Improving hub-height wind forecasts in complex terrain

by

David Siuta

B.S. Earth System Science, University of Wyoming, 2011

M.S. Atmospheric Science, University of Wyoming, 2013

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Atmospheric Science)

The University of British Columbia

(Vancouver)

March 2017

Abstract

Wind-speed forecasts from numerical-weather-prediction (NWP) models are important for daily wind-resource generation planning. However, NWP models are imperfect. The ability of energy planners to efficiently manage resources is a function of the accuracy of deterministic wind forecasts and of the associated probability estimates of forecast uncertainty. As the amount of energy generated from wind increases to significant levels, improving forecast accuracy and representation of forecast uncertainty is a key area of active research.

This dissertation advances wind forecasting over regions of complex topography using the Weather Research and Forecasting (WRF) model. The optimal WRF-model configuration is a function of planetary-boundary-layer (PBL) physics, grid length, and initial-condition choice. The first component of this work determines which of these three factors most influences forecast accuracy over complex terrain. The two largest factors influencing forecast accuracy are the PBL-physics scheme and the grid length, with the dominant factor being a function of location, season, and time of day.

The second component of the research addresses the need for probability-based forecast information, which is only recently being used within the industry. Wind forecasts from an ensemble using eight PBL schemes, three grid lengths, and two initial-conditions sources are converted into probability models that are then evaluated. Using the full, empirical ensemble distribution produces uncalibrated probabilistic forecasts. Prescribing a Gaussian probability distribution based on statistical moments of a past training dataset results in calibrated and sharp probabilistic forecasts. Such a method is also computationally cheap.

The final aspect of this study evaluates the role of boundary-layer static stability on forecast performance. Traditionally, empirical surface-layer similarity theory has been used to relate surface fluxes of heat, momentum, and moisture to vertical profiles of temperature and wind. To evaluate and improve surface-layer similarity theories over mountain ridges, a year-long field campaign of temperature and wind measurements was conducted at wind farms in British Columbia. New empirical equations for complex terrain are proposed based on the field data, and found to perform well at an independent test location.

Preface

This dissertation is composed of the original work of the author from three accepted or submitted journal articles, with minimal editing changes. These articles provide the contents of Chapters 2, 3, and 4, and are described in more detail below. The introduction and conclusion chapters, however, are unique to this dissertation.

Chapter 2

D. Siuta, 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,* in press. doi: 10.1175/WAF-D-16-0120.1. Copyright 2017 American Meteorological Society.

The idea to evaluate the performance of planetary-boundary-layer physics scheme choice in the WRF model (Chapter 2) was developed by D. Siuta with project funding provided by Dr. Roland Stull. Additionally, Dr. Greg West collaborated with D. Siuta and Dr. Stull throughout the project. D. Siuta carried out the research, analyzed the results, and wrote the submitted manuscript. Dr. Stull provided the computational resources for such an extensive study and also edited the contents of the manuscript. Dr. West provided background knowledge on the use of wind forecasts for wind resource planning at BC Hydro and edited the manuscript.

Chapter 3

D. Siuta, G. West, R. Stull, and T. Nipen. Calibrated probabilistic hub-height wind forecasts in complex terrain. *Weather and Forecasting,* in press. doi: 10.1175/WAF-D-16-0137.1. Copyright 2017 American Meteorological Society.

Drs. Greg West and Roland Stull identified the need to develop calibrated probabilistic hub-height wind-speed forecasts in locations of complex terrain (Chapter 3). D. Siuta thought of the idea for an ensemble forecast based on multiple planetary-boundary-layer schemes, grid lengths, and initial-conditions choices, and carried out the research, analysis, and authoring of the manuscript. D. Siuta collaborated with Dr. Thomas Nipen to use the Component-based Post-Processing System (COMPS) to carry out parts of the research. Dr. Nipen wrote the code used in the COMPS system and provided edits to the manuscript. Drs. West and Stull also provided edits to the manuscript. Dr. Stull wrote the appendix for the manuscript (Appendix B in this dissertation) and devised the new pq distribution.

Chapter 4

D. Siuta, R. Howard, and R. Stull. A flux-profile relationship for winds over mountain ridges. Submitted.

The idea to perform fieldwork to measure boundary-layer stability was by D. Siuta, who also carried out the field campaign. Dr. Stull provided input into fieldwork, instrument purchase, and calibration. Additional help in instrumentation, methodology, and editing was provided from Dr. Rosie Howard. Drs. Stull and Howard also provided some expertise in how to interpret the workings of the PBL.

Table of Contents

List of Tables

List of Figures

Glossary

Symbols

Below is a summary of all mathematical variables used in the dissertation:

A	Area under the un-normalized pq density curve
c_p	Specific heat capacity of air at constant pressure
C_D	Bulk drag coefficient for momentum
C_H	Bulk drag coefficient for heat
d	Zero-plane displacement distance
DMB_{t}	Degree of mass balance bias correction factor for the current day (t)
$DMB_{\text{t-1}}$	Degree of mass balance bias correction factor at time from previous day (t-1)
f_s	Functional form of pq frequency distribution
$f(x)$	Functional form of the pq probability density
F_{t}	Raw ensemble-member forecast
$F_{\text{t-1}}$	Mean of the previous day's forecasts
$\bar{F}_t$	Ensemble mean forecast
$\hat{F}_t$	Bias-corrected ensemble-member forecast
h	Canopy height
k	von Karman constant
L	Obukhov length
$MAESS_{BD}$	Mean absolute error skill score of the SHS relationship relative to the BD relationship
$MAESS_{ZA}$	Mean absolute error skill score of the SHS relationship relative to the ZA relationship
$\bar{O}$	Mean value of the observed wind profile
$O_{\text{t-1}}$	Mean of the previous day's observations
$\mathcal{N}_t$	Gaussian probability distribution for the current forecast time
p	Shape parameter in pq distribution
q	Shape parameter in pq distribution
r	Linear correlation coefficient
r^2	Coefficient of determination
RH	Relative humidity

Ri_b	Bulk Richardson number
S	Scaling parameter in pq distribution
t	time
x	Either the ensemble mean or ensemble variance within a linear regression against the forecast errors
x_o	Center-location parameter in pq distribution
u	Wind speed
u_*	Friction velocity
U_{BD}	Average wind speed produced from the BD flux-profile
U_{SHS}	Average wind speed produced from the SHS flux-profile
U_{ZA}	Average wind speed produced from the ZA flux-profile
z	Height above ground level
z_o	Aerodynamic roughness length
z_*	Depth of the transition layer (roughness sublayer)
α	Empirical constant in MOST
β	Empirical constant in MOST
γ	Empirical constant in MOST
Γ	The gamma function
μ_4	Fourth statistical moment
ϕ_h	Non-dimensional temperature gradient
ϕ_m	Non-dimensional wind shear
$\phi_{m,BD}$	Non-dimensional wind shear diagnosed from the BD relations
$\phi_{m,SHS}$	Non-dimensional wind shear diagnosed from the BD relations
ϕ_*	Stability-independent correction factor for non-dimensional wind shear
ψ_h	Integrated similarity function for heat
ψ_m	Integrated similarity function for momentum
ρ	Air density
σ_t^2	Variance of prescribed probability distribution at time t calculated from the past error distribution
σ_x^2	Variance of pq probability distribution
τ	Training period (Chapter 3), Momentum stress (Chapter 4)
θ	Potential temperature
θ_g	Potential temperature at the ground
θ_v	Virtual potential temperature
$\theta_v g$	Virtual potential temperature at the ground

θ_*	Temperature scale
ζ	Monin-Obukhov stability parameter

Abbreviations

Below is a summary of all abbreviations used in the dissertations.

