\phi_m = 1 + 5\frac{z}{L}. \quad (4.13)$$

Table 4.8: Experimental setup of WRF model optimizations for configuration WRF2. Roughness length z_0 in (4.8) is multiplied by a tuning parameter r , stability in the YSU PBL scheme is enhanced in the stable regime by parameter s in (4.14).

	WRF2	WRF2r2s1	WRF2r3s1	WRF2r2s0.5	WRF2r3s0.25	WRF2r3s0.1
r	1.0	2.0	3.0	2.0	3.0	3.0
s	1.0	1.0	1.0	0.5	0.25	0.1

This modification has been done ”to prevent smoothing nocturnal jets too much because the z/L term smooths less with height than $0.1h/L$ ” (personal communication with Jimy Dudhia, NCAR). This conclusion has been drawn on the CONUS domain with a different orography and climatology. Since this study observes the opposite, that is too strong nocturnal low level jets, it appears to be justified to reformulate (4.13) as

$$\phi_m = 1 + 5s \frac{z}{L}, \quad (4.14)$$

with $0 < s \leq 1$ being a tuning parameter. With typical PBL heights of 100 m during summerly stable conditions over Germany, (4.13) and the initial implementation (4.12) equal at the inversion layer height in case of $s = 0.1$. Although this range substantially deviates from other studies (e.g. *Foken* (2006) notes an accepted value of $r = 6$), such a choice still appears to be coherent for the YSU PBL scheme taking its initial implementation into account.

4.5.3 Experimental setup

In the previous section, the configuration WRF2 has been identified as the most favorable. This configuration of physical parameterizations is retained, and five optimizations of the WRF2 configuration are chosen to rerun the forecast periods 1 – 31 August 2014 and 1 – 30 November 2014. The optimizations are summarized in Table 4.8. The roughness length of the land use categories *Croplands* and *Cropland/Natural Vegetation Mosaic*, *Mixed Forests* and *Urban/Build-Up* is multiplied by either a factor of $r = 2.0$ or $r = 3.0$. For the November period, summerly values for land use categories *Croplands*, *Cropland/Natural Vegetation Mosaic* are taken as the initial point to account for an inhomogeneous land use, which is independent of the season. The stability parameter s is varied between $s = 1$, which correspond to the default setting of the YSU PBL scheme, and $s = 0.5, 0.25, 0.1$ to enhance mixing in the stable regime, with smaller values indicating stronger mixing.

Table 4.9: Verification statistics for the day-ahead wind power forecast with model configuration WRF2 and optimizations. The TSO forecast and the derived forecast with GFS model data is shown for comparison. The forecast period is 1 – 31 August 2014. All values are normalized with the installed wind power capacity. Bold numbers indicate statistics of the final model setup.

	TSO	GFS	WRF2	r2s1	r3s1	r2s0.5	r3s0.25	r3s0.1
RMSE	0.031	0.040	0.093	0.059	0.045	0.053	0.039	0.041
BIAS	-0.001	0.005	0.068	0.036	0.017	0.028	0.006	0.005
CORR	0.969	0.967	0.954	0.945	0.944	0.951	0.950	0.950

Table 4.10: Verification statistics for the day-ahead wind power forecast with model configuration WRF2 and optimizations. The TSO forecast and the derived forecast with GFS model data is shown for comparison. The forecast period is 1 – 30 November 2014. All values are normalized with the installed wind power capacity. Bold numbers indicate statistics of the final model setup.

	TSO	GFS	WRF2	r2s1	r3s1	r2s0.5	r3s0.25	r3s0.1
RMSE	0.0291	0.048	0.147	0.078	0.063	0.074	0.059	0.061
BIAS	-0.001	0.011	0.112	0.045	0.024	0.040	0.016	0.016
CORR	0.964	0.948	0.867	0.880	0.877	0.882	0.883	0.883

4.5.4 Results

Simulations with tuning parameters according to Table 4.8 have been carried out in the same manner as in the previous section. Yet, the discussion is restricted to the day-ahead wind power forecast. Table 4.9 and 4.10 list the aggregated verification statistics for the periods 1 – 31 August 2014 and 1 – 30 November 2014, respectively. The largest improvement in root mean square error is due to an increase in roughness length z_0 , as the positive wind bias is diminished. For the August period, the wind bias is reduced from 0.068 to 0.036 for $r = 2$ (WRF2r2s1), and further reduced to 0.017 for $r = 3$ (WRF2r3s1). For the November period, the wind bias is reduced from 0.112 to 0.045 for $r = 2$ (WRF2r2s1), and further reduced to 0.024 for $r = 3$ (WRF2r3s1). From this point on, the enhancement of mixing for stable conditions in the PBL further reduces the wind bias (WRF2rs0.5, WRF3rs0.25 and WRF3rs0.1). The wind bias is reduced to a negligible extent for the August period, but still apparent for the November period. The increase in roughness length has a negligible effect on correlation for both periods. However, a decrease in stability parameter s consistently increases correlations, indicating that the temporal evolution of the wind bias is corrected. This is confirmed by Figure 4.8 and 4.9, which show the diurnal

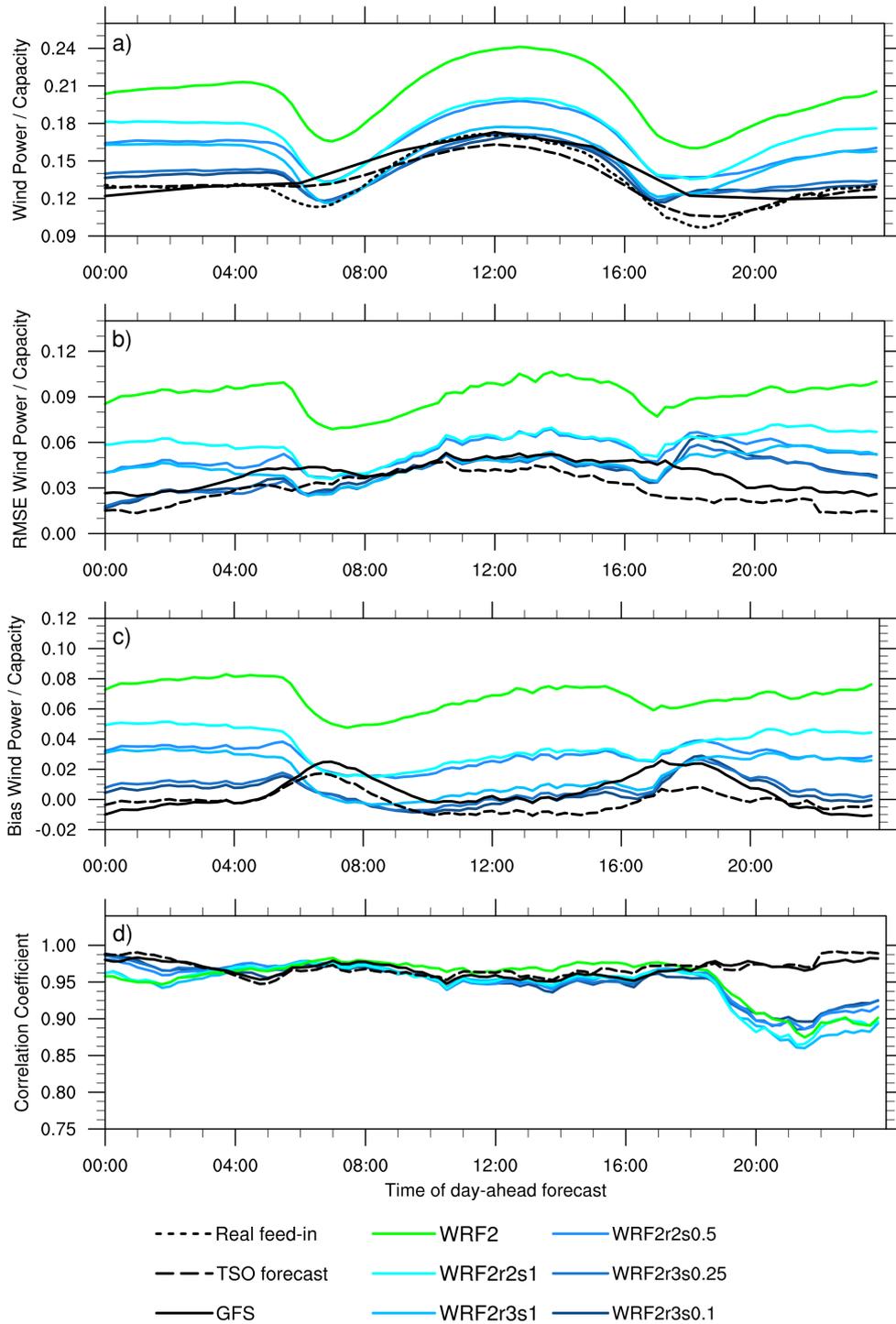


Figure 4.8: Diurnal cycle of verification statistics for the normalized day-ahead wind power forecast with optimized WRF configurations: (a) sample mean, (b) root mean square error, (c) bias and (d) correlation. The forecast period is 1 – 31 August 2014. The green line denotes the WRF2 configuration, blue lines of different shades the different optimizations. For comparison, the real power feed-in, the TSO forecast and the derived forecast with GFS model data and the IWES power model is shown.

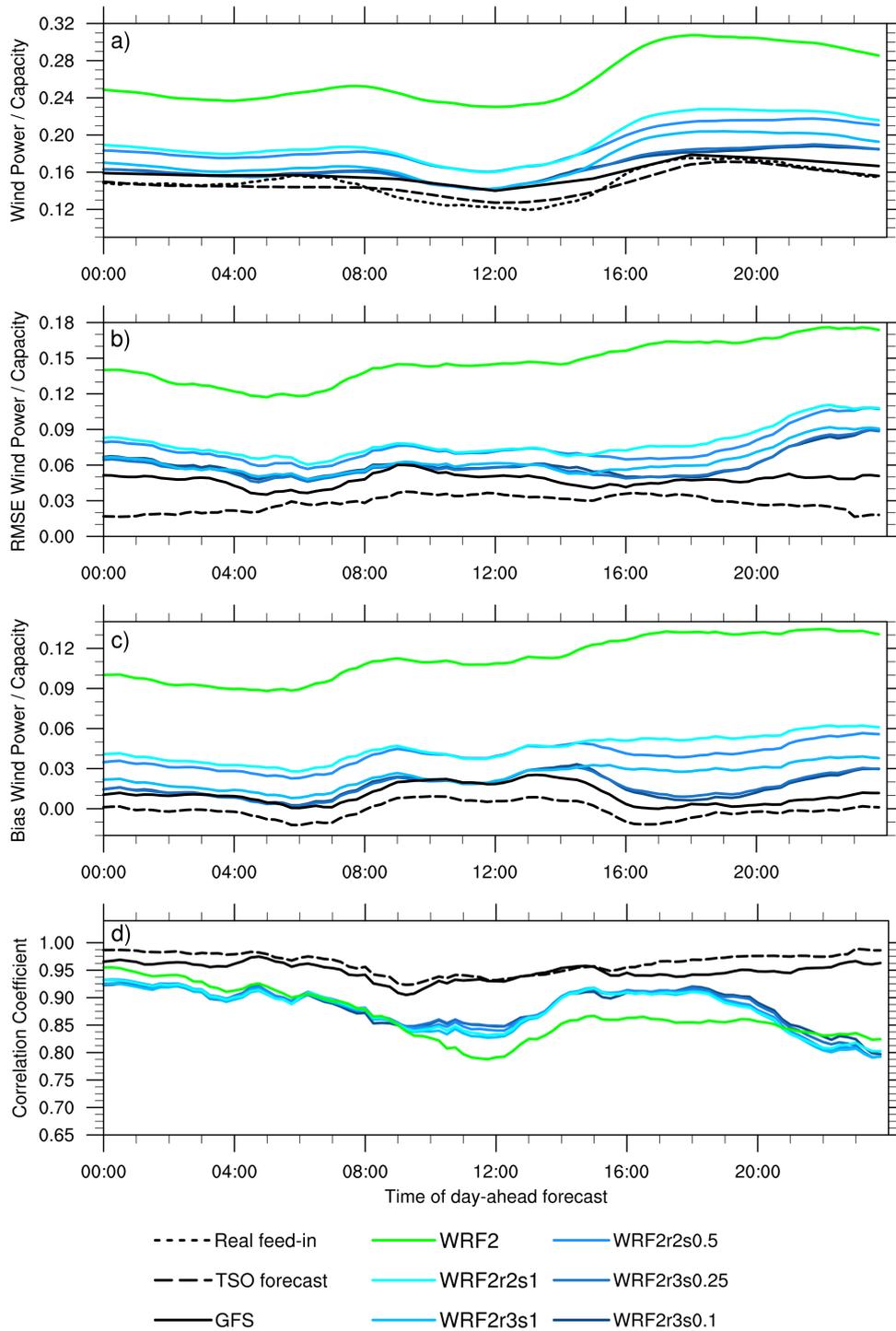


Figure 4.9: Diurnal cycle of verification statistics for the normalized day-ahead wind power forecast with optimized WRF configurations: (a) sample mean, (b) root mean square error, (c) bias and (d) correlation. The forecast period is 1 – 30 November 2014. The green line denotes the WRF2 configuration, blue lines of different shades the different optimizations. For comparison, the real power feed-in, the TSO forecast and the derived forecast with GFS model data and the IWES power model is shown.

cycle of the sample mean, root mean square error, bias and correlation, for both periods. Results for WRF3rs0.25 and WRF3rs0.1 are in good agreement with the mean diurnal cycle of the real feed-in for August (Figure 4.8a). It can be noted, that a variation of s has only an effect during nocturnal stable conditions, whereas wind power during day-time remains unaffected, as the boundary layer is well mixed.

Returning to Table 4.9 and 4.10, the root mean square error can be reduced to an extent comparable to the TSO forecast. For the August period, the lowest value can be noticed for WRF3rs0.25 with 0.039, slightly smaller than the root mean square error of the GFS model. For the November period, again, the lowest value can be noticed for WRF3rs0.25 with 0.059, noticeable larger than the root mean square error of the GFS model. A further decrease in the stability parameter s has a slightly negative effect on the root mean square error, and is therefore refused.

To conclude, the configuration WRF3rs0.25 has been chosen to serve as a deterministic model setup for the ultra large ensemble. By multiplying the roughness length z_0 by a factor of $r = 3$, and enhancing the the vertical mixing in the YSU PBL scheme with a stability parameter of $s = 0.25$, the forecast is of comparable skill to the forecast derived with the GFS model, and slightly worse than the postprocessed multi-model TSO forecast.

Chapter 5

Evaluation of an ultra large wind power ensemble

A predictability analysis of an exceptional error event in wind power forecasting is carried out with an ultra large meteorological ensemble of 1024 members. The experimental setup is summarized and the analysis to identify a major exceptional error event is described. The meteorological ensemble is evaluated in terms of ensemble dispersion of turbine hub height winds. Conversion to wind power is realized for the entire ensemble and evaluated in terms of higher order statistics. A case study from an exceptional error event in solar power forecasting supports the main findings.

5.1 Experimental setup

A 1024-member ensemble is set up with the ESIAS-met system. The domain configuration is described in detail in Section 4.1.1 and is depicted in Figure 4.1. The inner convection-permitting domain (nested domain) feeds back model values to the outer domain (parent domain) every single time step, known as two-way nesting. The WRF model configuration is chosen according to the results discussed in Chapter 4. Each ensemble member is initialized by the GEFS reforecast, with perturbed analysis fields distributed equally among the ensemble member. Initial condition perturbations of the GEFS reforecast have been generated by the Ensemble Transform method (*Wei et al.*, 2008), introduced in Section 2.2. Model uncertainty is represented by the SPPT (*Berner et al.*, 2011) and SKEB scheme (*Berner et al.*, 2015), introduced in Section 2.3.

Table 5.1 summarizes the stochastic perturbation parameters, while Figure 5.2a, 5.2b and 5.1 show instantaneous perturbation fields in physical space (the computation of perturbation patterns is realized in spectral space, see Section 2.3). The default settings of both schemes are varied only slightly, since simulation results from single sensitivity studies do not show a strong dependency on the parameter choice.

Table 5.1: Stochastic model parameter for SPPT and SKEB scheme. Subscripts ψ and θ denote parameter values for the streamfunction and potential temperature, respectively.

Parameter	Value
Temporal correlation length (SPPT)	$\tau = 21600 \text{ s}$
Spatial correlation length (SPPT)	$\kappa = 150 \text{ km}$
Grid point variance (SPPT)	$\eta = 0.125 - 0.25$
Temporal correlation length (SKEB)	$\tau_\psi = \tau_\theta = 10800 - 21600 \text{ s}$
Backscattered energy rates (SKEB)	$B_\psi = 10^{-5} \text{ m}^2/\text{s}^3, B_\theta = 10^{-6} \text{ m}^2/\text{s}^3$
Power law (SKEB)	$\beta_\psi = \gamma_\theta = -1.83$

This finding is in compliance with other studies using the WRF model at midlatitudes (*Berner et al., 2011; Romine et al., 2014; Berner et al., 2015; Duda et al., 2016; Jankov et al., 2017*). Despite different domain setups, all studies draw the conclusion, that the most favorable parameter choice remains close to the default values, indicating their universality. For the SPPT scheme, grid point variance κ is slightly reduced, as numerical instabilities have been observed on the nested domain. For the SKEB scheme, a power law of $\beta_\psi = \gamma_\theta = -1.83$ for the stochastic pattern results in a forcing with kinetic energy spectrum of $-5/3$ and a potential energy spectrum of $-10/3$. The steeper power law for the velocity leads to smaller spatial correlation lengths (see Figure 5.2a and 5.2b). Stochastic patterns are generated on the outer domain and interpolated onto the nested domain to preserve consistency. *Romine et al. (2014)* and *Jankov et al. (2017)* report on the SKEB scheme outperforming the SPPT scheme in terms of probabilistic skill scores. However, a combination of both schemes is favorable, as the ensemble dispersion in the planetary boundary layer is mainly induced by the SPPT scheme, and in the free atmosphere mainly by the SKEB scheme. Upon this well-established knowledge, 630 members are perturbed by the SKEB scheme and 384 members by the SPPT scheme. No a priori knowledge on the parameter is assumed, since parameter tuning to adjust the ensemble spread to the ensemble mean error for a single forecast period does not correspond to a realistic scenario. The first 11 ensemble members are dynamically downscaled without any model uncertainty representation.

5.2 Case study selection

The Fraunhofer IWES conducted an analysis to identify the days for which the TSO day-ahead forecast showed the most extreme errors. The TSO day-ahead forecast is a weighted mean of multiple NWP systems (*Good, 2017*). Hence, if the TSO day-ahead forecast is exceptionally erroneous, one can assume that at least the majority of involved NWP systems showed a poor forecast skill.

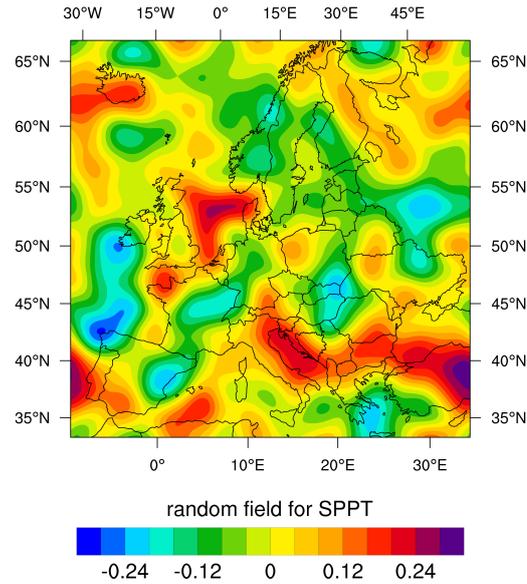


Figure 5.1: Instantaneous SPPT perturbation pattern on the outer domain.

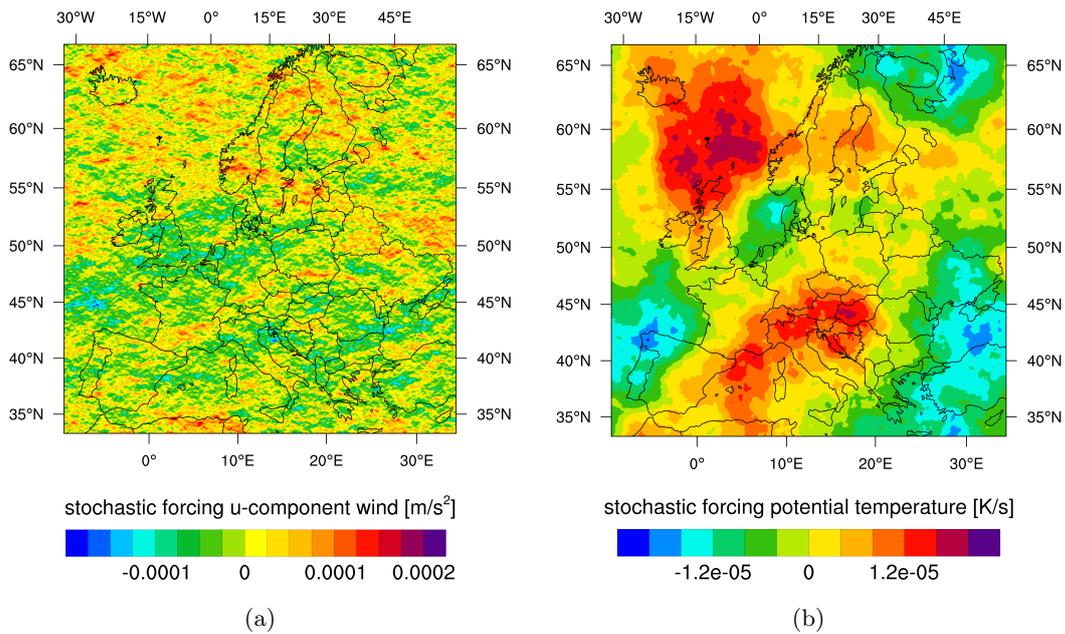


Figure 5.2: Instantaneous SKEB perturbation patterns on the outer domain for (a) u -component of wind vector and (b) potential temperature. The perturbation pattern for the v -component is generated by a rotation and phase shift of the u -component perturbation pattern.

The TSO day-ahead forecast has been compared to the real power feed-in (for convenience termed the true power supply) for the period 1 January 2014 – 31 August 2015. Both data sets are made publicly available by the TSOs. Moments of errors $\sqrt[n]{(F - O)^n}$ for each day-ahead forecast have been calculated, including data

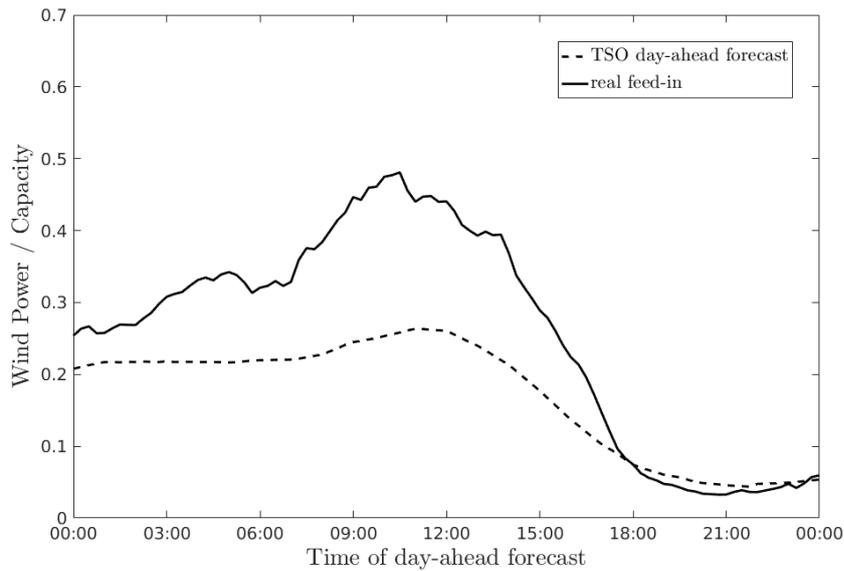


Figure 5.3: TSO day-ahead power forecast and real feed-in of wind power on 9 August 2014. Values are normalized by the installed wind power capacity (35 GW). Times are in UTC.

points of 15-minutes resolution, where F denotes the power forecast and O the real power feed-in. All values have been normalized by the root mean square error of the entire period. In principle, the higher the moment n of the error, the higher is the weight on more extreme error increments. Results show, that errors grow approximately linear in magnitude in the range of medium to moderate percentiles of the distribution, whereas the errors increase sharply above the 95% percentile. Hence, long tails are present in the distribution of errors, defining exceptional error events.

The day-ahead forecast of 9 August 2014 proved to be exceptionally poor. Statistically, it distinguishes itself from the remaining analysis period by multiple 6-hours intervals, which are arranged in the 99.8% percentiles of errors for the moments $n = 2, \dots, 6$. Therefore, the day-ahead forecast of 9 August 2014 did not only show single, extremely large forecast-observation increments, but also severe forecast errors throughout the day. Figure 5.3 shows the TSO day-ahead power forecast and the real power feed-in, where the wind power is normalized by the installed capacity. A distinct underestimation of wind power can be noticed at all lead times of relevance. The true power production increases sharply during the morning hours, while the TSO day-ahead forecast anticipates only a slight increase, with maximum error values around noon. With an installed wind power capacity of 35 GW at this time, forecast-observation increments increase from 1.6 GW at 00:00 UTC to 7.8 GW at 10:45 UTC. The root mean square error between 03:00 UTC and 15:00 UTC is thereby 5.4 GW. For comparison, typical root mean square errors are approximately

1.0 GW (see Table 4.9 and 4.10). The accumulated forecast error is 34% of the total production throughout the day. Due to the severity of the error event, it has been subject to further studies in the realm of energy meteorology, concerning its predictability with a cyclone detection system (*Steiner et al.*, 2017) or energy market effects (*Stark*, 2015).

5.3 Meteorological situation

From 8 – 9 August 2014, the meteorological condition over Central and Western Europe was dominated by two low pressure systems (see Figure 5.4a and 5.4b for the NCEP GFS analysis). At 500 hPa aloft, a stationary primary long-wave trough extended from Southern Greenland to Western Ireland and the Norwegian Sea. At 00 UTC 9 August, its low pressure system was located about 300 km northeast of Ireland, with a central pressure of 995 hPa. A secondary short-wave upper air trough developed on 8 August over the Eastern Atlantic, with its low pressure system just off the coast of Brittany at 00 UTC 8 August. This secondary cyclone moved rapidly around the primary system and approached Southern Britain by the end of 8 August. During that time, its central surface pressure dropped from 1100 hPa to 1000 hPa. Both systems were connected by an occlusion directed along Scotland at 00 UTC 9 August. Within the following 12 hours, the central pressure of the secondary cyclone dropped considerably further to 985 hPa, with the centre moving towards the North Sea. The upper air trough developed further downward, with strong frontogenesis aligned in North to South direction of Central Europe, propagating from the Benelux to Denmark and North-East Germany later in the day (see Figure 5.5a and 5.5b). Pressure gradients were strongly increasing from the cyclone centre towards south-east in the direction of Germany. 10-minute intervals of 10 m wind speeds reached maximum values of 21.1 m/s in Büsum, 15.7 m/s on the island of Sylt and 12.9 m/s in Berlin-Tegel (*German Weather Service*, 2017).

5.4 Results

5.4.1 Meteorological evaluation

The 1024-member ensemble is initialized at 00 UTC 8 August 2014, where forecast lead times between 00 UTC 9 August – 00 UTC 10 August correspond to the day-ahead forecast. The following discussion is restricted to the predictability analysis of the secondary cyclone. Figure 5.6a shows the difference in mean sea level pressure between the control member forecast and the NCEP GFS analysis at 12 UTC 9 August. This time is included in the hours of extreme errors in the day-ahead forecast (see Figure 5.3). The control member denotes the WRF model run initialized by the unperturbed analysis and a control model formulation (without model uncertainty

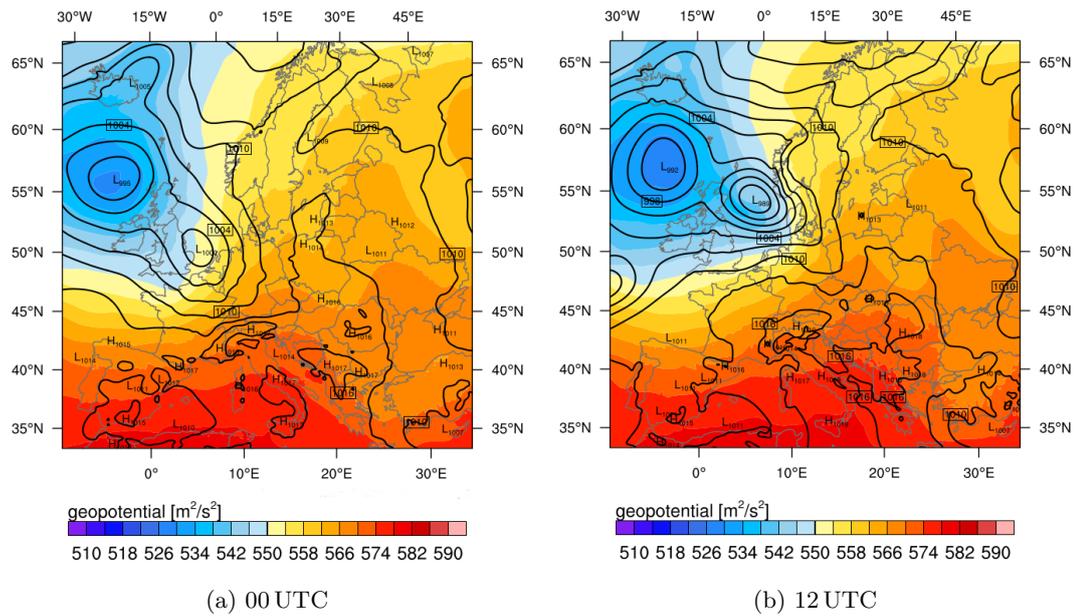


Figure 5.4: NCEP GFS analysis at (a) 00 UTC 9 August 2014 and (b) 12 UTC 9 August 2014. Geopotential height is shown at 500 hPa and surface pressure contours are shown each 4 hPa.

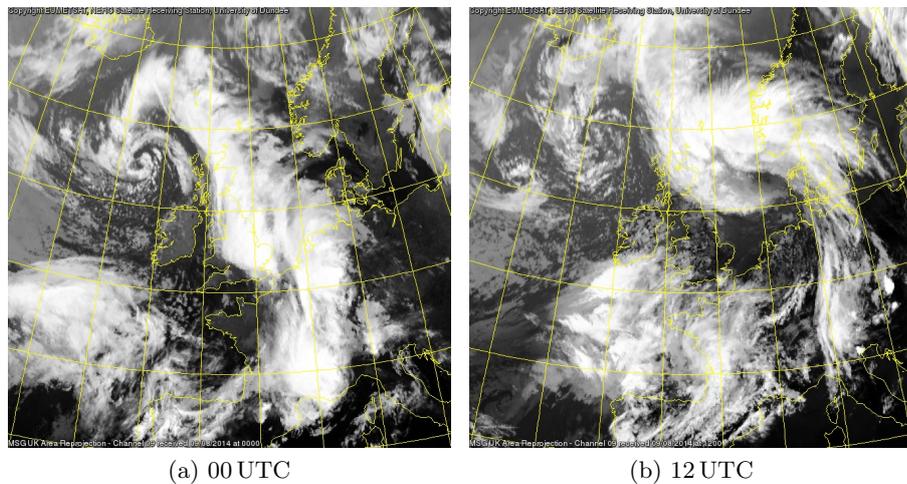


Figure 5.5: Meteosat Second Generation SEVIRI satellite imagery, channel 9, $10.8 \mu\text{m}$, UK projection: (a) 00 UTC 9 August 2014, (b) 12 UTC 9 August 2014. Source: <http://www.sat.dundee.ac.uk/>.

representation). The position of the central pressure is represented fairly well by the control member, between Northern Scotland and Southern Norway, only slightly shifted to the north compared to the GFS analysis. However, the central pressure is significantly overestimated by 4 hPa. Further, the control member overpredicts the local decrease in pressure gradients around the centre, resulting in a strong under-

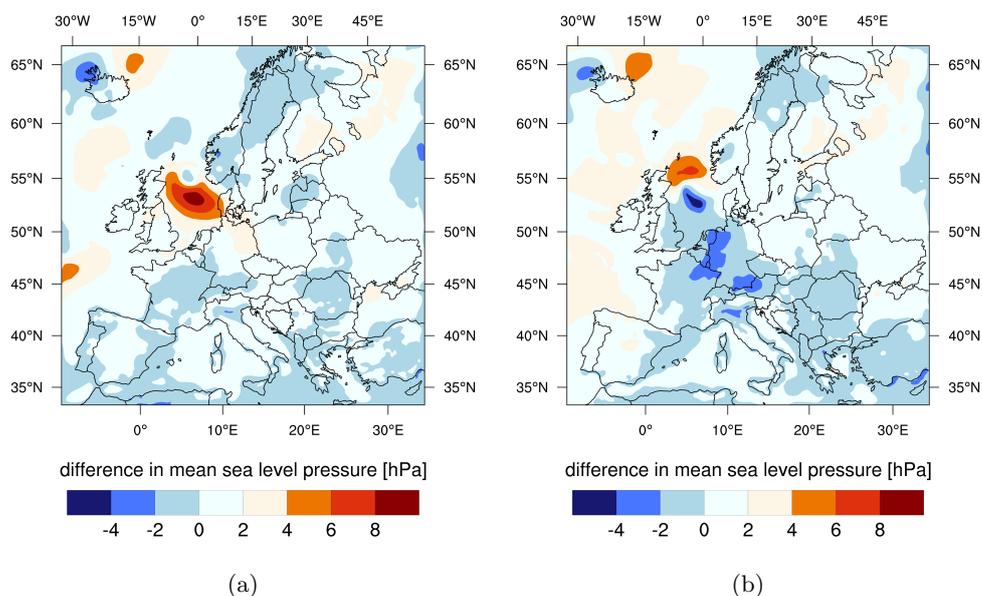


Figure 5.6: Difference in mean sea level pressure between forecast and GFS NCEP analysis at 12 UTC 9 August 2014, where in (a) the forecast is the control member and in (b) the dynamically downscaled fourth GEFS member without model perturbation.

estimation of pressure gradients further south-east. The combination of both results in forecast-minus-analysis increments exceeding 8 hPa at the latitudes of Denmark. Noticeably, an underestimation of pressure gradients extends further south-east to coastal areas of Germany and regions of frontal movement (see Figure 5.5b for comparison). Results shown in Figure 5.6a are well in agreement with results obtained by the COSMO-EU model, depicted in Figure 3b of *Steiner et al. (2017)*. Hence, the conclusion can be drawn, that the low predictability of the surface low shape and central pressure represents a certain universality for today's NWP systems.

Figure 5.6b shows the forecast-minus-analysis increments in mean sea level pressure for the ensemble member initialized by the fourth GEFS ensemble member (without model uncertainty representation). A distinct dipole pattern can be noticed in the area of the secondary cyclone, indicating that the low pressure system is located considerably further south compared to the GFS analysis. The central pressure is predicted fairly well, since the absolute values of minima and maxima increments are approximately of the same size. However, the dipole structure is not symmetric, indicating a surface low of different shape. Consequently, the evaluation of both examples depicted in Figure 5.6 suggests that the forecast errors are strongly sensitive to the location, shape, central pressure and local pressure gradients of the surface low.