ACM2	Asymmetric convection model version 2 scheme
ACM2GFS4	WRF 4-km grid output using the ACM2 PBL scheme initialized by the GFS
ACM2GFS12	WRF 12-km grid output using the ACM2 PBL scheme initialized by the GFS
ACM2GFS36	WRF 36-km grid output using the ACM2 PBL scheme initialized by the GFS
ACM2NAM4	WRF 4-km grid output using the ACM2 PBL scheme initialized by the NAM
ACM2NAM12	WRF 12-km grid output using the ACM2 PBL scheme initialized by the NAM
ACM2NAM36	WRF 36-km grid output using the ACM2 PBL scheme initialized by the NAM
AGL	Above ground level
ANOVA	Analysis of Variance
ARW	Advanced Research WRF dynamical core
ASL	Atmospheric Surface Layer
BD	Businger-Dyer flux-profile relationships
BMA	Bayesian model averaging
CASES99	Cooperative Atmospheric Surfaces-Exchange Study 1999
COMPS	Component-based post-processing system
CRMSE	Centered Root Mean Square Error
CRPS	Continuous ranked probability score
DMB	Degree of Mass Balance
DWD	Deutscher Wetterdienst (German Weather Service)
ECMWF	European Centre for Medium-range Weather Forecasts
EMOS	Ensemble model output statistics
EPS	Ensemble prediction system
G1	Ensemble using 1 PBL scheme and prescribed GNS distribution around raw-ensemble mean
G2	Ensemble using 2 PBL schemes and prescribed GNS distribution around raw-ensemble mean
G3	Ensemble using 3 PBL schemes and prescribed GNS distribution around raw-ensemble mean

G4	Ensemble using 4 PBL schemes and prescribed GNS distribution around raw-ensemble mean
G5	Ensemble using 5 PBL schemes and prescribed GNS distribution around raw-ensemble mean
G6	Ensemble using 6 PBL schemes and prescribed GNS distribution around raw-ensemble mean
G7	Ensemble using 7 PBL schemes and prescribed GNS distribution around raw-ensemble mean
G8	Ensemble using 8 PBL schemes and prescribed GNS distribution around raw-ensemble mean
GB1	Ensemble using 1 PBL scheme and prescribed GNS distribution around bias-corrected ensemble mean
GB2	Ensemble using 2 PBL schemes and prescribed GNS distribution around bias-corrected ensemble mean
GB3	Ensemble using 3 PBL schemes and prescribed GNS distribution around bias-corrected ensemble mean
GB4	Ensemble using 4 PBL schemes and prescribed GNS distribution around bias-corrected ensemble mean
GB5	Ensemble using 5 PBL schemes and prescribed GNS distribution around bias-corrected ensemble mean
GB6	Ensemble using 6 PBL schemes and prescribed GNS distribution around bias-corrected ensemble mean
GB7	Ensemble using 7 PBL schemes and prescribed GNS distribution around bias-corrected ensemble mean
GB8	Ensemble using 8 PBL schemes and prescribed GNS distribution around bias-corrected ensemble mean
GBM	Grenier-Bretherton-Park scheme
GBMGFS4	WRF 4-km grid output using the GBM PBL scheme initialized by the GFS
GBMGFS12	WRF 12-km grid output using the GBM PBL scheme initialized by the GFS
GBMGFS36	WRF 36-km grid output using the GBM PBL scheme initialized by the GFS
GBMNAM4	WRF 4-km grid output using the GBM PBL scheme initialized by the NAM
GBMNAM12	WRF 12-km grid output using the GBM PBL scheme initialized by the NAM
GBMNAM36	WRF 36-km grid output using the GBM PBL scheme initialized by the NAM
GDPS	Global Deterministic Prediction System, also known as the GEM
GFS	Global forecast system

GL	Grid length
GNS	Gaussian non-scaling distribution
GSEM	Gaussian distribution scaled by ensemble mean
GSEV	Gaussian distribution scaled by ensemble variance
GSI	Gridpoint Statistical Interpolation
GSM1	Ensemble using 1 PBL scheme and prescribed GSEM distribution around raw-ensemble mean
GSM2	Ensemble using 2 PBL schemes and prescribed GSEM distribution around raw-ensemble mean
GSM3	Ensemble using 3 PBL schemes and prescribed GSEM distribution around raw-ensemble mean
GSM4	Ensemble using 4 PBL schemes and prescribed GSEM distribution around raw-ensemble mean
GSM5	Ensemble using 5 PBL schemes and prescribed GSEM distribution around raw-ensemble mean
GSM6	Ensemble using 6 PBL schemes and prescribed GSEM distribution around raw-ensemble mean
GSM7	Ensemble using 7 PBL schemes and prescribed GSEM distribution around raw-ensemble mean
GSM8	Ensemble using 8 PBL schemes and prescribed GSEM distribution around raw-ensemble mean
GSMB1	Ensemble using 1 PBL scheme and prescribed GSEM distribution around bias-corrected ensemble mean
GSMB2	Ensemble using 2 PBL schemes and prescribed GSEM distribution around bias-corrected ensemble mean
GSMB3	Ensemble using 3 PBL schemes and prescribed GSEM distribution around bias-corrected ensemble mean
GSMB4	Ensemble using 4 PBL schemes and prescribed GSEM distribution around bias-corrected ensemble mean
GSMB5	Ensemble using 5 PBL schemes and prescribed GSEM distribution around bias-corrected ensemble mean
GSMB6	Ensemble using 6 PBL schemes and prescribed GSEM distribution around bias-corrected ensemble mean
GSMB7	Ensemble using 7 PBL schemes and prescribed GSEM distribution around bias-corrected ensemble mean

GSMB8	Ensemble using 8 PBL schemes and prescribed GSEM distribution around bias-corrected ensemble mean
GSV1	Ensemble using 1 PBL scheme and prescribed GSEV distribution around raw-ensemble mean
GSV2	Ensemble using 2 PBL schemes and prescribed GSEV distribution around raw-ensemble mean
GSV3	Ensemble using 3 PBL schemes and prescribed GSEV distribution around raw-ensemble mean
GSV4	Ensemble using 4 PBL schemes and prescribed GSEV distribution around raw-ensemble mean
GSV5	Ensemble using 5 PBL schemes and prescribed GSEV distribution around raw-ensemble mean
GSV6	Ensemble using 6 PBL schemes and prescribed GSEV distribution around raw-ensemble mean
GSV7	Ensemble using 7 PBL schemes and prescribed GSEV distribution around raw-ensemble mean
GSV8	Ensemble using 8 PBL schemes and prescribed GSEV distribution around raw-ensemble mean
GSVB1	Ensemble using 1 PBL scheme and prescribed GSEV distribution around bias-corrected ensemble mean
GSVB2	Ensemble using 2 PBL schemes and prescribed GSEV distribution around bias-corrected ensemble mean
GSVB3	Ensemble using 3 PBL schemes and prescribed GSEV distribution around bias-corrected ensemble mean
GSVB4	Ensemble using 4 PBL schemes and prescribed GSEV distribution around bias-corrected ensemble mean
GSVB5	Ensemble using 5 PBL schemes and prescribed GSEV distribution around bias-corrected ensemble mean
GSVB6	Ensemble using 6 PBL schemes and prescribed GSEV distribution around bias-corrected ensemble mean
GSVB7	Ensemble using 7 PBL schemes and prescribed GSEV distribution around bias-corrected ensemble mean
GSVB8	Ensemble using 8 PBL schemes and prescribed GSEV distribution around bias-corrected ensemble mean
IC	Initial condition