The ensemble spread of 100 m wind speeds is displayed in Figure 5.7 at 06 UTC 9 August and 12 UTC 9 August for the complete 1024-member ensemble. Local maxima of ensemble spread are well in compliance with the locations of the primary and

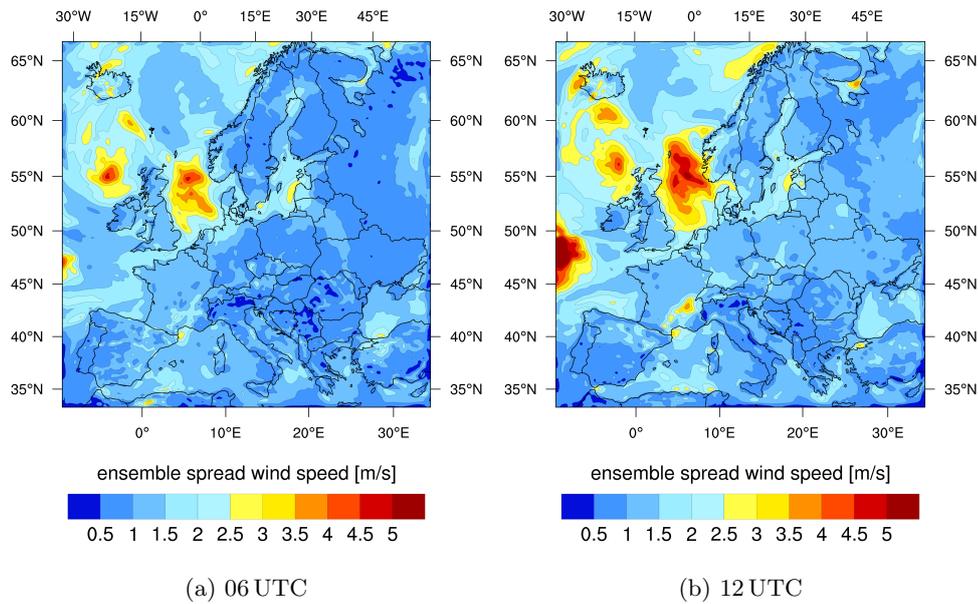


Figure 5.7: Ensemble spread of 100 m wind speed for the 1024-member ensemble: (a) 06 UTC 9 August 2014, (b) 12 UTC 9 August 2014.

secondary cyclones (see Figure 5.4b), and therefore with regions of low predictability. Within the short time period of 6 hours, a rapid increase in ensemble spread over the centre of the secondary cyclone can be observed. The increase is not only confined to the magnitude, but even more prominent in the spatial extent, representing the uncertainty in the low pressure system's location and shape. There is no pronounced ensemble spread visible in regions of frontogenesis, indicating an insufficient uncertainty representation by the ensemble spread. However, the frontal system is distinguishable in the ensemble spread of the wind components (not shown here), corresponding to uncertainty in the frontal orientation. In Figure 5.7b, a large ensemble spread between 45° and 50° latitude at the western domain boundary can be noticed. The ensemble dispersion results from the approaching extratropical remnant of hurricane Bertha, which is not of any relevance to the discussion here.

The spaghetti plot of the first 11 ensemble members at 12 UTC 9 August 2014 is depicted in Figure 5.8a, where each ensemble member is initialized by different initial conditions from the GEFS ensemble, yet without model uncertainty representation. Therein, isohypses are displayed at 850 hPa, as the initial upper air trough had already developed further downward at this time. The isohypse of the GFS analysis is shown for comparison. The green colored line corresponds to the ensemble member initialized by the fourth GEFS ensemble member, which is characterized by a slight displacement of the trough to the south, in compliance with Figure 5.6b. A distinct divergence of isohypses can be identified over the region of the secondary cyclone. The GFS analysis is enclosed in the ensemble realizations, however, all 11 member

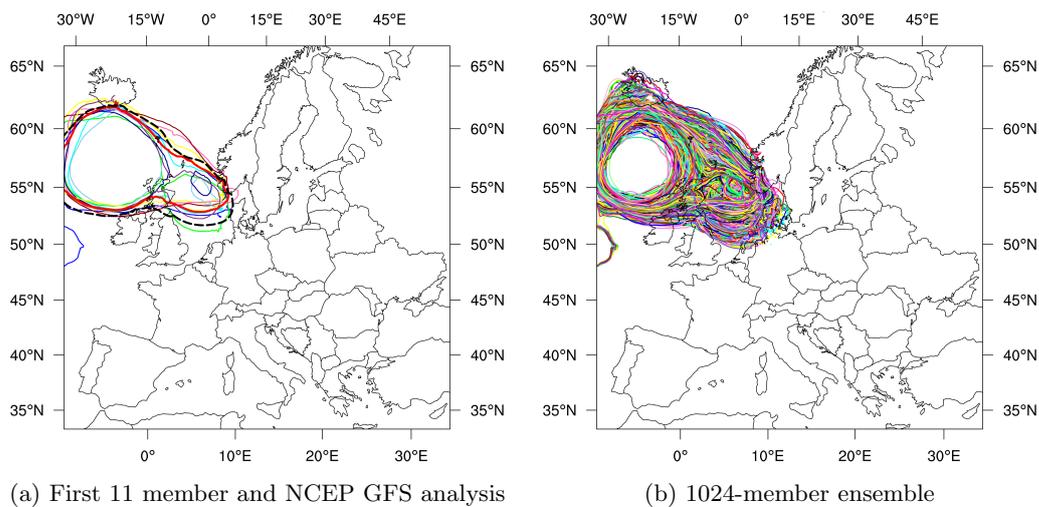


Figure 5.8: Spaghetti plot of geopotential height at 850 hPa, isohypses of 134 gpdam, at 12 UTC 9 August 2014. (a): 11 dynamically downscaled GEFS members (no model perturbation). The red line corresponds to the control member, the black dashed line to the NCEP GFS analysis. (b) The complete 1024-member ensemble.

exhibit a more narrow shape of the secondary trough. This fact is identified as the main reason of forecast failure. The ensemble's low pressure systems are all of smaller spatial extent. Therefore, smaller pressure gradients appear over Germany, resulting in weaker frontogenesis and hence, smaller wind speeds were induced. Figure 5.8b shows the spaghetti plot of the complete 1024-member ensemble. An even more distinct picture can be drawn in terms of forecast uncertainty by an increased isohypse dispersion. Model uncertainty schemes (SPPT or SKEB) eventually trigger instabilities in the flow, leading to pronounced differences in the model results, as numerous ensemble runs estimate the trough's location far off.

In wind and solar power forecasting, it is standard to interpret probabilistic forecasts by quantiles of the ensemble's distribution (*Möhrlen and Jørgensen, 2017*). Forecast intervals, known as quantiles, allow a comprehensible assessment of the cumulative distribution function and indicate a range of possible outcomes. Quantiles are frequently confused with confidence intervals, which aim to compute statistical metrics of a distribution and indicate how well the sample distribution reflects this metric. The latter is however not of primary interest for users of power forecasts. Hence, quantiles are considered here in accordance to the standard.

Nested quantiles of hub height wind speeds at the locations of measurement towers are shown in Figure 5.9, together with the correspondent observation. One can clearly identify the approach of the surface low towards Germany on 8 August, starting at the FINO1 tower at 10 UTC with a sharp increase in observed wind speed. During 9 August, the frontal system's progression from Western to Eastern Germany is visible

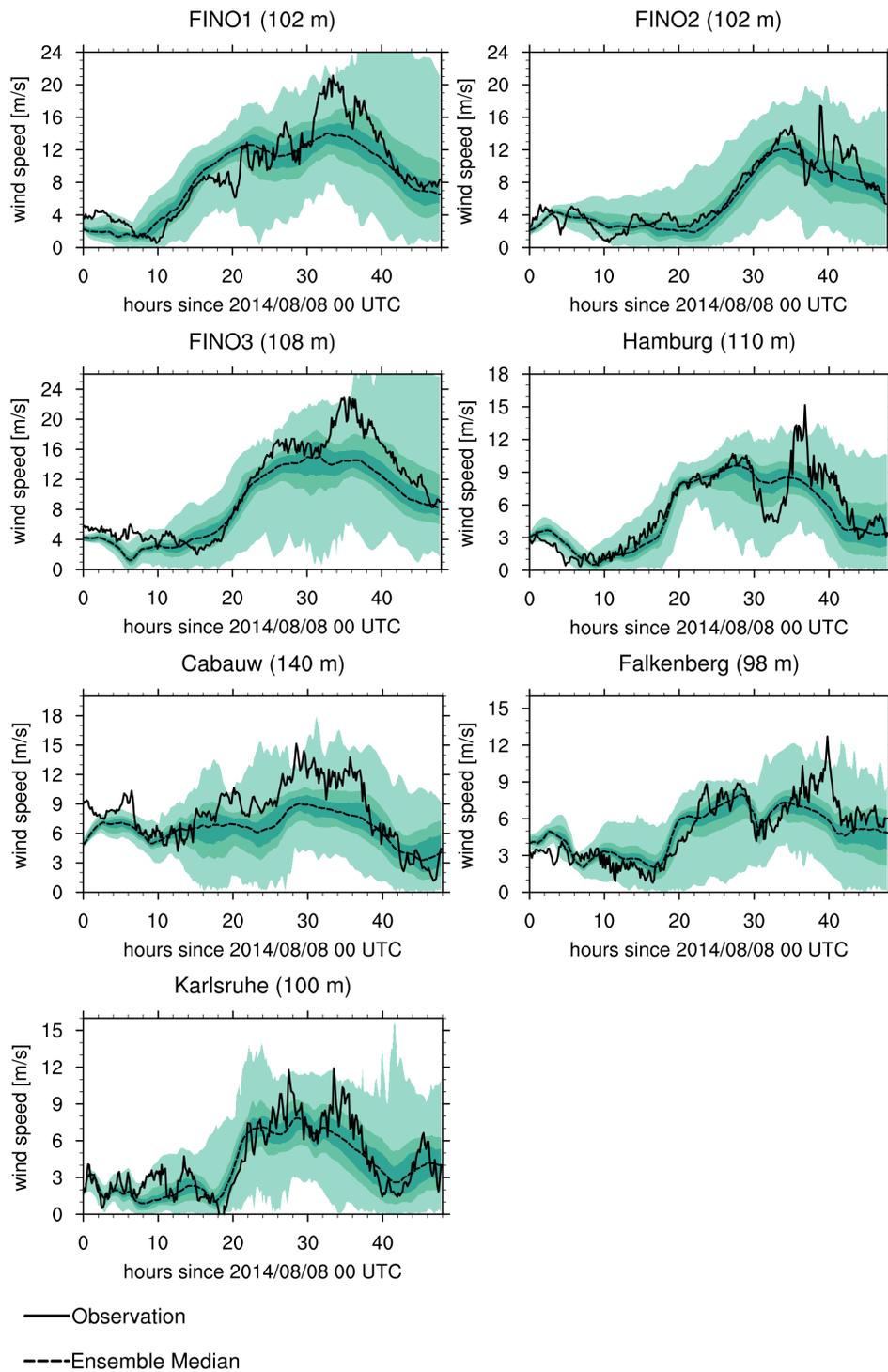


Figure 5.9: Nested quantiles of hub height wind speeds at measurement towers. The boundaries of shaded areas mark the minimum, 10 % percentile, 25 % percentile, 75 % percentile, 90 % percentile and the maximum in ascending order. The dashed line denotes the ensemble median and the solid line the observation.

in the temporal displacement of observed wind speed maxima among the measurement towers, which is also in compliance with the evolution of the error event (see Figure 5.3). The ensemble median indicates a clear underestimation of wind speeds during these times. Wind maxima of the ramping events are enclosed in the ensemble, except for Hamburg and Falkenberg, where a slight underestimation is identified. The distribution is clearly skewed towards overestimated wind speeds at FINO1 and FINO3 by the end of 9 August, mainly caused by ensemble members which share the initial conditions of the fourth GEFS ensemble member (see discussion above). Percentiles of 10 % and 90 % are closely grouped together in case of all measurement towers, with only a slight increase in divergence during times of exceptional errors. The ensemble spread indicates therefore an increased likelihood of forecast failure, but underestimates the values of extreme errors. In contrast, minimum and maximum values of wind speeds sharply increase with the onset of the exceptional error event. Hence, the conclusion can be drawn that the exceptional error event is solely indicated by long tails in the distribution of wind speeds due to single outliers. However, it should be noted, that the utilization of other ensemble generation techniques may come to a different conclusion.

5.4.2 Wind power evaluation

The ultra large meteorological ensemble is transformed into a wind power generation ensemble applying the physical wind power model utilized at Fraunhofer IWES. The wind power model is described in detail in Section 4.3. The evaluation in terms of wind power has been subject of *Good and Berndt (2017)* and is reinvestigated here.

In the following analysis, all values are normalized by the total installed wind power capacity of onshore German wind farms. The day-ahead forecast of the ultra large wind power ensemble for 9 August 2014 is shown in Figure 5.10, together with the ensemble mean, the TSO day-ahead forecast and the real power feed-in. A remnant of WRF's nocturnal wind bias can be observed in the ensemble mean between 00 UTC and 06 UTC. However, this bias is not carried throughout the day. The sharp decrease in wind power around sunrise is identified as a characteristic feature of stability transition in the PBL, followed by a reduction in wind bias (see Section 4.4). Hence, the forecast skill of the ensemble mean is of similar skill compared to the TSO day-ahead forecast, at least during times of most extreme errors.

The real power feed-in is enclosed by the ensemble at all forecast times. However, with respect to the size of the ensemble, only very few outliers capture the true power production during times of maximum forecast errors. Noticeably, there are more outliers generated in the direction of forecast error, indicating a proper skewness of the ensemble's PDF towards the true power production. This is further illustrated in Figure 5.11, which shows the nested quantiles of the ensemble. The temporal evolu-

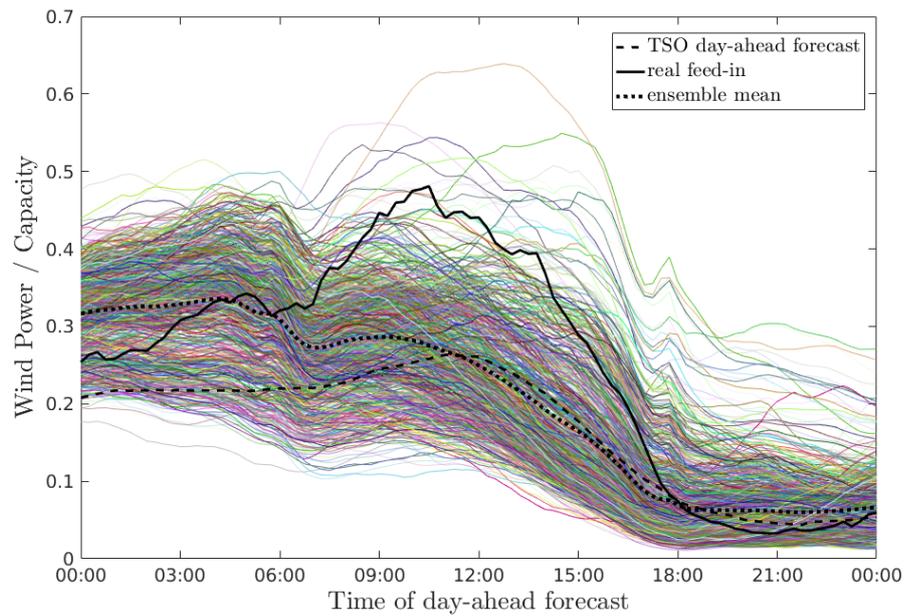


Figure 5.10: Day-ahead forecast of the ultra large wind power ensemble for 9 August 2014, including the ensemble mean, TSO day-ahead forecast and the real power feed-in.

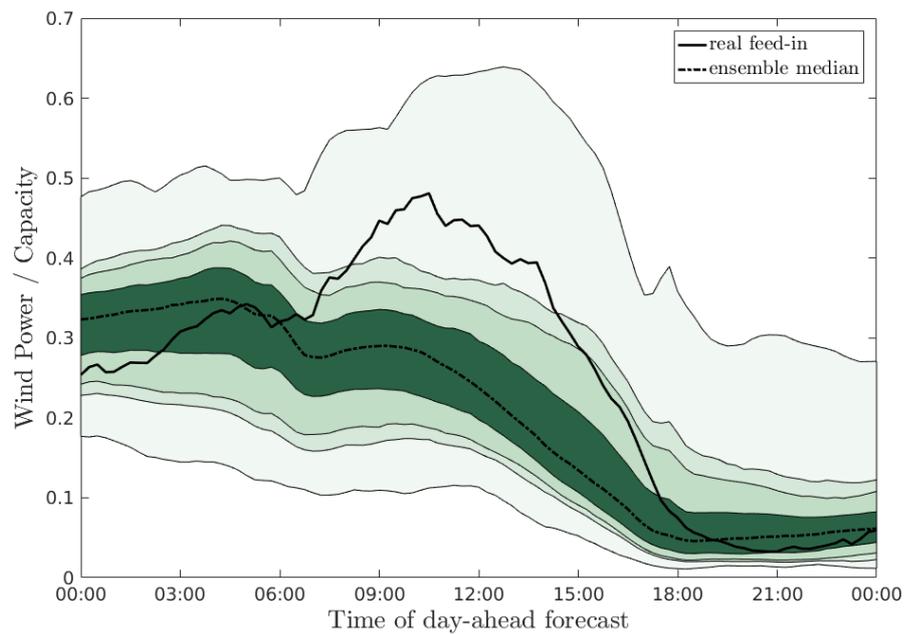


Figure 5.11: Nested quantiles of wind power. The boundaries of shaded areas mark the minimum, 5 % percentile, 10 % percentile, 25 % percentile, 75 % percentile, 90 % percentile, 95 % percentile and the maximum in ascending order. Additionally, the ensemble median and real power feed-in are displayed.