IFS	Integrated Forecast System
LES	Large Eddy Simulation
LLJ	Low-level jet
MAD	Mean absolute deviation
MAE	Mean absolute error
MAESS	Mean absolute error skill score
MM5	5th Generation of the Pennsylvania State University-National Center for Atmospheric Research Mesoscale Model
MOST	Monin-Obukhov Similarity Theory
MRF	Medium range forecast scheme
MRFGFS4	WRF 4-km grid output using the MRF PBL scheme initialized by the GFS
MRFGFS12	WRF 12-km grid output using the MRF PBL scheme initialized by the GFS
MRFGFS36	WRF 36-km grid output using the MRF PBL scheme initialized by the GFS
MRFNAM4	WRF 4-km grid output using the MRF PBL scheme initialized by the NAM
MRFNAM12	WRF 12-km grid output using the MRF PBL scheme initialized by the NAM
MRFNAM36	WRF 36-km grid output using the MRF PBL scheme initialized by the NAM
MSL	Elevation above mean sea level
MYJ	Mellor-Yamada-Janjic scheme
MYJGFS4	WRF 4-km grid output using the MYJ PBL scheme initialized by the GFS
MYJGFS12	WRF 12-km grid output using the MYJ PBL scheme initialized by the GFS
MYJGFS36	WRF 36-km grid output using the MYJ PBL scheme initialized by the GFS
MYJNAM4	WRF 4-km grid output using the MYJ PBL scheme initialized by the NAM
MYJNAM12	WRF 12-km grid output using the MYJ PBL scheme initialized by the NAM
MYJNAM36	WRF 36-km grid output using the MYJ PBL scheme initialized by the NAM
MYNN	Mellor-Yamade-Nakinishi-Nino level 2 scheme
MYNNGFS4	WRF 4-km grid output using the MYNN PBL scheme initialized by the GFS
MYNNGFS12	WRF 12-km grid output using the MYNN PBL scheme initialized by the GFS
MYNNGFS36	WRF 36-km grid output using the MYNN PBL scheme initialized by the GFS
MYNNNAM4	WRF 4-km grid output using the MYNN PBL scheme initialized by the NAM
MYNNNAM12	WRF 12-km grid output using the MYNN PBL scheme initialized by the NAM
MYNNNAM36	WRF 36-km grid output using the MYNN PBL scheme initialized by the NAM
NAM	North American mesoscale model
NAVGEM	Navy Global Environmental Model
NCEP	National centers for environmental prediction
NMM	Nonhydrostatic Mesoscale Model dynamical core

NWP	Numerical weather prediction
PBL	Planetary boundary layer
PIT	Probability integral transform
pq	pq distribution
R1	Ensemble using 1 PBL scheme and empirical raw-ensemble distribution
R2	Ensemble using 2 PBL schemes and empirical raw-ensemble distribution
R3	Ensemble using 3 PBL schemes and empirical raw-ensemble distribution
R4	Ensemble using 4 PBL schemes and empirical raw-ensemble distribution
R5	Ensemble using 5 PBL schemes and empirical raw-ensemble distribution
R6	Ensemble using 6 PBL schemes and empirical raw-ensemble distribution
R7	Ensemble using 7 PBL schemes and empirical raw-ensemble distribution
R8	Ensemble using 8 PBL schemes and empirical raw-ensemble distribution
RB1	Ensemble using 1 PBL scheme and empirical bias-corrected ensemble distribution
RB2	Ensemble using 2 PBL schemes and empirical bias-corrected ensemble distribution
RB3	Ensemble using 3 PBL schemes and empirical bias-corrected ensemble distribution
RB4	Ensemble using 4 PBL schemes and empirical bias-corrected ensemble distribution
RB5	Ensemble using 5 PBL schemes and empirical bias-corrected ensemble distribution
RB6	Ensemble using 6 PBL schemes and empirical bias-corrected ensemble distribution
RB7	Ensemble using 7 PBL schemes and empirical bias-corrected ensemble distribution
RB8	Ensemble using 8 PBL schemes and empirical bias-corrected ensemble distribution
RMSE	Root mean square error
QNSE	Quasi-normal scale elimination scheme
QNSEGFS4	WRF 4-km grid output using the QNSE PBL scheme initialized by the GFS
QNSEGFS12	WRF 12-km grid output using the QNSE PBL scheme initialized by the GFS
QNSEGFS36	WRF 36-km grid output using the QNSE PBL scheme initialized by the GFS
QNSENAM4	WRF 4-km grid output using the QNSE PBL scheme initialized by the NAM
QNSENAM12	WRF 12-km grid output using the QNSE PBL scheme initialized by the NAM
QNSENAM36	WRF 36-km grid output using the QNSE PBL scheme initialized by the NAM
SHS	Siuta-Howard-Stull flux-profile relationship
SKEBS	Stochastic kinetic energy backscatter
SPPT	Stochastic perturbation of physical tendencies
TKE	Turbulence Kinetic Energy
UBC	University of British Columbia
UKMO	UK Metoffice Unified Model
UTC	Coordinated Universal Time

UW	University of Washington scheme
UWGFS4	WRF 4-km grid output using the UW PBL scheme initialized by the GFS
UWGFS12	WRF 12-km grid output using the UW PBL scheme initialized by the GFS
UWGFS36	WRF 36-km grid output using the UW PBL scheme initialized by the GFS
UWNAM4	WRF 4-km grid output using the UW PBL scheme initialized by the NAM
UWNAM12	WRF 12-km grid output using the UW PBL scheme initialized by the NAM
UWNAM36	WRF 36-km grid output using the UW PBL scheme initialized by the NAM
WFIP2	Wind Forecast Improvement Project 2
WFRT	Weather forecast research team
WRF	Weather research and forecasting model
YSU	Yonsei University Scheme
YSUGFS4	WRF 4-km grid output using the YSU PBL scheme initialized by the GFS
YSUGFS12	WRF 12-km grid output using the YSU PBL scheme initialized by the GFS
YSUGFS36	WRF 36-km grid output using the YSU PBL scheme initialized by the GFS
YSUNAM4	WRF 4-km grid output using the YSU PBL scheme initialized by the NAM
YSUNAM12	WRF 12-km grid output using the YSU PBL scheme initialized by the NAM
YSUNAM36	WRF 36-km grid output using the YSU PBL scheme initialized by the NAM
ZA	Zhang and Anthes flux-profile relationships

Acknowledgments

I must first thank my research supervisor, Dr. Roland Stull, who provided me with endless motivation and encouragement over the last four years. His tremendous support and years of devotion to research in atmospheric science provided me with a great environment for my studies, for which I am forever indebted. I also wish to thank the other members of my supervisory committee, Drs. Phil Austin, Doug McCollor, and Magdalena Rucker, who gave input throughout my tenure as a student at UBC and provided me with insights into the use of weather forecasts for wind energy as well as access to key observational data.