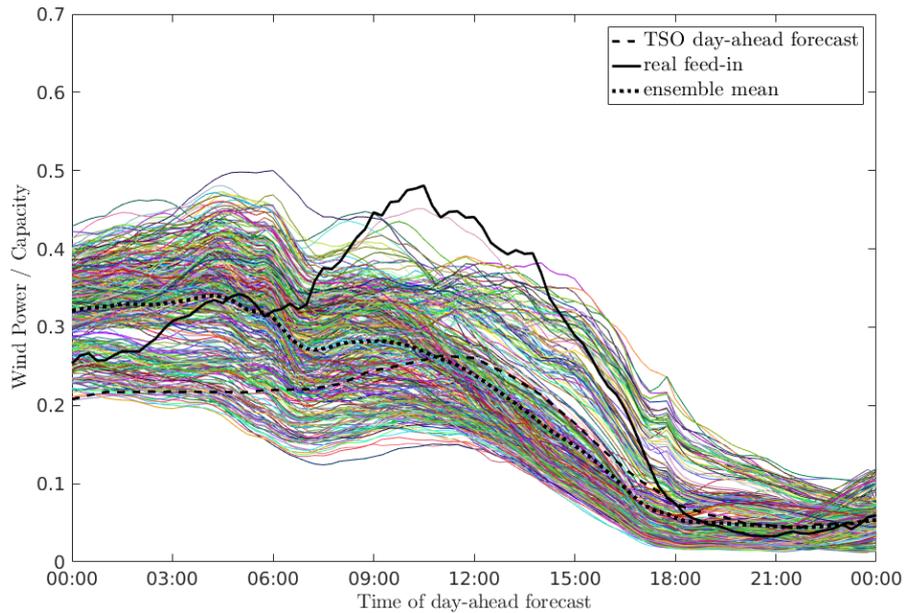


Figure 5.12: Same as Figure 5.10, but solely including ensemble members which share the SPPT scheme.

tion of the minimum and maximum values are well in qualitative agreement with the true power production. There is a large difference between the 95 % percentile and the ensemble maximum, indicating the long tails of the distribution into the direction of forecast error. In contrast, the 95 % percentile underestimates the true power production during most relevant times of the exceptional error event. In particular, during the times of maximum errors, only the 99.5 % percentile captures the true power production. In the previous section it is already noticed, that only a few single outliers out of the 1024-member ensemble indicate the exceptional error event. This conclusion is transferable to the wind power in a more pronounced manner. Thereby, an estimated size of 250 ensemble members appears to be sufficient to enclose the true power production at all forecast times.

Figure 5.12 shows a subset of the 1024-member ensemble by selecting solely the 384 members, which share the SPPT scheme for model uncertainty representation. One can clearly notice a decrease in outliers which are consistent with the true power production. This decrease is disproportional to the reduction in ensemble size, pointing to a general characteristic of model uncertainty schemes. As discussed in Section 2.3, different schemes address different shortcomings in NWP modeling, and therefore may induce fairly varying forcings, depending on the atmospheric conditions. It is well known, that the SPPT scheme is more effective in inducing ensemble dispersion in the PBL, while forcings induced by the SKEB scheme are stronger in the free atmosphere and during strong cyclogenesis (*Berner et al., 2011*). The latter is of

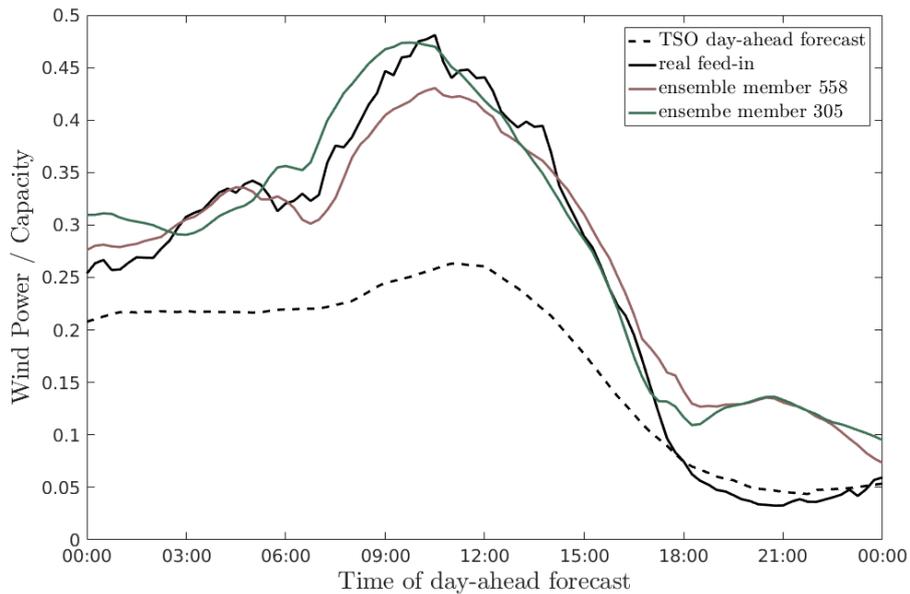


Figure 5.13: Day-ahead wind power forecast of ensemble member 305 and 558, which show the lowest root mean square error for 9 August 2014.

relevance here, making the SPPT scheme less effective. Instabilities of the underlying flow are more likely triggered by the SKEB scheme and evolve eventually more consistent with the true atmospheric state. These results are supported by *Berner et al.* (2015), who show that "the merits of model-error representation go beyond increasing spread and removing the mean error and can account for certain aspects of structural model uncertainty".

A spatially aggregated value of wind power allows for a convenient determination of the ensemble member which is most in compliance with the true weather situation in Germany during this time. Figure 5.13 shows the 556th and 305th ensemble member, distinguished by the lowest root mean square error over the whole day, with values of 0.0457 and 0.0459, respectively. The root mean square error is thereby in the order of typical average values (see Table 4.9 and 4.10). Yet, a different synoptic situation has developed later in the day in case of both members, overestimating the calm winds in the evening. Both members have been initialized by the second GEFS member and share the SKEB scheme for model uncertainty representation.

It has been hypothesized before, that the exceptional error event can only be anticipated by single outliers in the ensemble, such that the tails of the distribution are of primary interest. Central sample moments $m_n = E((x - \mu)^n)$ are therefore calculated, where $E(x)$ is the expected value of x and μ the ensemble mean. The ensemble spread equals $\sqrt{m_2}$, odd moments measure the skewness of the PDF and even moments the kurtosis, with higher moments increasing the weight of the PDF's

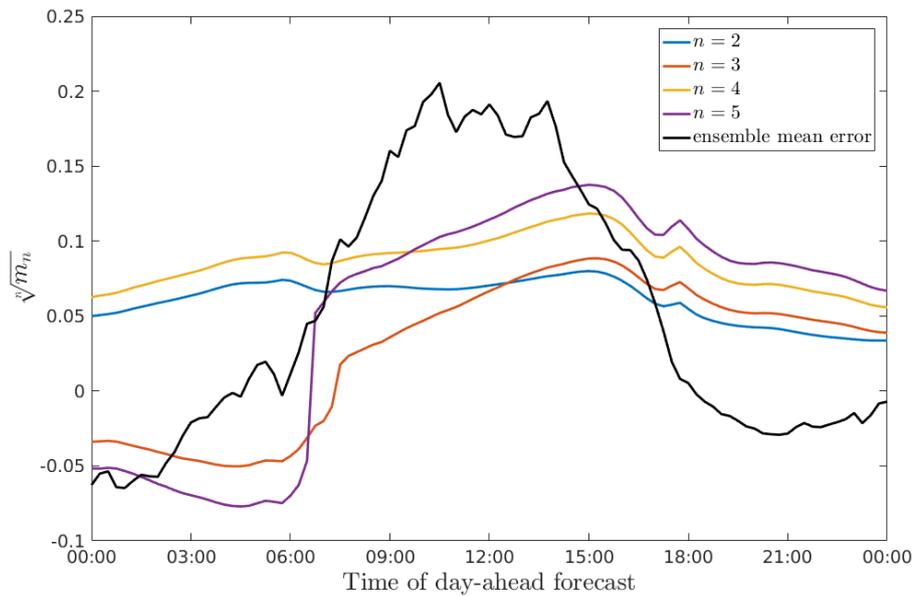


Figure 5.14: Root moments of wind power distribution from $n = 2$ to $n = 5$ for the wind power ensemble depicted in Figure 5.10. The ensemble mean error is shown for comparison.

tails. Moments are normalized to the same unit by the n th root and shown in Figure 5.14, together with the ensemble mean error for $n = 2, \dots, 5$. It is a common procedure in meteorological ensemble forecasting to compare the average ensemble spread to the root mean square error of the ensemble mean (*Grimit and Mass, 2007*), where a perfectly reliable ensemble shows a strict agreement of both. This comparison fails here due to the low predictability of the event. With increasing ensemble size, the majority of members arranges around the ensemble mean. Outliers, which are consistent with the true power production, are not well enough represented in the ensemble spread. Turning to the higher moments, a sharp increase of odd moments is most noticeable. This increase appears from negative to positive values at around 6 UTC, and is consistent with the frontal system's approach as well as the onset of most extreme errors. From 12 UTC on, odd moments are approximately of the same order as even moment, such that the even moments result mainly from the tail pointing into the direction of forecast error.

As mentioned above, the direction of skewness corresponds to the true direction of the forecast error. In this sense, Figure 5.14 reveals further information about the temporal evolution of the skewness, estimating the correct time when a different weather situation evolves, which distinguishes itself from the ensemble mean. Values of kurtosis ($n = 4$) and superkurtosis ($n = 6$) are many times the Gaussian values throughout the event. Thus, the ensemble shows a high probability of extreme error, compared to a normally distributed ensemble with the same ensemble spread. To

conclude, upon a skillfully estimated skewness and kurtosis, one can properly indicate a potential weather situation represented by the outliers, different from the situation characterized by the ensemble mean. Furthermore, the likelihood of the weather situation to result in an extreme error event can be assessed.

It remains to be shown, how many ensemble members are needed to resolve the ensemble statistics, as displayed in Figure 5.14. This is realized by randomly resampling the ultra large ensemble with replacement by multiple samples of smaller sizes. One may call this bootstrapping. However, in the classical sense bootstrapping aims to compute statistics of a single sampled distribution. Here, the aim is to compute statistics of subsets of a distribution and measure their variability. The subsets represent hypothetically reduced ensemble sizes. The complete 1024-member ensemble thereby replaces an infinitely large ensemble and a statistical moment is treated as a forecast variable. Hence, it is proceeded in the same manner as with the actual wind power forecast and quantiles are used to interpret the probabilistic forecast. For this purpose, subsets of sizes $N \in \{16, 64, 128, 256, 1024\}$ are formed. For each subset, 5000 samples are drawn from the ultra large ensemble, and for each sample, the second and third root moment are computed. Figure 5.15 and 5.16 show nested 5% and 95% percentiles of the resulting distributions for varying ensemble sizes. Nested percentiles of ensemble spread are arranged symmetrically around the true ensemble spread. The qualitative evolution with time is sufficiently resolved by all ensemble sizes. To accurately represent the ensemble spread of the 1024-member ensemble, 64 to 128 members are needed. Proceeding with the skewness, a larger ensemble size is necessary to make an accurate predication. Starting with 16 ensemble members, it is not obvious until 14 UTC, whether the distribution is skewed in the positive or negative direction. At this time, the most extreme errors of that day have already occurred. The same conclusion can be drawn for ensemble sizes of 64 and 128 members. At the time when the 5% percentile shows a sharp increase in skewness, parts of the extreme error events have already occurred. To conclude, at least 256 ensemble members are needed to resolve the skewness sufficiently well.

5.5 Outlook: Evaluation of an ultra large solar power ensemble

In the previous section, the kurtosis and in particular the skewness of the ensemble's distribution are identified as key indicators that contain information about the weather condition that might evolve differently from the bulk ensemble and lead potentially to an exceptional error event in wind power forecasting. This concept is taken further to a case study of solar power forecasting. The case study selection is conducted in the way described in Section 5.2, such that the day-ahead forecast

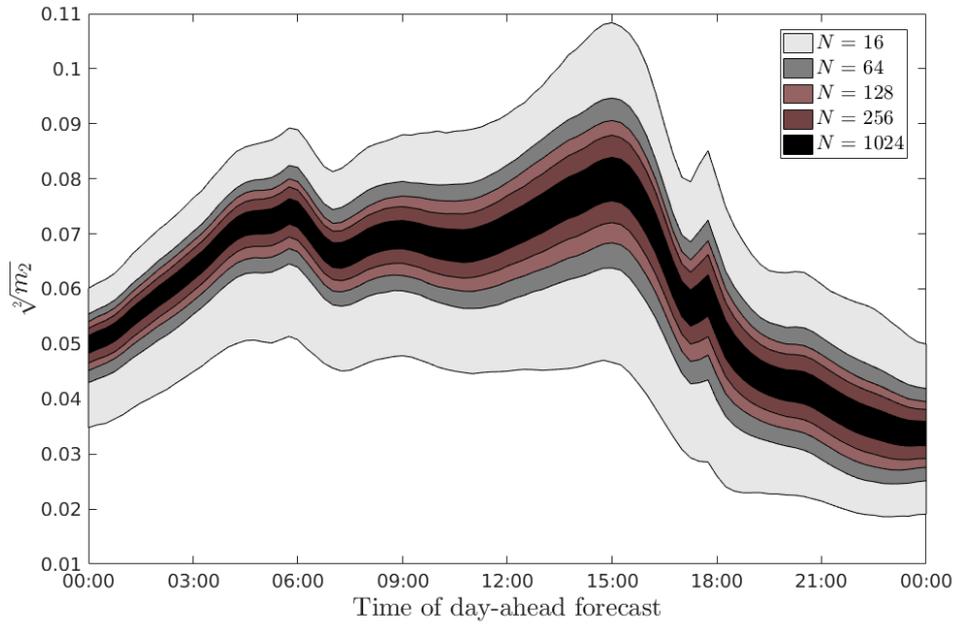


Figure 5.15: Nested 5% and 95% percentiles of the second root moment (ensemble spread) for bootstrapped samples of different ensemble sizes N . Wind power case study of 9 August 2014.

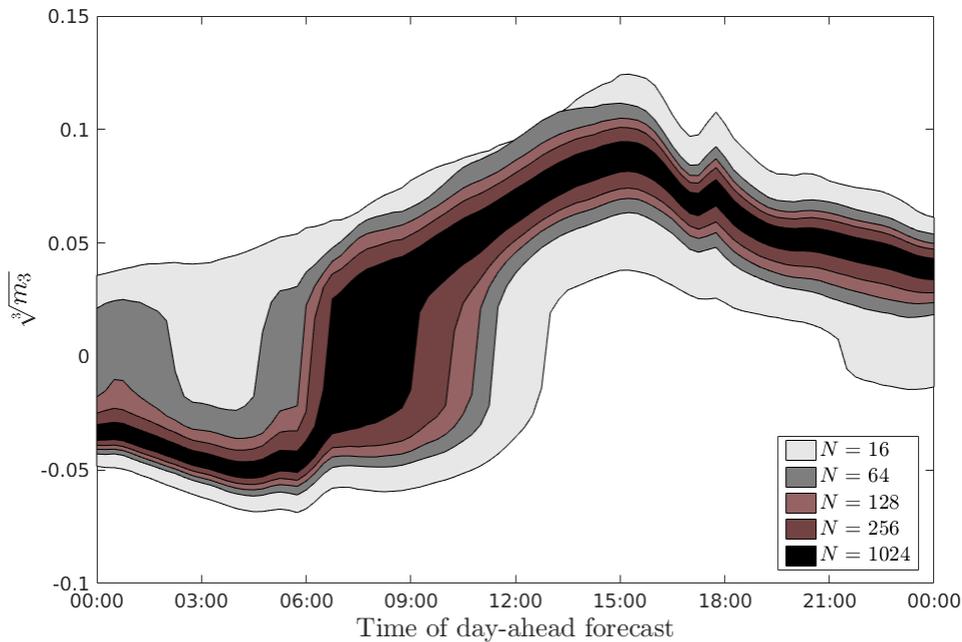


Figure 5.16: Nested 5% and 95% percentiles of the third root moment (skewness) for bootstrapped samples of different ensemble sizes N . Wind power case study of 9 August 2014.

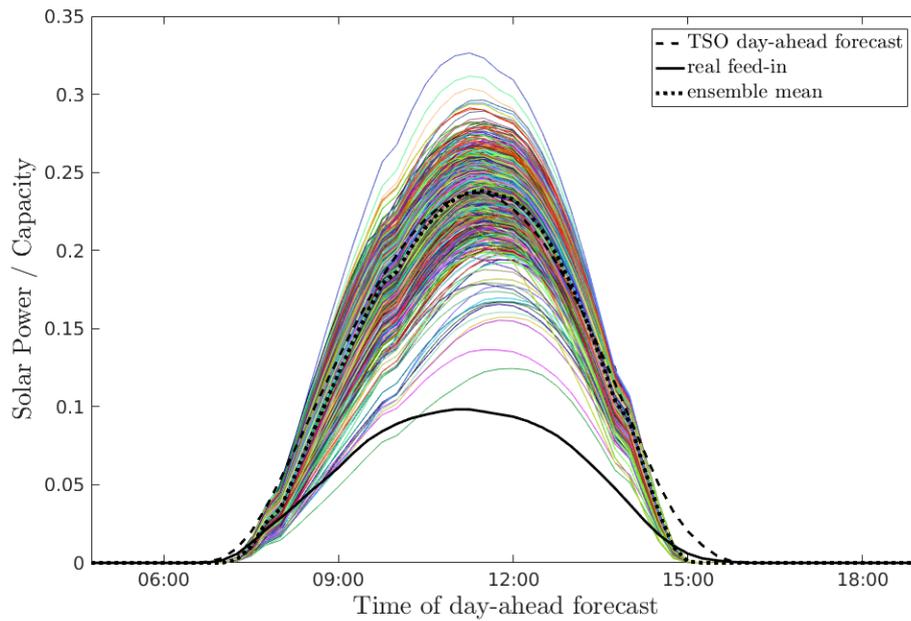


Figure 5.17: Day-ahead forecast of the ultra large solar power ensemble for 28 November 2014, including the ensemble mean, TSO day-ahead forecast and the real power feed-in.

of the 28 November 2014 is chosen. Results have been subject of *Good and Berndt* (2017) and are summarized in the following.

Between 28 – 29 November 2014, the weather situation in Germany was dominated by an anticyclonic southeasterly wind originating from a stationary high pressure system located over the Baltic States. Surface winds appeared to be easterly and advected cold air, while upper air winds at 300 hPa and 500 hPa were predominantly westerly with advection of warm air. Due to adiabatic heating of subsiding air masses, a distinct inversion layer evolved at approximately 900 hPa most notably over Western and Southern Germany. Hence, while at 14 UTC 28 November, a temperature of 6.0°C has been measured at the Zugspitze (2964 m) and 13.9°C at the Feldberg in the Black Forest (1490 m), temperatures appeared considerably lower in Freudenstadt (797 m, also located in the Black Forest) with 2.4°C , or Bamberg (240 m) with 3.5°C (*German Weather Service*, 2017). The temperature inversion led to low stratus clouds throughout the day in Western and Southern Germany.

An ultra large meteorological ensemble with 1024 members is set up in the exact same way as for the wind power case study. The WRF-Solar extension is utilized to estimate the global horizontal irradiance, direct normal irradiance and diffusive irradiance (*Jimenez et al.*, 2016). The conversion to solar power is realized with the Fraunhofer IWES regional Photovoltaic (PV) power model (*Saint-Drenan et al.*, 2017). Meteorological input fields are the components of solar radiation and 2 m

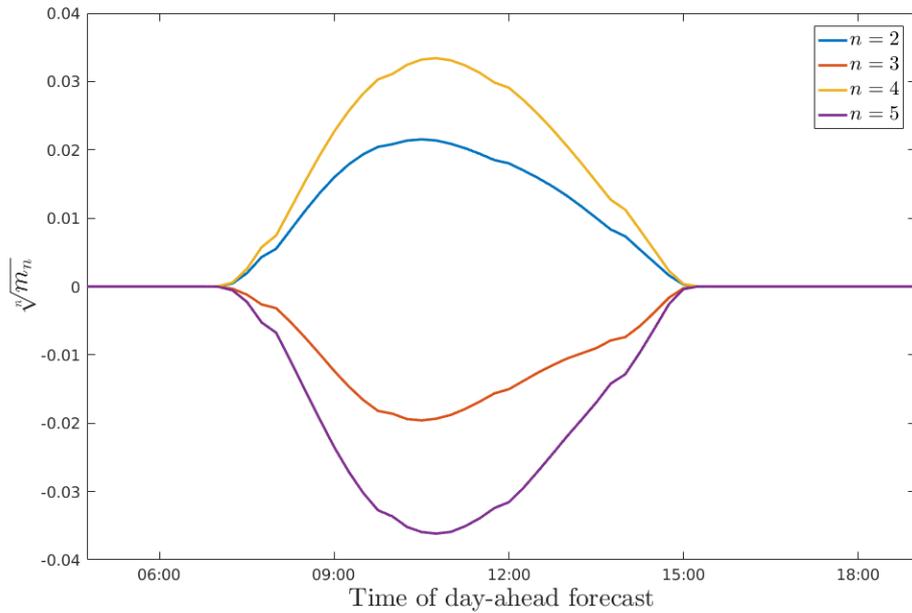


Figure 5.18: Root moments of solar power distribution from $n = 2$ to $n = 5$ for the solar power ensemble depicted in Figure 5.17.

temperature, as temperature decreases the power production due to module heating. The PV power model is probabilistic in the sense, that the power output of various possible orientations of PV modules is computed and trained on historical data.

Figure 5.17 shows the ultra large ensemble in comparison to the TSO day-ahead forecast and the real feed-in of solar power. The ensemble mean is in good agreement with the TSO day-ahead forecast. With an installed capacity of 38 GW, the maximum forecast error is 5.25 GW at 11:00 UTC. An overestimation of solar power generation is forecasted throughout the day, and even the ultra large ensemble does not enclose the true power production for most of the times.

The root moments of the distribution are displayed in Figure 5.18 for $n = 2, 3, 4, 5$. Skewness ($n = 3$) and superskewness ($n = 5$) increase rapidly in the direction of the forecast error, while values of kurtosis and superkurtosis are multiples of the Gaussian values. To assess the ensemble size, which is needed to resolve the proper direction of skewness, it is proceeded as in the previous section. Nested 5% and 95% percentiles of skewness are displayed in Figure 5.19 for different sample sizes. Up to a sample size of 1024 members, it cannot be determined with certainty for all forecast times, whether the ensemble is skewed into the direction of the forecast error. Hence, only very few outliers are responsible for the measure of the proper skewness, which cannot be represented sufficiently well by sample sizes smaller than 1024 member. This result is even more striking compared to the wind power case study, due to the fact

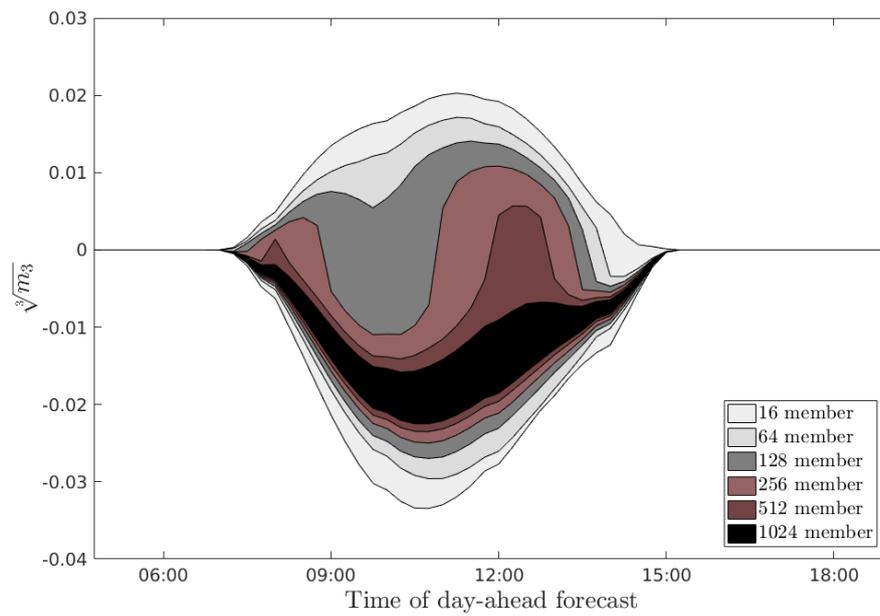


Figure 5.19: Nested 5% and 95% percentiles of the third root moment (skewness) for bootstrapped samples of different ensemble sizes N . Solar power case study of 28 November 2014.

that radiance forecasting is in general more nonlinear.

Chapter 6

Summary

The skill of meteorological ensembles is well known to be restricted by the available ensemble size. This work demonstrated the benefit of an increased ensemble size in the frame of wind power forecasting, an application where exceptional error events have a disproportionate effect on risk management and associated costs. To analyse the impact of an ultra large ensemble size of up to 1024 members on weather dependent power forecasting, the following developments and investigations have been carried out.

A novel approach of an ensemble control system has been developed on the Petaflop supercomputer JUQUEEN, a 5-dimensional interconnected Blue Gene/Q architecture. For particle filtering with an increased ensemble size, this control system extends the WRF model to a stand-alone ensemble version. By integrating a second stage of parallelism, ensemble members are executed concurrently within a single executable. This approach enables inter-communication between the ensemble members, allowing a monitoring of the particle filter's effective sample size during runtime. Moreover, numerous Sequential Importance Resampling steps can be performed during the model run. This software represents the meteorological part of Ensembles for Stochastic Integration of Atmospheric Systems (ESIAS-met).

The system's parallel scalability was demonstrated on the basis of a feasibility study with up to 4096 ensemble members utilizing 262,144 processors. The analysis showed that deviations from perfect scalability are of a negligible order of 1 – 3 % and are caused by limited bandwidth during I/O processes. Furthermore, a real data case study demonstrated that ESIAS-met efficiently realizes Sequential Importance Resampling computations within the frame of ultra large ensemble sizes. The conclusion can be drawn that for a fixed problem size, the computational expenses of a resampling step with ESIAS-met depend exclusively on the maximum duplication number n_{\max} and are therefore independent of the ensemble size. Moreover, existing model uncertainty schemes (SPPT and SKEB) have been made computationally feasible on the Blue Gene/Q architecture.

As a basis for the ultra large ensemble, it has further been the objective to achieve a satisfactory deterministic forecast skill of the WRF model with regard to the wind power forecast in Germany. A first assessment showed a distinct sensitivity of hub height wind speeds to the choice of physical parameterizations, in particular to the land surface model and planetary boundary layer scheme. However, a large positive wind speed bias is apparent for all tested parameterization suites. The wind bias is observed solely over land, throughout the whole day and with maximum values during the night. These results led to the conclusion of a systematic model deficit, at least over Germany. Thereupon, the best performing parameterization suite served as a configuration for further model optimization. Based upon the notion that the WRF model is optimized for the Contiguous United States, the surface roughness length has been increased. Further, within the YSU boundary layer scheme, the mixing of vertical momentum in the stable regime has been enhanced. As a result, an optimized WRF model configuration could be provided, which shows competitive forecast skill of hub height wind speeds when transformed to a corresponding wind power forecast.

Based upon the previous developments and results, a major exceptional error event in wind power forecasting has been analyzed. A convection-permitting ensemble of 1024 members has been set up with perturbed initial conditions from the GEFS reforecast and state-of-the-art model uncertainty schemes (SPPT and SKEB). The ultra large ensemble has been transformed to a corresponding day-ahead wind power forecast and compared to the true power production. The analysis revealed that only few outliers could capture the error event, while exclusively assessing the ensemble spread remained insufficient for the event's identification, caused by its low predictability. The analysis of the ensemble distribution's tails by higher order moments proved to be more suitable for this purpose. Here, the skewness provided correct information about the direction of forecast error given by the outliers, while the kurtosis indicated that these outliers are arranged in the long tails of the distribution, evolving very differently from the ensemble mean. In other words, the ultra large ensemble size reveals a model state distribution, that is skewed by long tails in the proper direction of the true state. This is what ultimately defines an exceptional error event in power forecasting — an event of low probability that leads to a very different outcome than predicted by the majority of available forecast systems. A statistical evaluation based on bootstrapped samples determined the necessary ensemble sizes to resolve the higher order moments sufficiently well in magnitude and temporal evolution. The analysis indicated, that at least a size of $O(100)$ ensemble members is needed.

The model uncertainty schemes used to generate the ensemble performed quite differently. SKEB proved to be more effective in triggering instabilities in the

underlying flow, thus induced more ensemble spread and generated outliers which captured the error event. Since the error event was caused by the low predictability of an evolving upper air trough, the influence of the SPPT scheme is limited by the perturbations of parameterizations which effect this process.

To verify the main findings, a second case study of an exceptional error event in solar power forecasting has been carried out. Here, the results showed again that higher order moments correctly indicated the direction and likelihood of the forecast error. In a pronounced manner, only the full 1024-member ensemble could properly indicate the error event.

To conclude, this work showed that an ultra large meteorological ensemble allowed for more reliable identification of the likelihood and character of exceptional error events in weather dependent power forecasting, which are most likely to be erroneously represented by smaller ensemble sizes.

Chapter 7

Conclusion and outlook

The results obtained in this work suggest valuable implications for future warning systems. If the ensemble size is large enough to warn of an imminent error event and to estimate the direction of forecast error, this gives an opportunity to transmission system operators to take increased and specific action on safeguard strategies to ensure the electrical grid stability. In this case, energy market participants would be supported to reassess their risk management on a more reliable basis, which in the end leads to a more economic integration of weather dependent energy sources into the electrical grid.

Although results in this work are derived from two case studies, they demonstrated the potential of an increased ensemble size in an obvious manner. Further research is needed upon long-term periods to support the results and provide a more statistically significant analysis. Larger ensemble sizes will increasingly become available with the progress towards exascale supercomputing. Here, the combination of different parallel granularities has evolved as the preferable path. Already promising performance improvements have been reported for atmospheric models, utilizing for example graphics processing units (*Huang et al., 2015; Fuhrer et al., 2017*) or Intel Xeon Phi processors (*Michalakes, 2016; Wang et al., 2017*). Within the Energy oriented Centre of Excellence for computing applications (EoCoE) project, a study with the ESIAS-met system and an ultra large ensemble size investigating a one year period is in the planning stage. For this purpose, model resolution will necessarily be reduced, such that it has to be investigated if the results obtained with higher resolution still hold. Further, the probabilistic skill of the ultra large ensemble may be improved by postprocessing techniques. A long-term period of an ultra large ensemble will provide the necessary basis to exploit the skill of non-parametric postprocessing techniques (*Taillardat et al., 2016*) concerning their ability to improve the uncertainty representation of exceptional error events.

The software developments achieved in this work will provide a basis for further

investigations of nonlinear data assimilation in the realm of energy meteorology related matters. ESIAS-met has been designed flexible for this purpose, allowing for the application of different particle filtering techniques. Most notably, an ultra large ensemble size offers the opportunity to exploit the benefit of Big Data Analytics methods (*Riedel et al.*, 2016). Here, the potential lies in model parameter estimation as well as suitable ensemble selection within the particle filter algorithm. These approaches shall be further investigated within the EoCoE project to improve forecasts of convective cloud systems. A coupling with the ESIAS-chem system (*Franke*, 2018) is planned to realize a probabilistic forecast system with uncertainty representation in the meteorological and chemical model. Thereby, both models will exploit the skill of nonlinear data assimilation techniques. Within the realm of energy meteorology, this coupling appears promising with respect to Saharan dust outbreaks, which have shown to be a further candidate for exceptional error events in solar power forecasts (*Steiner et al.*, 2016).

Appendix A

Verification statistics at measurement towers

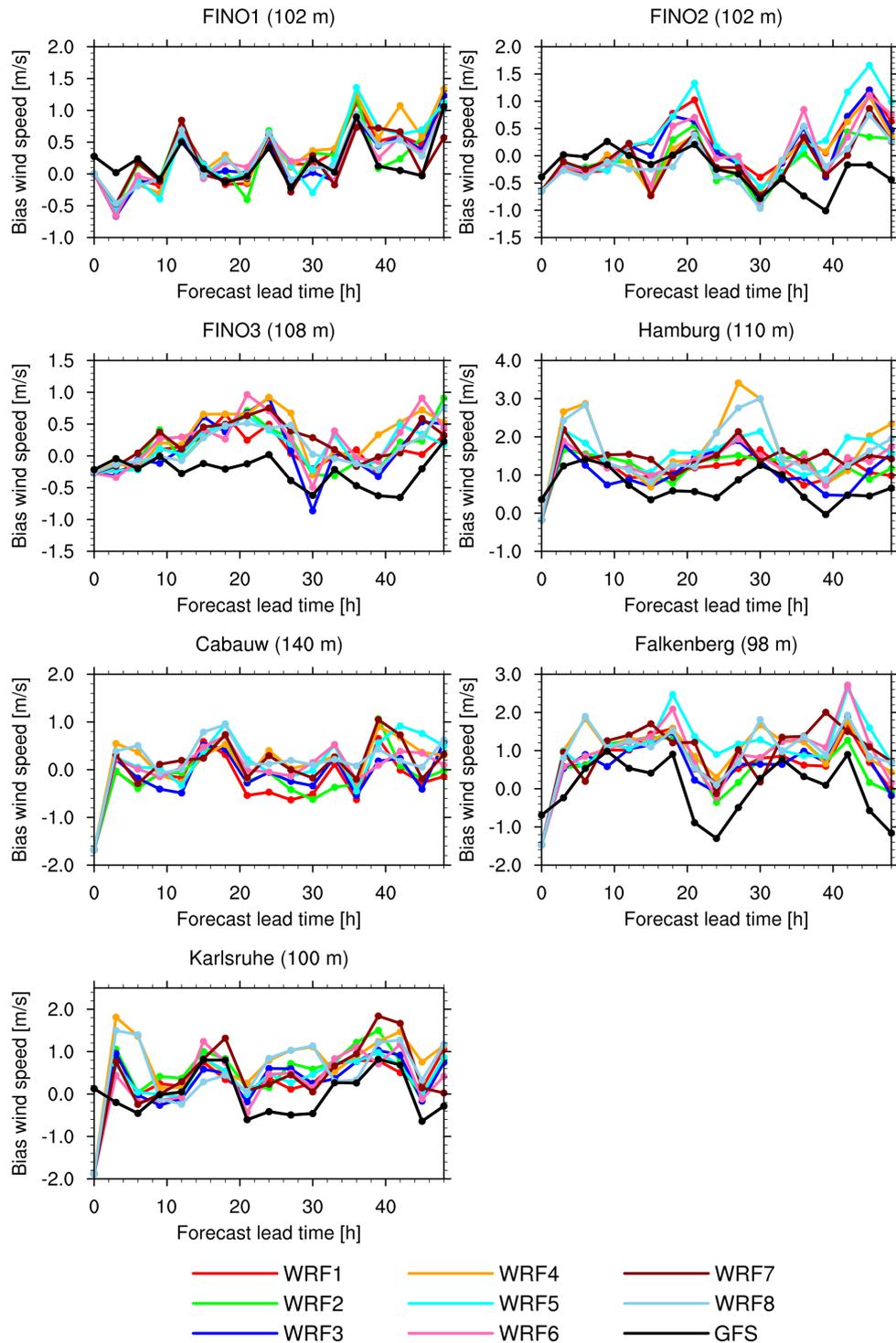


Figure A.1: Bias of hub height wind speeds at measurement towers for 1 – 31 August 2014. Different colors correspond to eight WRF model configurations. Results for GFS forecast fields are shown for comparison.