I am truly grateful for the help of many of the members of the Weather Forecast Research Team, without which my successes would not have been possible. I must specifically thank Dr. Greg West who spent many late nights with me going over manuscript edits, traveled across British Columbia to help with my field work, and whose conversation helped spur many of the ideas presented here. Dr. Rosie Howard, who provided me with key guidance on the setup of my field campaign and instrumentation, is also greatly acknowledged. I appreciate team members Dr. Dominique Bourdin, Pedro Odon, Maggie Campbell, Jesse Mason, Tim Chui, and Nadya Moisseeva for always allowing me to stop by their desks to chat about my work. Former team member Dr. Thomas Nipen is acknowledged for spending several of his evenings in Norway chatting with me online about creating probabilistic forecasts and statistical verification methods. Dr. Henryk Modzelewski and Roland Schigas made such an extensive modeling study possible through excellent assistance with computing resources and coding.

I thank Pedro Odon, Tyler Gifford, and my father, Albert Siuta for their willingness to spend several days with me in remote sections of British Columbia to set up, retrieve data from, and disassemble my field campaign. They stood outside for many hours in sub-freezing temperatures, snowfall, and wind-driven rains for the betterment of my research.

BC Hydro, the Canadian Natural Science and Engineering Research Council, and Mitacs are acknowledged for providing the funding to make this research possible. Energia (formerly GDF SUEZ North America), Alterra Power Corp, Capital Power, and Altagas Ltd are thanked for allowing me access to their sites to install instrumentation and meteorological data.

Finally, I thank my parents, Albert and Annmarie, and my sister Rose for being sources of inspiration through the last four years.

Chapter 1

Introduction

The research goal of this dissertation is to improve wind-speed forecasts at wind-turbine hub (axle) height (80 - 100 m) in mountainous terrain. The method is numerical weather prediction (NWP), where an ensemble of multiple NWP-model configurations is run to provide different estimates of forecast wind speed every hour. This work is motivated by the fact that utility companies need to anticipate how much contingency energy-reserve generation is needed to fill in when the winds decrease. Small errors in anticipated wind speed can cost utility companies thousands of dollars each day. The procedure is first to do sensitivity studies to find the optimum NWP-model configuration for the ensemble members. The result is used to give an ensemble-average (best deterministic) wind-speed forecast and a probability distribution of likely wind-speed errors. A field study is also conducted at four wind farms in British Columbia, Canada, to relate hub-height wind speed to static stability near the ground to identify the best NWP model options. The resulting findings present significant innovations in wind-profile estimation and NWP hub-height wind-speed forecasting for complex mountainous terrain. Details of the goals, method, motivation, procedure, and significance are explained below.

1.1 Use of Wind Forecasts for Wind Energy

Energy planners rely on wind forecasts from NWP models to anticipate electrical power generated from renewable wind resources. These estimates are used for two primary purposes: (1) scheduling of spinning (online) and stand-by contingency energy reserves (DeMeo et al., 2007; Corbus et al., 2009), and (2) market trading (Draxl et al., 2014). Contingency reserves are other energy resources that are available to dispatch energy to the electric transmission grid to ensure reliable service in the event that wind generation is less than expected, or if other maintenance issue arises (DeMeo et al., 2007). Namely, they add flexibility to the grid to accommodate natural wind variability and forecast uncertainty.

Accurate forecasts of hub-height wind speeds are critical for efficient reserve planning because they determine how much reserve capacity is needed. While contingency reserves serve important roles in maintaining a reliable electric grid, they operate at a significant cost to utility companies

and scheduling too many reserves can result in missed market trading opportunities (Clement, 2012). Additionally, these reserves require 'spin-up' time that can range from hours to over a day, which is one reason why wind forecasts are needed for short-term and long-term forecast horizons (Ahlstrom et al., 2013).

For energy-market trading, analysts need to know how much excess electricity will be available to sell to adjacent utility districts, or, if there will be a need to import energy. Trading also occurs at both short and long-term intervals, with bids updated as short as half an hour prior to delivery deadlines (Draxl et al., 2014). In some cases, traders can sell wind-generated electricity at a premium to adjacent areas needing to meet green-energy quotas. Penalties can be incurred for not being able to deliver the bid power amount (Monteiro et al., 2009).

As the amount of the total electric supply generated from wind resources increases, the economic value of accurate forecasts from NWP models also increases (Corbus et al., 2009). Marquis et al. (2011) detail that small reductions in the mean absolute error (MAE) of hub-height wind forecasts on the order of 10-20% result in millions of dollars of savings. They find that there is between $1.6 and $4.1 billion dollars of savings that can be achieved by improving current state-of-the-art forecast systems to a 'perfect forecast', assuming approximately 20% wind penetration (the portion of electrical generation produced by wind). The United States aims to reach this level of wind penetration by 2030, with a further increase to 35% wind penetration by 2050 (United States Department of Energy, 2015). Canada has a similar goal of 35% wind penetration by 2025 (Canadian Wind Energy Association, 2016). Marquis et al. (2011) also mention that additional savings could be achieved in energy planning by incorporating measures of forecast uncertainty.

Wind forecasts from NWP models are imperfect, leading to uncertainty in expected wind generation estimates. Such uncertainty is not unique to wind, and arises from several NWP model deficiencies. Errors arise from incorrect/incomplete depictions of atmospheric state during data assimilation, incomplete understanding of atmospheric physics, and because of spatial discretization of atmospheric flows over complex terrain (Warner, 2011; Candille, 2009; Mass et al., 2002). These factors have led to the use of ensemble-forecasting techniques, which quantify forecast uncertainty by running several NWP models defined by different initial conditions, physics parameterizations, and grid lengths. Since the processes describing the evolution of the atmosphere are non-linear, these small model differences cause the forecast differences between ensemble members to grow with time. Thus, the overall spread of the ensemble members provides an indication of forecast uncertainty (Warner, 2011), with larger spread implying greater uncertainty.

Over areas of complex (mountainous) terrain, uncertainties may be amplified by sparse observations, by the use of inappropriate physical parameterizations that have been derived over homogeneous flat terrain, and through poor representation of topography. Together, these question the

ability of NWP models to accurately represent typical flows that arise in complex topography, such as thermally-induced circulations, dynamically-forced flow channeling, and mountain-wave breaking. For wind-energy sources located on ridge tops and in other complex terrain, this could result in less certain forecasts.

Nonetheless, locations of complex terrain often provide some of the richest wind resources. Building wind farms in mountainous areas, however, has been limited by poor access (limited or non-existent roads) and lack of nearby electric-transmission infrastructure. Thus, studies on forecasting wind-turbine hub-height winds in areas of complex terrain have been limited. The Wind Forecast Improvement Project 2 (WFIP2) recently began a large field campaign in the United States to address this knowledge gap (see http://www.esrl.noaa.gov/gsd/renewable/wfip2.html).

There are three goals of this dissertation. The first is to identify the sensitivity of deterministic hub-height wind-forecast accuracy to the selection of three Weather Research and Forecasting (WRF) model configurations: (1) the initial-condition (IC) source, (2) the grid length (GL), and (3) the planetary-boundary-layer (PBL) scheme. The second goal is to evaluate the performance of a multi-PBL, multi-GL, multi-IC ensemble forecast to obtain calibrated probabilistic wind forecasts in regions of complex terrain. The final goal is to determine the role of planetary-boundary-layer static stability on the performance of WRF hub-height wind forecasts. Each of these three goals are introduced in the following three subsections, with details presented in subsequent chapters.