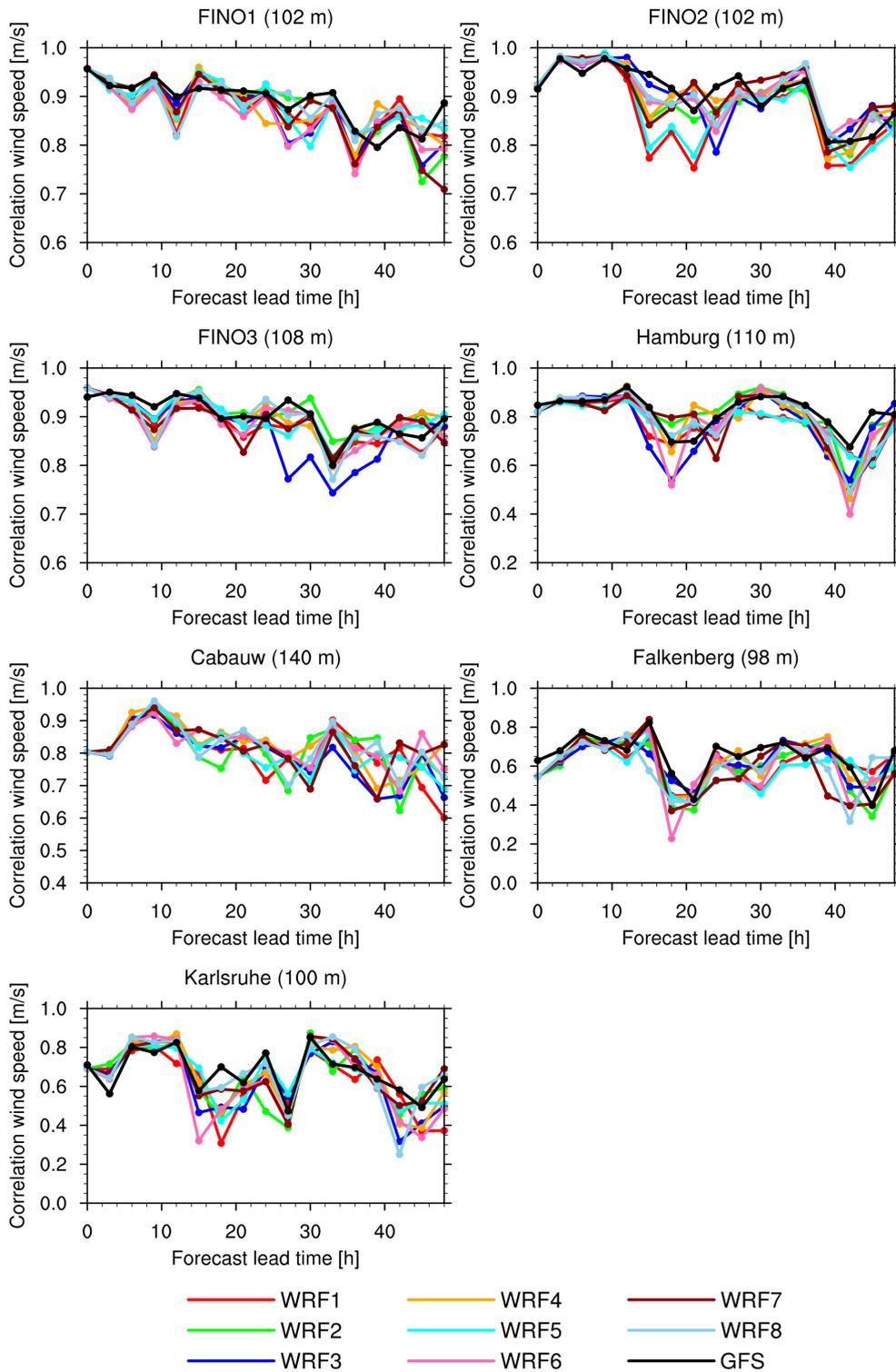


Figure A.2: Correlation between hub height wind speeds and model data at measurement towers for 1 – 31 August 2014. Different colors correspond to eight WRF model configurations. Results for GFS forecast fields are shown for comparison.

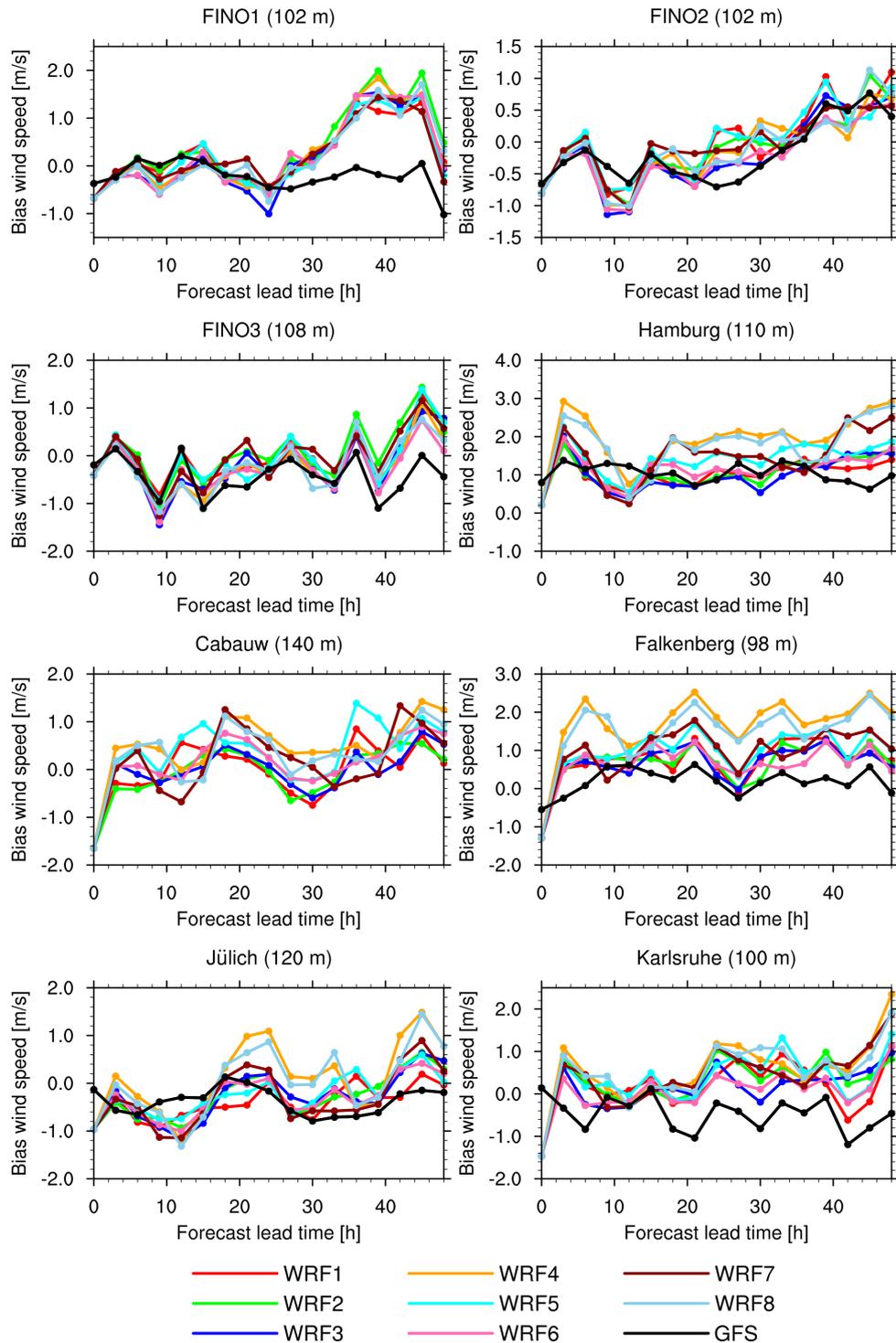


Figure A.3: Bias of hub height wind speeds at measurement towers for 1 – 30 November 2014. Different colors correspond to eight WRF model configurations. Results for GFS forecast fields are shown for comparison.

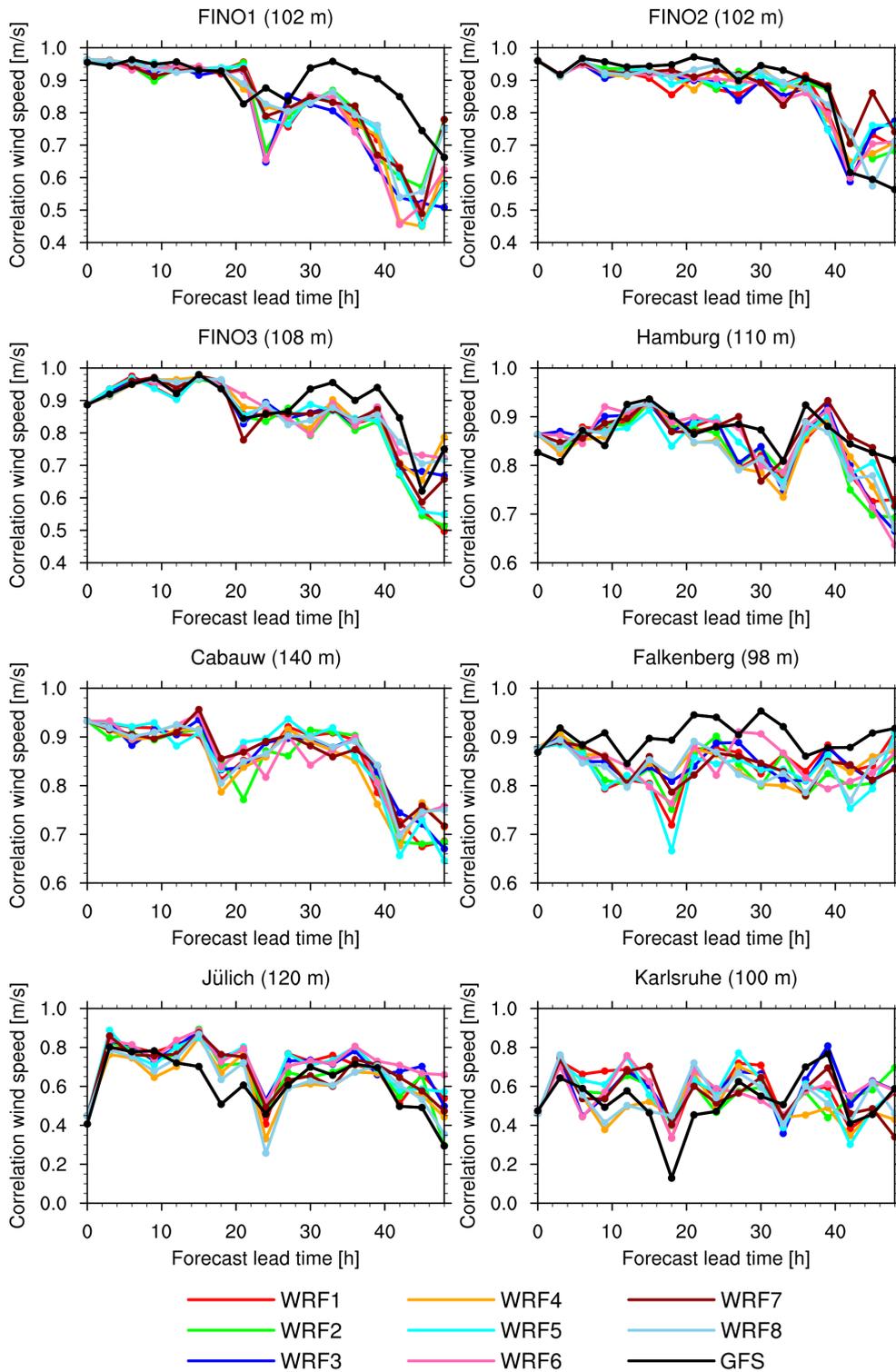


Figure A.4: Correlation between hub height wind speeds and model data at measurement towers for 1 – 30 November 2014. Different colors correspond to eight WRF model configurations. Results for GFS forecast fields are shown for comparison.

Bibliography

- Arakawa, A. and R. V. Lamb**, Computational design of the basic dynamical processes of the UCLA general circulation model, *Methods of computational physics*, 17, 173–265, 1977.
- Arakawa, A. and W. H. Schubert**, Interaction of a Cumulus Cloud Ensemble with the Large-Scale Environment, Part I, *Journal of the Atmospheric Sciences*, 31, 674–701, 1974.
- Barker, D. M., W. Huang, Y.-R. Guo, A. J. Bourgeois and Q. N. Xiao**, A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results, *Monthly Weather Review*, 132, 897–914, 2004.
- Beljaars, A. C. M. and A. A. M. Holstag**, Flux parameterization over land surfaces for atmospheric models, *Journal of Applied Meteorology*, 30, 327–341, 1991.
- Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Smirnova and R. Bleck**, Mesoscale Weather Prediction with the RUC Hybrid Isentropic-Terrain-Following Coordinate Model, *Monthly Weather Review*, 132, 473–494, 2004.
- Bernardet, L., S. Koch, E. Szoke, A. Loughe, J. L. Mahoney, L. Nance, M. Demirtas, T. Fowler, R. Gall, H.-Y. Chuang and M. Pyle**, The Developmental Testbed Center and its Winter Forecasting Experiment, *Bulletin of the American Meteorological Society*, 89, 611–627, 2008.
- Berner, J., G. J. Shutts, M. Leutbecher and T. N. Palmer**, A Spectral Stochastic Kinetic Energy Backscatter Scheme and its Impact on Flow-Dependent Predictability in the ECMWF Ensemble Prediction System, *Journal of the Atmospheric Sciences*, 66, 603–626, 2009.
- Berner, J., S.-Y. Ha, J. P. Hacker, A. Fournier and C. Snyder**, Model Uncertainty in a Mesoscale Ensemble Prediction System: Stochastic versus Multiphysics Representations, *Monthly Weather Review*, 139, 1972–1995, 2011.
- Berner, J., K. R. Fossell, S.-Y. Ha, J. P. Hacker and C. Snyder**, Increasing the Skill of Probabilistic Forecasts: Understanding Performance Improve-

- ments from Model-Error Representations, *Monthly Weather Review*, *143*, 1295–1320, 2015.
- Beyrich, F. and W. K. Adam**, Site and data report for the Lindenberg reference site in CEOP: Phase I. Berichte des Deutschen Wetterdienstes (Reports of the German Weather Service), *Issue 230*, 55, 2007.
- Bishop, C. H. and D. Hodyss**, Adaptive Ensemble Covariance Localization in Ensemble 4D-VAR State Estimation, *Monthly Weather Review*, *139*, 1241–1255, 2011.
- Borge, R., V. Alexandrov, J. J. del Vas, J. Lumbreras and E. Rodríguez**, A comprehensive sensitivity analysis of the WRF model for air quality applications over the Iberian Peninsula, *Atmospheric Environment*, *42*, 8560–8574, 2008.
- Bouttier, F., B. Viè, O. Nuissier and L. Raynaud**, Impact of Stochastic Physics in a Convection-Permitting Ensemble, *Monthly Weather Review*, *140*, 3706–3721, 2012.
- Bremen, L. V. and A. Wessel**, Kurzfrist-Windleistungsvorhersage, *Promet: Meteorologische Fortbildung. Meteorologische Aspekte der Nutzung erneuerbarer Energien*, *39*, 181–192, 2015.
- Brümmer, B., I. L. H. and H. Konow**, Atmospheric boundary layer measurements at the 280 m high Hamburg weather mast 1995-2011: mean annual and diurnal cycles, *Meteorologische Zeitschrift*, *21*, 319–335, 2012.
- Buizza, R. and T. N. Palmer**, Impact of Ensemble Size on Ensemble Prediction, *Monthly Weather Review*, *126*, 2503–2518, 1998.
- Buizza, R., M. Milleer and T. N. Palmer**, Stochastic representation of model uncertainties in the ECMWF ensemble prediction system, *Quarterly Journal of the Royal Meteorological Society*, *125*, 28872908, 1999.
- Caplan, P., J. Derber, W. Gemmill, S.-Y. Hong, H.-L. Pan and D. Parrish**, Changes to the 1995 NCEP Operational Medium-Range Forecast Model Analysis-Forecast System, *Weather and Forecasting*, *12*, 581–594, 1997.
- Carns, P., R. Latham, R. Ross, K. Iskra, S. Lang and K. Riley**, 24/7 characterization of petascale I/O workloads, *Proceedings of 2009 Workshop on Interfaces and Architectures for Scientific Data Storage*, September 2009.
- Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell and L. Michelin**, Toward Random Sampling of Model Error in the Canadian Ensemble Prediction System, *Monthly Weather Review*, *138*, 1877–1901, 2010.

- Chen, D., N. Eisley, P. Heidelberger, R. M. Senger, Y. Sugawara, S. Kumar, V. Salapura, D. L. Satterfield, B. D. Steinmacher-Burow and J. Parker**, The IBM Blue Gene/Q Interconnection Fabric, *IEEE Micro*, 32–43, 2012.
- Cheng, W. Y. Y. and J. Steenburgh**, Evaluation of surface sensible weather forecast by the WRF and the Eta Models over the western United States, *Weather Forecasting*, 20, 812–821, 2005.
- Chorin, A., M. Morzfeld and X. Tu**, Implicit particle filters for data assimilation, *Communication in Applied Mathematics and Computational Science*, 5, 221–240, 2010.
- Christensen, H. M., I. M. Moroz and T. N. Palmer**, Stochastic and Perturbed Parameter Representations of Model Uncertainty in Convection Parameterization, *Journal of the Atmospheric Sciences*, 72, 2525–2544, 2015.
- Cohen, A. E., S. M. Cavallo, M. C. Coniglio and H. E. Brooks**, A Review of planetary boundary layer parameterization schemes and their sensitivity in simulating Southeastern U.S. cold season severe weather environments, *Weather and Forecasting*, 30, 591–612, 2015.
- Consentec**, Gutachten zur Dimensionierung des Regelleistungsbedarfs unter dem NRV, *Consulting für Energiewirtschaft und -technik GmbH*, 2010.
- Consentec**, Weiterentwicklung des Ausgleichsenergie-Preissystems im Rahmen des Verfahrens BK6-12-024 der Bundesnetzagentur, *Consulting für Energiewirtschaft und -technik GmbH*, 2012.
- Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch and A. Staniforth**, The Operational CMC MRB Global Environmental Multiscale (GEM) Model. Part I: Design Considerations and Formulation, *Monthly Weather Review*, 126, 1373–1395, 1998.
- Courtier, P., J.-N. Thépaut and A. Hollingsworth**, A strategy for operational implementation of 4D-Var, using an incremental approach, *Quarterly Journal of the Royal Meteorological Society*, 120, 1367–1387, 1994.
- Daley, R.**, Basic theory of data assimilation, *Cambridge University Press, New York*, pp. 457, 1991.
- Denhard, M., C. Primo, F. Fundel, A. Rhodin, A. F. del Rio, H. Anlauf, A. Cress, M. Buchhold and R. Potthast**, The global ICON Ensemble at DWD, *16th EMS Annual Meeting & 11th European Conference on Applied Climatology (ECAC)*, Trieste, Italy, 2016.

- Dobschinski, J., R. Bessa, P. Du, K. Geisler, S. E. Haupt, M. Lange, C. Möhrlen, D. Nakafuji and M. de la Torre Rodriguez**, Uncertainty Forecasting in a Nutshell: Prediction Models Designed to Prevent Significant Errors, *IEEE Power and Energy Magazine*, 15, 40–49, 2017.
- Douc, R., O. Cappé and E. Moulines**, Comparison of Resampling Schemes for Particle Filtering, *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, 64–69, 2005.
- Drechsel, S., G. J. Mayr, J. W. Messner and R. Stauffer**, Wind Speeds at Heights Crucial for Wind Energy: Measurements and Verification of Forecasts, *Journal of Applied Meteorology and Climatology*, 51, 1602–1617, 2012.
- Duda, J. D., X. Wang, F. Kong, M. Xue and J. Berner**, Impact of a Stochastic Kinetic Energy Backscatter Scheme on Warm Season Convection-Allowing Ensemble Forecasts, *Monthly Weather Review*, 144, 1887–1908, 2016.
- Dudhia, J.**, Numerical Study of Convection Observed during the Winter Monsoon Experiment Using a Mesoscale Two-Dimensional Model, *Journal of the Atmospheric Sciences*, 46, 3077–3107, 1989.
- Dudhia, J.**, A Multi-layer Soil Temperature Model for MM5, *The Sixth PSU/NCAR Mesoscale Model Users' Workshop*, 1996.
- Eckel, F. and C. Mass**, Aspects of Effective Mesoscale, Short-Range Ensemble Forecasting, *Weather and Forecasting*, 20, 328–350, 2005.
- Elbern, H., A. Strunk, H. Schmidt and O. Talagrand**, Emission rate and chemical state estimation by 4-dimensional variational inversion, *Atmospheric Chemistry and Physics*, 7, 3749–3769, 2007.
- Epstein, E. S.**, Stochastic dynamic prediction, *Tellus*, 21, 1969.
- Evensen, G.**, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research*, 99, 10143–10162, 1994.
- Fisher, M., Y. Trémolet, H. Auvinen, D. Tan and P. Poli**, Weak-Constraint and Long-Window 4D-Var, *Technical Memorandum*, 2011.
- Focken, U., M. Lange and H.-P. Waldl**, Previento - A Wind Power Prediction System with an innovative upscaling algorithm, *Proceedings of the European Wind Energy Conference, Copenhagen, Denmark*, 826–829, 2001.

- Focken, U., M. Lange, K. Mönnich, H.-P. Waldl, H. G. Beyer and A. Luig**, Short-term prediction of the aggregated power output of wind farms - a statistical analysis of the reduction of the prediction error by spatial smoothing effects, *Journal of Wind Engineering and Industrial Aerodynamics*, 90, 231–246, 2002.
- Foken, T.**, 50 Years of the Monin-Obukhov Similarity Theory, *Boundary-Layer Meteorology*, 119, 431–447, 2006.
- Franke, P.**, Quantitative estimation of unexpected emissions in the atmosphere by stochastic inversion techniques, *PhD Thesis, University of Cologne, Germany*, submitted 23 Oct 2017, 2018.
- Frederiksen, J. S. and A. G. Davies**, The Regularized Dia Closure for Two-Dimensional Turbulence, *Geophysical and Astrophysical Fluid Dynamics*, 98, 203–223, 2004.
- Fuhrer, O., T. Chadha, T. Hoefler, G. Kwasniewski, X. Lapillonne, D. Leutwyler, D. Lüthi, C. Osuna, C. Schär, T. C. Schulthess and H. Vogt**, Near-global climate simulation at 1 km resolution: establishing a performance baseline on 4888 GPUs with COSMO 5.0, *Geoscientific Model Development Discussions*, 2017, 1–27, 2017.
- García-Díez, M., J. Fernández, L. Fita and C. Yagüe**, Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe, *Quarterly Journal of the Royal Meteorological Society*, 139, 501–514, 2013.
- Geimer, M., F. Wolf, B. J. N. Wylie, E. Ábrahám, D. Becker and B. Mohr**, The Scalasca performance toolset architecture, *Concurrency and Computation: Practice and Experience*, 22, 702–719, 2010.
- German Weather Service**, *Climate Data Center*, 2017.
- Giebel, G., G. Kariniotakis and R. Brownsword**, The state-of-the-art in short-term prediction of wind power: A literature overview, *Position paper created for the ANEMOS project*, 2005.
- Giebel, G., R. Brownsword, G. Kariniotakis, M. Denhard and C. Draxl**, The state-of-the-art in short-term prediction of wind power: A literature overview, 2nd edition, *Project report for the ANEMOS.plus and SafeWind projects, Risø DTU.*, 2011.
- Good, G. H.**, Power Ensemble Evaluation, *Technical Report, Energy oriented Center of Excellence for computing applications*, pp. 22, 2017.

- Good, G. H. and J. Berndt**, Anticipating extreme wind and pv power forecast errors using ultra-large ensembles, *European Conference for Applied Meteorology and Climatology, Dublin, Ireland*, 2017.
- Grell, G. A. and S. R. Freitas**, A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling, *Atmospheric Chemistry and Physics*, 14, 5233–5250, 2014.
- Grell, G. A., J. Dudhia and D. R. Stauffer**, A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5), *NCAR Technical Note TN-398 + STR*, 1994.
- Grimit, E. P. and C. F. Mass**, Measuring the Ensemble Spread-Error Relationship with a Probabilistic Approach: Stochastic Ensemble Results, *Monthly Weather Review*, 135, 203–221, 2007.
- Hacker, R.**, Certification of Algorithm 112: Position of point relative to polygon, *Communications of the ACM*, 12, 606, 1962.
- Ha, S., J. Berner and C. Snyder**, A Comparison of Model Error Representations in Mesoscale Ensemble Data Assimilation, *Monthly Weather Review*, 143, 3893–3911, 2015.
- Hagelin, S., J. Son, R. Swinbank, A. McCabe, N. Roberts and W. Tennant**, The Met Office convective-scale ensemble, MOGREPS-UK, *Quarterly Journal of the Royal Meteorological Society*, 2077–2107, 2017.
- Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. G. Jr., Y. Zhu and W. Lapenta**, NOAA’s second-generation global medium-range ensemble reforecast data set, *Bulletin of the American Meteorological Society*, 94, 1553–1565, 2013.
- Hansen, F. V.**, Surface roughness lengths, *U.S. Army Research Laboratory, U.S. Army Technical Report No. ARL-TR-61, White Sands Missile Range, New Mexico, USA*, 1962.
- Hirth, L. and I. Ziegenhagen**, Balancing power and variable renewables: A glimpse at german data, *Technical report, Vattenfall GmbH and Potsdam-Institute for CLimate IMPact Reserach and Prognos AG*, 2013.
- Hong, S.-Y. and J.-O. J. Lim**, The WRF single-moment 6-class microphysics scheme (WSM6), *Journal of the Korean Meteorological Society*, 42, 129–151, 2006.
- Hong, S.-Y., J. Dudhia and S.-H. Chen**, A Revised Approach to Ice Microphysical Processes for the Bulk Parameterization of Clouds and Precipitation, *Monthly Weather Review*, 132, 103–120, 2004.

- Hong, S.-Y., Y. Noh and J. Dudhia**, A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes, *Monthly Weather Review*, *134*, 2318–2341, 2006.
- Houtekamer, P. L., L. Lefaitvre, J. Derome, H. Ritchie and H. L. Mitchell**, A System Simulation Approach to Ensemble Prediction, *Monthly Weather Review*, *124*, 1225–1242, 1996.
- Huang, X.-Y., Q. Xiao, D. M. Barker, X. Zhang, J. Michalakes, W. Huang, T. Henderson, J. Bray, Y. Chen, Z. Ma, J. Dudhia, Y. Guo, X. Zhang, D.-J. Won, H.-C. Lin and Y.-H. Kuo**, Four-Dimensional Variational Data Assimilation for WRF: Formulation and Preliminary Results, *Monthly Weather Review*, *137*, 299–314, 2009.
- Huang, M., J. Mielikainen, B. Huang, H. Chen, H.-L. A. Huang and M. D. Goldberg**, Development of efficient GPU parallelization of WRF Yonsei University planetary boundary layer scheme, *Geoscientific Model Development*, *8*, 2977–2990, 2015.
- Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough and C. W. D.**, Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, *Journal of Geophysical Research: Atmospheres*, *113*, D13103, 2008.
- Jacob, M.**, Beeinflussung von Windmessungen an einem Rohrmast durch die Maststruktur, *Bachelor thesis, University of Hamburg, Germany*, 2013.
- Janjić, Z. I.**, The Step-Mountain Eta Coordinate Model: Further Developments of the Convection, Viscous Sublayer, and Turbulence Closure Schemes, *Monthly Weather Review*, *122*, 927–945, 1994.
- Jankov, I., J. Berner, J. Beck, H. Jiang, J. B. Olson, G. Grell, T. G. Smirnova, S. G. Benjamin and J. M. Brown**, A Performance Comparison between Multiphysics and Stochastic Approaches within a North American RAP Ensemble, *Monthly Weather Review*, *145*, 1161–1179, 2017.
- Jiménez, P. A. and J. Dudhia**, Improving the Representation of Resolved and Unresolved Topographic Effects on Surface Wind in the WRF Model, *Journal of Applied Meteorology and Climatology*, *51*, 300–316, 2012.
- Jimenez, P. A., J. P. Hacker, J. Dudhia, S. E. Haupt, J. A. Ruiz-Arias, C. A. Gueymard, G. Thompson, T. Eidhammer and A. Deng**, WRF-Solar: Description and Clear-Sky Assessment of an Augmented NWP Model for Solar Power Prediction, *Bulletin of the American Meteorological Society*, *97*, 1249–1264, 2016.

- Jiming, J., N. Miller and N.-J. Schlegel**, Sensitivity Study of Four Land Surface Schemes in the WRF Model, *Advances in Meteorology*, 2010, pp. 11, 2010.
- Jülich Supercomputing Centre**, JUQUEEN: IBM Blue Gene/Q Supercomputer System at the Jülich Supercomputing Centre, *Journal of large-scale research facilities*, 1, (A1), 2015.
- Kain, J. S.**, The Kain-Fritsch Convective Parameterization: An Update, *Journal of Applied Meteorology*, 43, 170–181, 2004.
- Kalnay, E.**, Atmospheric modelling, data assimilation and predictability, *Cambridge University Press*, pp. 341, 2003.
- Kessler, E.**, On the distribution and continuity of water substance in atmospheric circulations, *American Meteorological Society, Meteorological Monographs*, 32, 129–151, 1969.
- K. Gilliland, E. and C. Rowe**, A comparison of cumulus parameterization scheme in the WRF model, *AMS Annual Meeting, 87, and Conference on Hydrology, 21, Proceeding, San Antonio, Texas, USA*, 2007.
- Kitagawa, G.**, Monte-Carlo filter and smoother for non-Gaussian non-linear state-space models, *Journal of Computational and Graphical Statistics*, 5, 1–25, 1998.
- Klemp, J. B., W. C. Skamarock and J. Dudhia**, Conservative split-explicit time integration methods for the compressible nonhydrostatic equations, *Monthly Weather Review*, 135, 2897–2913, 2007.
- Knüpfer, A., C. Rössel, D. an Mey, K. Diethelm, D. Eschweiler, M. Geimer, M. Gerndt, D. Lorenz, A. Malony, W. E. Nagel, Y. Oleynik, P. Philippen, P. Saviankou, D. Schmidl, S. Shende, R. Tschüter, M. Wagner, B. Wesarg and F. Wolf**, Score-P - A Joint Performance Measurement Run-Time Infrastructure for Periscope, Scalasca, TAU, and Vampir, *Proceedings of 5th Parallel Tools Workshop, Springer*, 79–91, 2012.
- Konow, H. M.**, Tall Wind Profiles in Heterogeneous Terrain, PhD thesis, *University of Hamburg, Germany*, 2015.
- Krieger, J., J. Zhang, D. Atkinson, X. Zhang and M. Shulski**, Sensitivity of WRF model forecasts to different physical parameterizations in the Beaufort sea region, *Proceedings of the 8th conference on coastal atmospheric and oceanic prediction and processes Phoenix, Ariz, USA, Phoenix, Arizona, USA*, 2009.
- Laprise, A.**, The Euler Equations of motion with hydrostatic pressure as in independent variable, *Monthly Weather Review*, 120, 197–207, 1992.

- Lee, P., D. Kang, J. McQueen, M. Tsidulko, M. Hart, G. DiMego, N. Seaman and P. Davidson**, Impact of Domain Size on Modeled Ozone Forecast for the Northeastern United States, *Journal of Applied Meteorology and Climatology*, *47*, 443–461, 2008.
- Leith, C. E.**, Theoretical Skill of Monte Carlo Forecasts, *Monthly Weather Review*, *102*, 409–418, 1974.
- Li, X., M. Charron, L. Spacek and G. Candille**, A Regional Ensemble Prediction System Based on Moist Targeted Singular Vectors and Stochastic Parameter Perturbations, *Monthly Weather Review*, *136*, 443–462, 2008.
- Lin, Y.-L., R. D. Farley and H. D. Orville**, Bulk Parameterization of the Snow Field in a Cloud Model, *Journal of Climate and Applied Meteorology*, *22*, 1065–1092, 1983.
- Lorenz, E. N.**, Deterministic nonperiodic flow, *Journal of Atmospheric Sciences*, *20*, 130–141, 1963.
- Lorenz, E.**, The predictability of a flow which possesses many scales of motion, *Tellus*, *21*, 289–307, 1969.
- Lowrey, M. R. K. and Z.-L. Yang**, Assessing the Capability of a Regional-Scale Weather Model to Simulate Extreme Precipitation Patterns and Flooding in Central Texas, *Weather and Forecasting*, *23*, 1102–1126, 2008.
- Lui, J. S. and R. Chen**, Sequential Monte-Carlo methods for dynamical systems, *Journal of the American Statistical Association*, *90*, 567576, 1998.
- Lundgren, K.**, Aktuelle Arbeiten im Rahmen von EWeLiNE, *5. Treffen der Industrie- und Forschungsplattform Prognose, EWeLiNE project, Kassel, Germany*, 2015.
- Lydia, M., S. S. Kumar, A. I. Selvakumar and G. E. P. Kumar**, A comprehensive review on wind turbine power curve modeling techniques, *Renewable and Sustainable Energy Reviews*, *30*, 452–460, 2014.
- Mason, P. J. and D. J. Thomson**, Stochastic backscatter in large-eddy simulations of boundary layers, *Journal of Fluid Mechanics*, *242*, 51–78, 1992.
- Mass, C. and D. Ovens**, Fixing WRFs high speed wind bias: A new subgrid scale drag parameterization and the role of detailed verification, *Preprints, 24th Conference on Weather and Forecasting/20th Conference on Numerical Weather Prediction, Seattle, USA, American Meteorological Society*, 2011.

- Mass, C. F., D. Ovens, K. Westrick and B. A. Colle**, Does increasing horizontal resolution produce more skillful forecasts?, *Bulletin of the American Meteorological Society*, *83*, 407–430, 2002.
- Masson, V., J.-L. Champeaux, F. Chauvin, C. Meriguet and R. Lacaze**, A Global Database of Land Surface Parameters at 1-km Resolution in Meteorological and Climate Models, *Journal of Climate*, *16*, 1261–1282, 2003.
- McCabe, A., R. Swinbank, W. Tennant and A. Lock**, Representing model uncertainty in the Met Office convection-permitting ensemble prediction system and its impact on fog forecasting, *Quarterly Journal of the Royal Meteorological Society*, *142*, 2897–2910, 2016.
- McLean, J.**, Further developing europe’s power market for large scale integration of wind power: Wp 2.6 equivalent power curves, *Technical report, TradeWind Consortium*, 2008.
- Michalakes, J.**, Experiences Optimizing NWP for Intels Xeon Phi Processor, *17th Workshop on High Performance Computing in Meteorology, Boulder, USA*, 2016.
- Min, K.-H., S. Choo, D. Lee and G. Lee**, Evaluation of WRF Cloud Microphysics Schemes Using Radar Observations, *Weather and Forecasting*, *30*, 1571–1589, 2015.
- Miyoshi, T., K. Kondo and T. Imamura**, The 10,240-member ensemble Kalman filtering with an intermediate AGCM, *Geophysical Research Letters*, *41*, 5264–5271, 2014.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono and C. A. Shepard**, Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, *Journal of Geophysical Research: Atmospheres*, *102*, 16663–16682, 1997.
- Möhrlen, K. and J. U. Jørgensen**, Forecasting Wind Power in High Wind Penetration Markets using Multi-Scheme Ensemble Prediction Methods, *Proceedings of the German Wind Energy Conference, Bremen, Germany*, 2008.
- Möhrlen, C. and J. U. Jørgensen**, The Role of Ensemble Forecasting in Integrating Renewables into Power Systems: From Theory to Real-Time Applications, *Integration of Large-Scale Renewable Energy into Bulk Power Systems*, 2017.
- Molteni, F., R. Buizza, T. N. Palmer and T. Petroliagis**, The ECMWF Ensemble Prediction System: Methodology and validation, *Quarterly Journal of the Royal Meteorological Society*, *122*, 73–119, 1996.

- Montani, A., D. Cesari, C. Marsigli and T. Paccagnella**, Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges, *Tellus*, *63a*, 605–624, 2011.
- Mullen, S. L. and D. P. Baumhefner**, Sensitivity of Numerical Simulations of Explosive Oceanic Cyclogenesis to Changes in Physical Parameterizations, *Monthly Weather Review*, *116*, 2289–2329, 1988.
- Mullen, S. L. and R. Buizza**, The Impact of Horizontal Resolution and Ensemble Size on Probabilistic Forecasts of Precipitation by the ECMWF Ensemble Prediction System, *Weather and Forecasting*, *17*, 173–191, 2002.
- Nakanishi, M. and H. Niino**, Development of an improved turbulence closure model for the atmospheric boundary layer, *Journal of the Meteorological Society Japan*, *87*, 895912, 2009.
- Nastrom, G. D. and K. S. Gage**, A Climatology of Atmospheric Wavenumber Spectra of Wind and Temperature Observed by Commercial Aircraft, *Journal of the Atmospheric Sciences*, *42*, 950–960, 1985.
- Noh, Y., W. G. Cheon, S. Y. Hong and S. Raasch**, Improvement of the K-profile Model for the Planetary Boundary Layer based on Large Eddy Simulation Data, *Boundary-Layer Meteorology*, *107*, 401–427, 2003.
- Nutter, P., D. Stensrud and M. Xue**, Effects of Coarsely Resolved and Temporally Interpolated Lateral Boundary Conditions on the Dispersion of Limited-Area Ensemble Forecasts, *Monthly Weather Review*, *132*, 2358–2377, 2004.
- Olsen, B. T., A. N. Hahmann, A. M. Sempreviva, J. Badger and H. E. Joergensen**, Intercomparison of state-of-the-art models for wind energy resources with mesoscale models, *European Geosciences Union General Assembly 2016, Vienna, Austria*, 2016.
- Palmer, T. N. and R. Hagedorn**, Predictability of Weather and Climate, *Cambridge University Press*, pp. 702, 2006.
- Palmer, T., G. Shutts, R. Hagedorn, F. Doblas-Reyes, T. Jung and M. Leutbecher**, Representing model uncertainty in weather and climate prediction, *Annual Review of Earth and Planetary Sciences*, *33*, 163–193, 2005.
- Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer and A. Weisheimer**, Stochastic parameterization and model uncertainty, *Technical Memorandum*, *598*, 42, 2009.
- Park, S. K. and K. W. Miller**, Random number generators: good ones are hard to find, *Communication of the ACM*, *31*, 1192–1201, 1988.