1.2 Deterministic Hub-Height Wind Forecasting and Forecast Accuracy

Historically, vertical extrapolation methods like the 1/7 power law and logarithmic profile (Peterson and Hennessey, 1978; Stull, 1988) have been used to convert winds at 10-m wind to those expected at wind-turbine hub height. However, these methods make assumptions about atmospheric static stability and have been derived from observations in flat terrain. The 1/7 power law in particular has no physical basis for its formulation (Arya, 2001). Nonetheless, both these methods are easy to use and have stood the test of time.

Recently, forecasts from NWP models such as WRF have taken the forefront of hub-height wind-forecast research. Vertical levels in the WRF model can be fine-tuned to produce output directly at hub heights, thus reducing the dependence on surface-based extrapolation. In addition, while most NWP models used by national meteorological centers have been configured with the specific goal of improving precipitation and temperature forecasts, the WRF model has numerous physics packages that can be used to improve forecast performance with wind-energy applications in mind. Siuta (2013) found that WRF wind-forecast performance was more accurate than that of the

US National Centers for Environmental Prediction (NCEP) operational North American Mesoscale Model (NAM) at several surface locations in Wyoming. Finding the best WRF configuration for wind forecasts is an active area of research.

Since wind-turbine hub-heights are located within the planetary boundary layer (approximately the bottom 3 km of the atmosphere), a few studies have evaluated the effects of PBL-scheme choice on forecast hub-height wind speeds. Deppe et al. (2012) and Shin and Hong (2011) evaluated several PBL schemes in WRF over flat terrain, with neither study finding any particular scheme that performs best. Deppe et al. (2012) found an ensemble of PBL schemes provided higher accuracy than any individual scheme, although there was very little difference between the individual schemes. Since these early studies, options for the PBL scheme have increased, and a bug in one of the most widely used schemes has been identified and fixed (Hu et al., 2013). A recent study by Draxl et al. (2014) showed that the use of different PBL schemes affects modeled wind shear. This is important since shear-induced stress has been shown to result in damaged wind-turbine gear boxes. Additionally, the difference in vertical wind shear between PBL schemes implies that verification of wind forecasts at hub height (as opposed to verification of 10-m winds) is a necessity for wind-energy applications when using NWP models.

The studies mentioned were all conducted over flat terrain, and literature is limited with respect to studies over complex terrain. One study by Jiménez and Dudhia (2011) finds WRF to have a surface wind bias that depends on the terrain type, with winds too fast over mountain tops. Chow et al. (2013) indicate the need for studies over complex terrain as current flat-terrain model physics parameterizations may not apply. None of the PBL schemes currently available in WRF have been derived from data over mountainous terrain. The relative importance of PBL scheme, GL, and IC choice is necessary knowledge for those needing to set up wind forecasts in areas of complex terrain, and is addressed by new research described in Chapter 2.

1.3 Probabilistic Hub-Height Wind Forecasting

Research on the use of probabilistic wind forecasts in the wind-energy community is approximately 10 years old with pioneering papers written by Gneiting et al. (2005) and Pinson et al. (2006), who recognized the value of quantifying forecast uncertainty for wind-energy purposes. Namely, quantifying forecast uncertainty provides added value to energy planners through optimized resource planning and market trading abilities (Monteiro et al., 2009). Gneiting et al. (2005) and Pinson et al. (2006) describe that for probabilistic forecasts to be useful to the wind-energy community, such forecasts must primarily be probabilistically calibrated (i.e., forecast probabilities match observed frequencies), with added value by forecast systems that have sharper probabilistic spread

(i.e., narrower range of uncertainty).

Several methods exist for deriving probabilistic forecasts, including parametric and non-parametric approaches (Zhang et al., 2014). Chapter 3 focuses on the use of ensemble prediction systems (EPSs) to provide probabilistic forecasts through both parametric and non-parametric methods.

Literature on the use of EPSs and probability-based hub-height wind forecasts is limited (Junk et al., 2015). While ensemble-forecast methods have been around since the 1960s-1970s (Warner, 2011), their application to wind energy is a new area of research. For example, Wilczak et al. (2015) recently evaluated the benefits of a nine-member ensemble for improving ramp-event prediction and found an approximate 20% improvement in the ranked probability score when compared to the best deterministic forecast. However, Wilczak et al. (2015) evaluated ensemble forecasts for only 0-6 hour lead times and found quick reduction in the skill of ensemble-based forecast systems for longer forecast horizons. This result leads one to question the use of ensembles for hub-height wind forecasts at longer lead times, including those for same-day and day-ahead planning. For meteorological variables that are highly influenced by the Earth's surface, such as hub-height winds, the usefulness of ensembles has been previously questioned (Eckel and Mass, 2005; Stensrud et al., 2000).

EPSs have primarily been shown to improve deterministic forecasts through the use of the ensemble mean and for quantifying forecast uncertainty based on the spread of the ensemble members (Warner, 2011; Bakhshaii and Stull, 2009). Empirical ensemble distributions, however, are often underdispersive. This makes them a poor representation of forecast uncertainty once converted to a probability distribution (probabilistic forecast) (Warner, 2011; McCollor and Stull, 2008b; Eckel and Mass, 2005). The fact that empirical ensemble distributions tend to be underdispersive complicates the probabilistic forecast process since calibration steps must be taken. Probability distributions should fit the typical error distribution calculated around the deterministic ensemble mean. Only when forecast probabilities match the observed frequency of occurrence, is the forecast said to be calibrated (Nipen and Stull, 2011).

Two main frameworks exist for converting a distribution of ensemble members to a probability distribution representing forecast uncertainty (uncertainty model). The first method uses ensemble model output statistics (EMOS), which parameterizes a probability distribution based on error statistics of a training dataset (Gneiting et al., 2005; Nipen and Stull, 2011). The second is Bayesian Model Averaging (BMA), which fits a Gaussian distribution around each ensemble member before weighting the distributions (Courtney et al., 2013; Sloughter et al., 2010; Raftery et al., 2005). EMOS is a simpler method to implement operationally.

Junk et al. (2015) compare forecasts produced from EPSs consisting entirely of global NWP models to those consisting only of local area models. They find the global ensemble to produce

lower MAE and continuously ranked probability scores (CRPS) when using the EMOS method. They also improve ensemble performance by using a multi-model approach by combining members from the global ensemble and a local-area ensemble.

The new research described in Chapter 3 furthers probabilistic hub-height wind forecasts in two main areas. First, I evaluate the performance of a multi-PBL, multi-GL, multi-IC ensemble, which is posited to capitalize on uncertainty in the evolution of the PBL over complex terrain. Second, I evaluate the EMOS method for hub-height winds in regions of complex terrain.

1.4 The Role of Boundary-Layer Stability

WRF and other NWP models rely on empirical relationships within surface-layer and PBL schemes to parameterize the surface stress, vertical mixing, and turbulence, all of which affect wind speed as a function of height. Surface-layer schemes parameterize surface fluxes of heat, momentum, and moisture between the ground and the first model level, and provide the lower boundary conditions for the PBL scheme (Shin and Hong, 2011). PBL-schemes parameterize sub-grid-scale fluxes and vertical diffusion from the first model level up to model top. Both the surface-layer and PBL schemes are a factor in the shape of vertical wind profiles (Shin and Hong, 2011; Jiménez et al., 2012). Depending on model configurations, actual wind-turbine hub-heights can be lower or higher than the first model layer.

Surface-layer parameterizations are based on empirically-derived flux-profile relationships, with perhaps the most well-known being the Businger-Dyer relations (Businger et al., 1971; Dyer, 1974). These relationships rely on the assumptions of Monin-Obukhov Similarity Theory (MOST), and were designed to describe the shape of wind, temperature, and moisture profiles in the lowest 10% of the planetary boundary layer using experimental observations over flat homogeneous terrain. Other profiles have been derived by Paulson (1970); Webb (1970); Beljaars (1995); Zhang and Anthes (1982); Arya (2001); Santoso and Stull (2001); Optis et al. (2016). Flux-profile relationships recognize the effects of boundary-layer stability on vertical profiles through the use of stability correction factors. However, because of the assumptions made in similarity theory, and because no relationships have been derived explicitly from data over complex terrain, use of existing relationships over such locations is questionable. Chapter 4 describes new fieldwork to determine a better flux-profile relationship for two wind farms in mountainous British Columbia.

1.5 Dissertation Organization

Aside from the Introduction and Conclusion chapters (Chapters 1 and 5, respectively), this dissertation is composed of submitted or accepted peer-reviewed journal manuscripts. While these

manuscripts have been reformatted for this dissertation, the contents are the same as the submitted or published versions, with some minor edits.

Chapter 2 addresses the sensitivity of deterministic WRF hub-height wind forecasts in complex terrain to PBL scheme, GL, and IC choice. This work has been accepted for publication in *Weather and Forecasting* (Siuta et al., 2017b). Chapter 3 describes a method to obtain calibrated hub-height wind forecasts and has also been accepted for publication in *Weather and Forecasting* (Siuta et al., 2017a). Chapter 4 explores the validity of current flux-profile relationships in complex terrain and suggests new relationships that better fit observed wind profiles. The work of Chapter 4 has been submitted for peer review.

Chapter 2

WRF Hub-Height Wind Forecast Sensitivity to PBL Scheme, Grid Length, and Initial-Condition Choice in Complex Terrain

2.1 Introduction

Many methods exist to approximate hub-height wind speeds from 10-m winds. Typical methods are the 1/7 Power Law and Adiabatic Profile (logarithmic law) (Peterson and Hennessey, 1978; Stull, 1988; Arya, 2001), and Radix profile (Santoso and Stull, 1998, 2001). However, these methods are based on implicit assumptions about the boundary layer including atmospheric stability, which are not always valid. Siuta (2013) showed for the Cooperative Atmospheric Surface-Exchange Study 1999 (CASES99, Poulos et al., 2002) that the assumptions in these wind-profile laws can break down in differing boundary-layer static stability regimes, producing large error in the estimated hub-height wind.

NWP models, such as the Weather Research and Forecasting model (WRF) (Skamarock et al., 2008), have taken the forefront of wind-speed forecast research and can be used to directly forecast winds at hub-height to avoid vertical interpolation. WRF has two dynamical cores, the Nonhydrostatic Mesoscale Model (NMM) and the Advanced Research WRF (ARW). The NMM core is used by the U.S. National Centers for Environmental Prediction (NCEP) North American Mesoscale Model (NAM), while the ARW core has been primarily used by academic research. Both cores have several physics and dynamics options that can be individually selected by the user, which can influence the hub-height wind-speed forecast and lead to improvements over standard operational models from national meteorological centers. Siuta (2013) found WRF-ARW simulations over the Wyoming Wind Corridor (Martner and Marwitz, 1982) to be superior to those from the NCEP op-

erational NAM forecasts over the same area. In some cases, the difference in estimated wind power density varied by several hundred W m^{-2}, both seasonally and daily.

Several studies have recently been done to evaluate the effects of WRF planetary-boundary-layer (PBL) physics parameterization scheme choice on wind forecasts. Shin and Hong (2011) evaluated five PBL schemes for a single-day case study in the CASES99 dataset and found larger differences and higher biases in 10-m wind forecasts produced by different PBL schemes during night-time stable regimes. Deppe et al. (2012) evaluated six PBL schemes over Iowa at wind-turbine hub height and found an ensemble of schemes to outperform the deterministic forecasts from any one PBL scheme. They also found the non-local-mixing PBL schemes (detailed in the next section) to have better accuracy.

Draxl et al. (2014) found biases over flat terrain to be strongly dependent on atmospheric static stability, with the non-local schemes performing best during unstable and near-neutral conditions, and a local scheme performing best during stable conditions. Draxl et al. (2014) also highlight the strong variation in the vertical wind shear between seven PBL schemes, indicating that evaluating PBL scheme performance on the 10-m winds alone is not sufficient when trying to use forecasts for hub-height winds.

Storm et al. (2009) evaluate two PBL schemes for simulating the Great Plains nocturnal low-level jet (LLJ) at two locations, and find different positional and magnitude errors, depending on the scheme. The LLJ is often at sufficiently low altitudes to affect wind turbines. Finally, Hahmann et al. (2015) determine that hub-height wind forecasts over sea are most sensitive to the choice of PBL scheme. While these studies provide useful results to the wind energy and NWP communities, they leave knowledge gaps in areas of complex terrain.

Jiménez and Dudhia (2011) showed WRF-ARW to have a surface (10 m) wind bias dependent on the terrain type, highlighting important deficiencies in wind forecasts over complex terrain. However, given the Draxl et al. (2014) findings, hub-height winds (not 10-m winds) should be evaluated for wind-energy purposes. Yang et al. (2013) provides one such study over the Columbia River Basin for three PBL schemes, with the finding that each scheme allowed winds to be too fast during ramp events (rapid increases or decreases in wind speed) under stable conditions. Yang et al. (2016) find strong sensitivity to the adjustment of 26 parameters within the Mellor-Yamada-Nakanishi-Niino PBL scheme, but do not consider other schemes.

One of the primary problems with producing hub-height wind forecasts from NWP models over complex terrain is that none of the available PBL physics parameterization schemes (in WRF) were derived from data taken over complex terrain, which I conclude from reading the source publications cited in the next section. Instead, model physics that are responsible for evolving boundary-layer processes were formulated mainly from field data over flat terrain, with heavy focus on improving

precipitation forecasts.

Atmospheric flows in mountainous terrain are complicated by the presence of both thermally-induced circulations and dynamically forced flows (Markowski and Richardson, 2010). Thermally-induced circulations include slope flows, mountain/valley winds, and mountain convective boundary layers (De Wekker and Kossmann, 2015; Rucker et al., 2008; De Wekker et al., 2004). For example, slope-wind systems have been shown to result in turbulent mixing above the typical convective boundary-layer along mountain tops due to slope-flow convergence (De Wekker and Kossmann, 2015). Closer to the surface, the development of upslope and downslope flows impacts wind speeds and directions at wind-turbine hub heights due to differential heating (Markowski and Richardson, 2010; Rucker et al., 2008). Dynamically-forced flows affect turbine-level winds under specific synoptic conditions that support flow blockage, channeling, and mountain-top acceleration (Markowski and Richardson, 2010). Additionally, mountain-wave activity may be present, resulting in enhanced turbulence, for flow going over mountain ridges under statically-stable conditions (Stull, 1988).

Thus, the assumptions made in PBL schemes may not be appropriate for complex terrain, and their performance should be evaluated (Chow et al., 2013). This is especially true for hub-height wind forecasts where small improvements (around 10-20% improvement in mean absolute error) can lead to millions of dollars in cost savings (Marquis et al., 2011). Because of these limitations, the improvement of hub-height wind forecasts in regions of complex terrain is the goal of the Wind Forecast Improvement Project 2 (http://www.esrl.noaa.gov/gsd/renewable/wfip2.html), which recently commenced field work.