- Pennelly, C., G. Reuter and T. Flesch**, Verification of the WRF model for simulating heavy precipitation in Alberta, *Atmospheric Environment*, 135-136, 172–192, 2014.
- Pielke, R. A.**, Mesoscale Meteorological Modelling, *Academic Press*, pp. 612, 1984.
- Pieri, A. B., J. von Hardenberg, A. Parodi and A. Provenzale**, Sensitivity of Precipitation Statistics to Resolution, Microphysics, and Convective Parameterization: A Case Study with the High-Resolution WRF Climate Model over Europe, *Journal of Hydrometeorology*, 16, 1857–1872, 2015.
- Pitt, M. K. and N. Shephard**, Filtering via simulation: Auxiliary particle filters, *Journal of the American Statistical Association*, 94, 590–599, 1999.
- Poterjoy, J.**, A Localized Particle Filter for High-Dimensional Nonlinear Systems, *Monthly Weather Review*, 144, 59–76, 2016.
- Powers, J. G., J. B. Klemp, W. C. Skamarock, C. A. Davis, J. Dudhia, D. O. Gill, J. L. Coen, D. J. Gochis, R. Ahmadov, S. E. Peckham, G. A. Grell, J. Michalakes, S. Trahan, S. G. Benjamin, C. R. Alexander, G. J. Dimego, W. Wang, C. S. Schwartz, G. S. Romine, Z. Liu, C. Snyder, F. Chen, M. J. Barlage, W. Yu and M. G. Duda**, The Weather Research and Forecasting (WRF) Model: Overview, System Efforts, and Future Directions, *Bulletin of the American Meteorological Society*, 98, 1717–1737, 2017.
- Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf and A. Simmons**, The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics, *Quarterly Journal of the Royal Meteorological Society*, 126, 1143–1170, 2000.
- Richardson, D. S.**, Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size, *Quarterly Journal of the Royal Meteorological Society*, 127, 2473–2489, 2001.
- Riedel, M., T. Lippert, D. Mallmann and G. Cavallaro**, Scientific Big Data Analytics by HPC, *Innovatives Supercomputing in Deutschland*, 14, 70–74, 2016.
- Ritchie, H., C. Temperton, A. Simmons, M. Hortal, T. Davies, D. Dent and M. Hamrud**, Implementation of the Semi-Lagrangian Method in a High-Resolution Version of the ECMWF Forecast Model, *Monthly Weather Review*, 123, 489–514, 1995.
- Romine, G. S., C. S. Schwartz, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson and M. L. Weisman**, Representing Forecast Error in a Convection-Permitting Ensemble System, *Monthly Weather Review*, 142, 4519–4541, 2014.

- Roux, G., Y. Liu, D. L. Monache, R.-S. Sheu and T. T. Warner**, Verification of high resolution WRF-RTFDAA surface forecasts over mountains and plains, *Preprints, 2009 9th Annual WRF Users Workshop, Boulder, CO, NCAR.*, 2009.
- Ruiz, J. J., C. Saulo and J. Nogués-Paegle**, WRF Model Sensitivity to Choice of Parameterization over South America: Validation against Surface Variables, *Monthly Weather Review*, 138, 3342–3355, 2010.
- Saint-Drenan, Y. M., G. H. Good and M. Braun**, A probabilistic approach to the estimation of regional photovoltaic power production, *Solar Energy*, 276, 147–257, 2017.
- Sanchez, C., K. D. Williams and M. Collins**, Improved stochastic physics schemes for global weather and climate models, *Quarterly Journal of the Royal Meteorological Society*, 142, 147–159, 2016.
- Schroedter-Homscheidt, M., H. Deifel and D. Heinemann**, Grundlagen des Energiesystems für Energiemeteorologen, *Promet: Meteorologische Fortbildung. Meteorologische Aspekte der Nutzung erneuerbarer Energien*, 39, 138–150, 2015.
- Shutts, G.**, A kinetic energy backscatter algorithm for use in ensemble prediction systems, *Quarterly Journal of the Royal Meteorological Society*, 131, 3079–3102, 2005.
- Shutts, G.**, Coarse Graining the Vorticity Equation in the ECMWF Integrated Forecasting System: The Search for Kinetic Energy Backscatter, *Journal of the Atmospheric Sciences*, 70, 1233–1241, 2013.
- Shutts, G.**, A stochastic convective backscatter scheme for use in ensemble prediction systems, *Quarterly Journal of the Royal Meteorological Society*, 141, 2602–2616, 2015.
- Siuta, D., G. West and R. Stull**, WRF Hub-Height Wind Forecast Sensitivity to PBL Scheme, Grid Length, and Initial Condition Choice in Complex Terrain, *Weather and Forecasting*, 32, 493–509, 2017.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang and J. G. Powers**, A Description of the Advanced Research WRF Version 3, *AVAILABLE FROM NCAR; P.O. BOX 3000; BOULDER, CO*, 2008.
- Snyder, C., T. Bengtsson, P. Bickel and J. Anderson**, Obstacles to High-Dimensional Particle Filtering, *Monthly Weather Review*, 136, 4629–4640, 2008.
- Stark, M.**, Prognoseoptimierung vs. Wirtschaftlichkeit-Nutzen von Prognosequantilen im Intra-day Handel, 5. *Treffen der Industrie- und Forschungsplattform Prognose, EWeLiNE project, Kassel, Germany*, 2015.

- Steiner, A., V. Wehner and J. Förstner**, Forecasting the reduction in photovoltaic power production during Saharan dust outbreaks, *16th EMS Annual Meeting & 11th European Conference on Applied Climatology (ECAC), Trieste, Italy*, 2016.
- Steiner, A., C. Köhler, I. Metzinger, A. Braun, M. Zirkelbach, D. Ernst, P. Tran and B. Ritter**, Critical weather situations for renewable energies - part a: Cyclone detection for wind power, *Renewable Energy*, 41–50, 2017.
- Stensrud, D. J., J.-W. Bao and T. T. Warner**, Using Initial Condition and Model Physics Perturbations in Short-Range Ensemble Simulations of Mesoscale Convective Systems, *Monthly Weather Review*, 128, 2077–2107, 2000.
- Strogatz, S. H.**, Nonlinear Dynamics and Chaos, *Westview Press*, pp. 513, 2014.
- Stull, R.**, An Introduction to Boundary Layer Meteorology, *Springer, Berlin*, 1988.
- Taillardat, M., O. Mestre, M. Zamo and P. Naveau**, Calibrated Ensemble Forecasts Using Quantile Regression Forests and Ensemble Model Output Statistics, *Monthly Weather Review*, 144, 2375–2393, 2016.
- Tennant, W.**, Improving initial condition perturbations for MOGREPS-UK, *Quarterly Journal of the Royal Meteorological Society*, 141, 2324–2336, 2015.
- Tennant, W. J., G. J. Shutts, A. Arribas and S. A. Thompson**, Using a Stochastic Kinetic Energy Backscatter Scheme to Improve MOGREPS Probabilistic Forecast Skill, *Monthly Weather Review*, 139, 1190–1206, 2011.
- Tewari, M., F. Chen, W. Wang, J. Dudhia, M. A. LeMone, K. Mitchell, M. Ek, G. Gayno, J. Wegiel and R. H. Cuenca**, Implementation and verification of the unified NOAA land surface model in the WRF model, *20th conference on weather analysis and forecasting/16th conference on numerical weather prediction*, 11–15, 2004.
- Theis, S. E., A. Hense and U. Damrath**, Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach, *Meteorological Applications*, 12, 257–268, 2005.
- Thomas, P. and S. Vogt**, Variances of the vertical and horizontal wind measured by tower instruments and SODAR. An intercomparison, *Applied Physics B: Lasers and Optics*, 57, 19–26, 1993.
- Thuburn, J., J. Kent and N. Wood**, Cascades, backscatter and conservation in numerical models of two-dimensional turbulence, *Quarterly Journal of the Royal Meteorological Society*, 140, 2013.

- Toth, Z. and E. Kalnay**, Ensemble Forecasting at NMC: The Generation of Perturbations, *Bulletin of the American Meteorological Society*, 74, 2317–2330, 1993.
- Toth, Z. and E. Kalnay**, Ensemble Forecasting at NCEP and the Breeding Method, *Monthly Weather Review*, 125, 3297–3319, 1997.
- Troen, I. B. and L. Mahrt**, A simple model of the atmospheric boundary layer sensitivity to surface evaporation, *Boundary Layer Meteorology*, 37, 129–148, 1986.
- van Dop, H.**, Terrain classification and derived meteorological parameters for inter-regional transport models, *Atmospheric Environment*, 17, 1099–1105, 1983.
- van Leeuwen, P. J.**, Particle Filtering in Geophysical Systems, *Monthly Weather Review*, 137, 4089–4114, 2009.
- van Leeuwen, P. J.**, Nonlinear data assimilation in geosciences: an extremely efficient particle filter, *Quarterly Journal of the Royal Meteorological Society*, 136, 2010.
- van Ulden, A. P. and J. Wieringa**, Atmospheric boundary layer research at Cabauw, *Boundary Layer Meteorology*, 78, 39–69, 1996.
- Vezzole, P.**, IBM Blue Gene/Q Overview, *Prace Winter School. Bologna, Italy*, 2012.
- Vogt, S., J. Dobschinski, T. Kanefendt, S. Otterson and Y. Saint-Drenan**, A Hybrid Physical and machine Learning Based Forecast of Regional Wind Power, *15th Wind Integration Workshop*, 2016.
- Walaszek, K., M. Kryza and M. Werner**, A Sensitivity Analysis of the WRF Model to Shortwave Radiation Schemes for Air Quality Purposes and Evaluation with Observational Data, *Air Pollution Modeling and its Application*, 539–543, 2014.
- Wang, X. and T. Lei**, GSI-Based Four-Dimensional Ensemble-Variational (4DEnsVar) Data Assimilation: Formulation and Single-Resolution Experiments with Real Data for NCEP Global Forecast System, *Monthly Weather Review*, 142, 3303–3325, 2014.
- Wang, X., D. M. Barker, C. Snyder and T. M. Hamill**, A Hybrid ETKF3DVAR Data Assimilation Scheme for the WRF Model. Part I: Observing System Simulation Experiment, *Monthly Weather Review*, 136, 5116–5131, 2008.
- Wang, H., H. Chen, Q. Wu, J. Lin, X. Chen, X. Xie, R. Wang, X. Tang and Z. Wang**, GNAQPMS v1.1: accelerating the Global Nested Air Quality Prediction Modeling System (GNAQPMS) on Intel Xeon Phi processors, *Geoscientific Model Development*, 10, 2891–2904, 2017.

- Warner, T. T., R. A. Peterson and E. T. Russell**, 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, 2599–2617, 1997.
- Weaver, A. and P. Courtier**, Correlation modelling on the sphere using a generalized diffusion equation, *Quarterly Journal of the Royal Meteorological Society*, 127, 1815–1846, 2001.
- Wei, M., Z. Toth, R. Wobus and Y. ZHU**, Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system, *Tellus A*, 60, 2008.
- Wenzel, A., Kalthoff and V. Horlacher**, On the profiles of wind velocity in the roughness sublayer above a coniferous forest, *Boundary Layer Meteorology*, 84, 219–230, 1991.
- Wharton, S., M. Simpson, J. Osuna, J. Newman and S. Biraud**, Assessment of Land Surface Model Performance in WRF for Simulating Wind at Heights Relevant to the Wind Energy Community, *Lawrence Livermore National Laboratory, United States Department of Energy*, 52, pp. 35, 2013.
- Wilks, D. S.**, Effects of stochastic parameterizations in the Lorenz '96 system, *Quarterly Journal of the Royal Meteorological Society*, 131, 389–407, 2005.
- Yang, Q., L. K. Berg, M. Pekour, J. D. Fast, R. K. Newsom, M. Stoelinga and C. Finley**, Evaluation of WRF-Predicted Near-Hub-Height Winds and Ramp Events over a Pacific Northwest Site with Complex Terrain, *Journal of Applied Meteorology and Climatology*, 52, 1753–1763, 2013.
- Zolotarev, P. and M. Gökeler**, Netzregelverbund - Koordinierter Einsatz von Sekundärregelleistung, *Proceedings of the 10. ETG/GMM-Fachtagung "Netzregelung und Systemführung, Munich, Germany*, 2011.

Acknowledgements

The author gratefully acknowledges the Gauss Centre for Supercomputing (GCS) for providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS share of the supercomputer JUQUEEN at Jülich Supercomputing Centre (JSC). GCS is the alliance of the three national supercomputing centres HLRS (Universität Stuttgart), JSC (Forschungszentrum Jülich), and LRZ (Bayerische Akademie der Wissenschaften), funded by the German Federal Ministry of Education and Research (BMBF) and the German State Ministries for Research of Baden-Württemberg (MWK), Bayern (StMWFK) and Nordrhein-Westfalen (MIWF).

Onshore meteorological measurements at the measurement masts Cabauw, Karlsruhe, Falkenberg, Hamburg and Jülich were provided by the Royal Netherlands Meteorological Institute (KNMI), the German Weather Service (DWD), the Karlsruhe Institute of Technology (KIT), the Meteorological Institute (MI) of the University of Hamburg, and the Forschungszentrum Jülich (FZJ), for which the author is very grateful. The Bundesministerium für Wirtschaft und Energie (BMWi), the Bundesamt für Seeschifffahrt und Hydrographie (BSH), and the Projektträger Jülich (PTJ) are acknowledged for providing measurements of the offshore research platforms Fino1, Fino2, and Fino3.

The author wishes to acknowledge the Fraunhofer-Institut für Windenergie und Energiesystemtechnik for carrying out the wind power calculations.

This work was supported by the Energy oriented Centre of Excellence (EoCoE), grant agreement number 676629, funded within the Horizon2020 framework of the European Union.

Persönliche Danksagung

Allen, die mich bei dieser Arbeit unterstützt haben, möchte ich meinen Dank aussprechen.

In erster Linie gilt mein Dank meinem Doktorvater Dr. Hendrik Elbern, für seine persönliche Art der Betreuung, seinen stetigen Zuspruch und sein Vertrauen in meine Arbeit. Durch die von ihm ermöglichte Einblicke in den Wissenschaftsbetrieb fühlte ich mich auch über meine Doktorarbeit hinaus gefördert. Seine ideenreiche wissenschaftliche Denkweise hat mich beeindruckt und ich hoffe, für meinen weiteren Weg etwas davon mitnehmen zu können.

Prof. Dr. Yaping Shao danke ich herzlich für das Erstellen des Zweitgutachtens. Ich hoffe, das Lesen hat neben dem Aufwand auch sein fachliches Interesse wecken können.

Dem Institut IEK-8, Forschungszentrum Jülich, danke ich für die Möglichkeit diese Arbeit anfertigen zu können. Danke an Frau Prof. Dr. Astrid Kiendler-Scharr und Herrn Prof. Dr. Andreas Wahner.

Ein besonderer Dank geht an Garrett Good vom Fraunhofer IWES, der stets geduldig Energievorhersagen produziert hat und damit zum Gelingen dieser Arbeit maßgeblich beigetragen hat.

Sebastian Lührs danke ich für seine technische Hilfe und ständige Bereitschaft, lästige Fragen ausführlich zu beantworten.

Prof. Dr. Roel Neggers sei für die hilfreiche Diskussion über Grenzschicht Parameterisierung gedankt.

Ohne Systemadministrator wären wohl schon so einige Arbeiten gescheitert, deshalb einen großen Dank an Georg Piekorz.

Allen Mitarbeitern im Rheinischen Institut für Umweltforschung möchte ich danken. Insbesondere sollen erwähnt sein: der harte Kern der Mensa Gruppe, Isabel Ribeiro und Johannes Klimpt; Philipp Franke für Diskussionen rund um das Filtern; Kai Krajsek für das Bearbeiten großer Datenmengen; sowie Elmar Frieze für geduldiges Herunterladen von Modelldaten. Anne Caroline Lange danke ich sehr für ihre

Unterstützung und Aufmunterung in der gemeinsamen Endphase der Promotion.

Der letzte Dank geht an meine Familie, für ihre Unterstützung und Geduld. Danke an meine Eltern, meinen Bruder, Ketu und Janina.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von PD Dr. Hendrik Elbern betreut worden.

Köln, den 05.12.2018

Jonas Berndt

Lebenslauf

Jonas Berndt

Persönliche Daten

Adresse Petersbergstraße 88
50939 Köln
Deutschland

Telefon +49 (0)221 / 16821889
E-Mail jonas.berndt@gmx.de

Geburtsdatum 06.03.1985
Geburtsort Köln
Staatsangehörigkeit deutsch

Schulausbildung

08.1991–06.1995 Gemeinschaftsgrundschule Köln Longerich, Köln
08.1995–06.2004 Erich-Kästner Gymnasium, Köln
Schulabschluss: Allgemeine Hochschulreife

Studium

10.2005–01.2006 Diplomstudiengang Umweltwissenschaften an der Universität Trier

10.2006–12.2009 Bachelorstudium in Geophysik und Meteorologie an der Universität zu Köln
Abschluss: Bachelor of Science

10.2009–08.2013 Masterstudium in Geophysik an der Universität Münster
Abschluss: Master of Science

seit 12.2013 Promotionsstudium an der Universität zu Köln

Berufstätigkeit

12.2008–10.2009 Studentische Hilfskraft am Rheinischen Institut für Umweltforschung an der Universität zu Köln

10.2011–10.2012 Studentische Hilfskraft am Institut für Geophysik an der Universität Münster

seit 12.2013 Wissenschaftlicher Mitarbeiter (Doktorand) am Institut für
Energie und Klimaforschung – Troposphäre (IEK-8) am
Forschungszentrum Jülich