The goal of the study presented here is to quantify the sensitivity of hub-height wind-speed forecasts in complex terrain to the PBL-scheme, initial-condition, and grid-length selection in the WRF model. Answering this question will fill a knowledge gap on the use of different PBL schemes over complex terrain. It will also be of use to those who need to create WRF-model hub-height wind forecasts over mountainous areas by addressing which factors are the most important to consider during the initial stages of the modeling process (e.g., the influence of initial-condition, PBL-scheme, and grid-length selection on forecast accuracy). The rest of this chapter is organized as follows. Section 2.2 describes the methodology used, including WRF model set-up and brief descriptions of PBL schemes, initial conditions, and grid lengths tested. Section 2.3 provides results from the study period, followed by a discussion in section 2.4. Section 2.5 concludes with a summary of the findings and suggestions for future work.

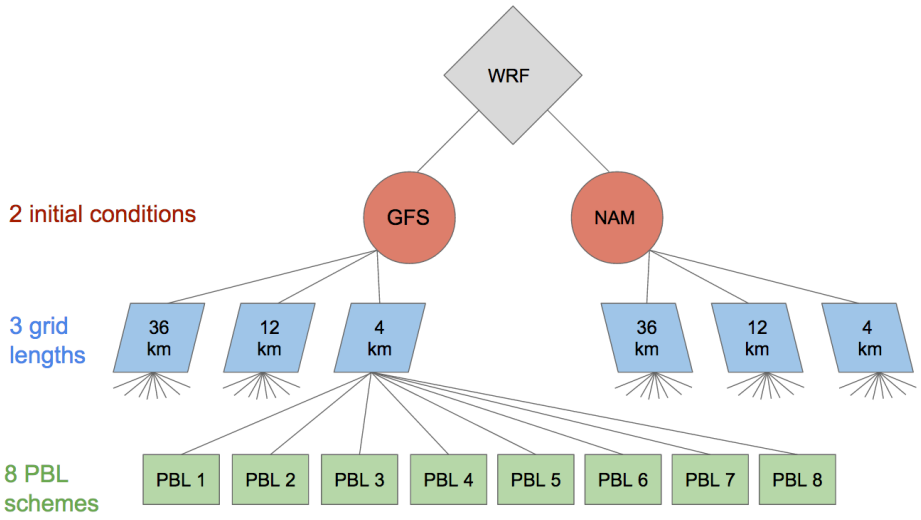

Figure 2.1: Method used to produce 48 daily forecasts at each wind farm. The names of the PBL schemes are detailed in Table 2.1.

2.2 Methodology

The WRF-ARW model was used to produce 48 forecasts per day for one year (June 2014-May 2015) by using different initial-condition sources, horizontal grid lengths, and PBL schemes (Fig. 2.1, Table 2.1). Eight available PBL schemes, detailed later in this section, were each used in sixteen daily WRF models. Eight runs were initialized by the 32-km NAM and another eight by the 0.5 degree Global Forecast System (GFS). Each of these runs contained domains of three different grid lengths, yielding the 48 total forecasts. The model forecast used four, two-way nested domains: a 36-km parent over western Canada, a 12-km nest over British Columbia and Alberta, and two 4-km nests containing wind farms of interest (Fig. 2.2). Forecasts covered the important 0-24 hour forecast horizon, after allowing for 9 hours of model spin-up. With a 24-hour forecast horizon, such forecasts would be used for same-day reserve scheduling and market trading.

Spin-up time is important for higher resolution local-area models like WRF because observation networks and relatively coarse initial conditions do not contain enough information to describe typical small scale features (e.g., tight gradients around fronts, terrain-induced local flows, and boundary-layer circulations) (Warner, 2011). Instead, WRF must develop, or *spin-up*, these features with time, taking advantage of the filtered digital elevation model in WRF. Thus, the output from the first few hours may contain unrealistic features and should be discarded. Spin-up times generally range from 6-12 hours (Warner, 2011; Skamarock, 2004), during which the output should not be used.

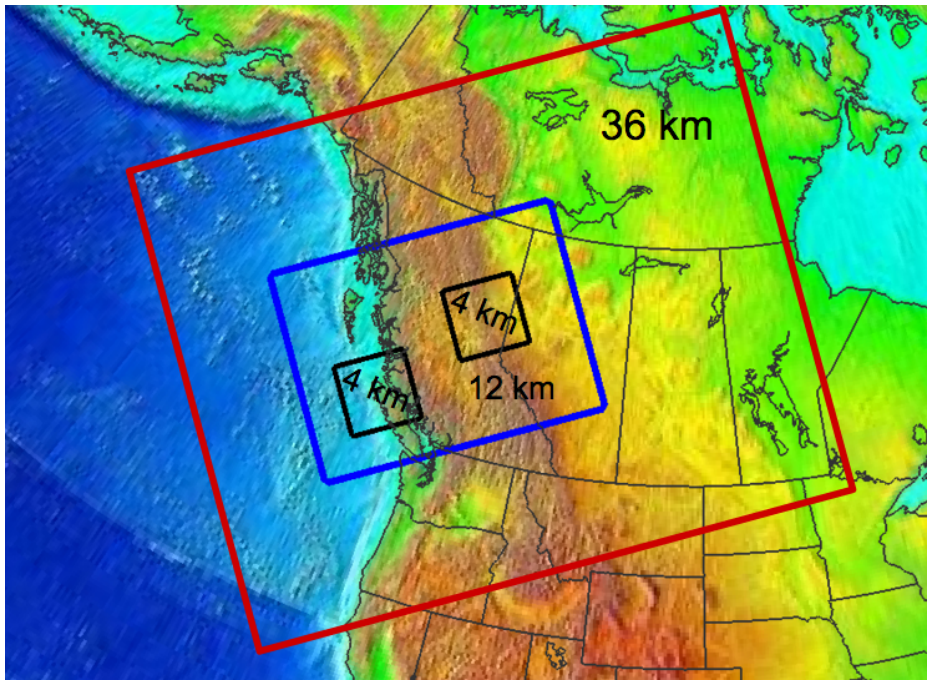

Figure 2.2: WRF domains used in this study. The 36-, 12-, and 4-km grids are bound by the red, blue, and black boxes, respectively.

Vertical levels in the model were tuned to have fine resolution around wind-turbine hub-height with 62 vertical levels total and six levels in the lowest 100 m. Forecast data was extracted without horizontal interpolation (using nearest neighbor) or vertical interpolation (using model level closest to hub-height) for each wind farm location to allow for differences to be attributed only to the PBL scheme. The authors recognize that interpolations can further improve the hub-height wind forecasts.

2.2.1 Planetary-Boundary-Layer Schemes

There are 11 PBL schemes available for use in the WRF-ARW core, version 3.5.1. Of those, eight were tested here. Three schemes were not tested (Mellor-Yamada Nakanishi and Niino Level 3, Total Energy-Mass Flux, and Boulac schemes), because they caused WRF to crash, and due to computing resource constraints.

A PBL scheme is ultimately responsible for the simulated development of the planetary boundary layer, and for sub-grid-scale fluxes of heat, momentum, and moisture by vertical turbulent transport from the first model level to model top (Yang et al., 2013; Cheng et al., 2013; Shin and Hong, 2011). Wind-turbine hub-heights and blade-spans fall within the PBL during most times of the day,

Table 2.1: WRF model configurations used for this study.

Model Detail	Setting
WRF Core	ARW Version 3.5.1
Grid Dimensions	36 km: 100x76, 12 km: 136x103, 4km: 103x103
Land Surface	Noah
Microphysics	WRF Single-Moment 5-class
Cumulus	Kain-Fritsch
Shortwave Radiation	Dudhia
Longwave Radiation	Rapid Radiative Transfer Model
Planetary Boundary Layer	Yonsei University (YSU), Medium Range Forecast (MRF), Asymmetric Convective Model (ACM2), Mellor-Yamada-Janjic (MYJ), Quasi-Normal Scale Elimination (QNSE), Mellor-Yamada Nakanishi Niino Level 2.5 (MYNN), Grenier-Bretherton-McCaa (GBM), and University of Washington (UW)
Surface Layer	MM5 similarity (used with YSU, ACM2, MRF, MYNN, GBM, and UW PBL schemes) or Eta similarity (MYJ PBL scheme only) or QNSE surface layer (QNSE PBL scheme only)

thus making the PBL scheme of crucial importance in understanding WRF strengths and deficiencies in simulating the hub-height wind. Each PBL scheme is generally compatible with at least one (but sometimes up to three) surface-layer scheme(s). The surface-layer scheme was kept constant (Table 2.1) for consistency when possible. For more information on how the PBL and surface layer are linked, see Skamarock et al. (2008). The following are brief explanations of the main components of each PBL scheme.

(i) Medium Range Forecast Model (MRF) scheme: A first-order closure, non-local eddy-diffusivity (K-theory) approach that uses a countergradient correction term based on the definition of PBL top (Hong and Pan, 1996). PBL top is defined by a critical value of the bulk Richardson number (Ri_b=0.5). The countergradient correction term accounts for contributions to the total flux by large eddies that span the boundary layer, making this a non-local scheme.

(ii) Yonsei University (YSU) scheme: A first-order closure, non-local eddy-diffusivity (K-theory)

approach with explicit treatment of the PBL top (Hong et al., 2006). This scheme provides an updated countergradient flux term originally provided in the MRF scheme. Here, the PBL top is determined by the level at which minimum turbulent flux exists (heat, momentum, moisture). The YSU scheme was updated in WRF version 3.4 to account for a bug in the code in the calculation of the momentum diffusion coefficient (Hu et al., 2013). Hu et al. (2013) claims this improves wind speed estimation in stable regimes by decreasing vertical mixing, and is one of the only PBL schemes to focus improvements in wind.

(iii) Mellor-Yamada-Janjic (MYJ) scheme: A local mixing, 1.5-order closure, prognostic turbulent kinetic energy (TKE) scheme, used operationally in the NAM. TKE vertical distribution is controlled via a master mixing length, which is used in the TKE budget equation (Janjić, 1994). Local closure implies only small turbulent-eddy contributions to the TKE distribution are accounted for.

(iv) Asymmetric Convective Model (ACM2) scheme: A combination non-local transilient turbulence (Stull, 1993) scheme during unstable conditions with local eddy diffusivity in stable conditions (Pleim, 2007). A stability parameter controls the contribution of the total mixing from non-local and local components. Downward mixing is controlled by local K-theory. Pleim (2007) show strong agreement between the ACM2 scheme and LES. The PBL height is determined by a critical bulk Richardson number.

(v) Quasi-Normal Scale Elimination (QNSE) scheme: The QNSE scheme follows the same theory as the MYJ scheme in unstable and neutral conditions but develops a spectral theory during stable periods (Deppe et al., 2012). The goal of this scheme is to more accurately develop the boundary-layer during stable regimes (Sukoriansky et al., 2005). This scheme treats turbulent eddies as waves during stable regimes.

(vi) Mellor-Yamada Nakanishi Niino Level 2.5 (MYNN) scheme: Also a local mixing, 1.5-order closure, prognostic TKE scheme, and a slight variation of the MYJ scheme. TKE vertical distribution is controlled via a master mixing length that differs from that in the MYJ scheme. For the MYNN scheme, the master mixing length is the sum of three individual length scales: the surface-layer length, the turbulent-layer length and the buoyancy length (Deppe et al., 2012).

(vii) Grenier-Bretherton-McCaa (GBM) scheme: The GBM scheme was formed to try and improve representations of cloud-topped marine boundary layers (Grenier and Bretherton, 2001). It uses moist thermodynamics in its turbulence scheme, unlike others that use unsaturated ther-

modynamics. The GBM scheme is also a 1.5-order closure local TKE prognosis scheme using K-theory to parameterize turbulent fluxes.

(viii) University of Washington (UW) scheme: A modified GBM scheme taking into account important cloud-top processes such as cloud-top cooling, evaporation and phase change (Bretherton and Park, 2009). This scheme uses moist thermodynamic variables as in GBM and slightly varies the master mixing length formulation.

2.2.2 Initial-Condition Sources

Initial conditions are used as the best estimate of the initial atmospheric state in a model simulation, whereas boundary conditions provide updates to the lateral boundaries of the outermost model domain during the simulation. In this dissertation I refer to both generally as *initial conditions*. In a similar study over uniform terrain, Deppe et al. (2012) show different initial-condition sources do lead to different solutions for hub-height winds after progressing forward in time.

For this study the NCEP GFS and NAM were used for the initial conditions. Each source was tested with all PBL schemes mentioned in section 2.2.1. GFS-initialized WRF runs used the 0000 UTC GFS output at 0.5 degree resolution, while NAM-initialized WRF runs used 0000 UTC 32-km NAM output. The NAM contains many different grid lengths ranging from 4 to 32 km at the time of this writing. The master grid of 32-km grid length was used for this study because it was the only NAM grid that captured the full study area. The NAM (NMM dynamical core) differs from WRF-ARW primarily in its grid configuration. The NAM is a grid box model operating on an Arakawa-B grid arrangement whereas the WRF-ARW core used in this study uses an Arakawa-C grid arrangement.

2.2.3 Verification Metrics

Murphy and Winkler (1987) detail a systematic framework for forecast verification. Namely, the joint distribution of forecasts and observations should be considered, and the forecast user or provider should pinpoint which aspects of the forecast are most relevant for their use. Wind power is proportional to the cube of wind speed for individual wind turbines, although the farm-averaged wind power may be slightly less than cubic. Therefore, errors in estimates of wind power are highly sensitive to differences between the forecast and observed wind speeds, which should also be highly correlated. State-of-the-art forecasts should capture wind-speed variations as well as forecast uncertainty due to imperfect models and initial conditions.

This chapter addresses short-term, deterministic forecast performance through the following metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and linear correlation co-

efficient (r). The best forecasts will have small MAE (high accuracy), small RMSE (small amount of large forecast errors), and r closest to 1.0 (forecasts and observations trend with each other). When gauging forecast performance of a large number of forecasts (in this case 48 forecasts) a summary plot such as a Taylor diagram (Taylor, 2001) is useful. It shows centered-RMSE (CRMSE), r, and standard deviation (the latter showing how well the standard deviation of the forecast values matches the standard deviation of the observations). CRMSE is calculated by removing the mean value of the forecasts and mean value of the observations from each forecast-observation pair when calculating RMSE [Eqn. (2) in Taylor (2001)].

I also use the standardized anomaly (z-score) to measure which model configurations are better or worse than average. In addition, MAE skill score (MAESS) is used to quantify the amount of improvement that can be gained using one configuration over another, and through the use of bias correction (detailed later).

Finally, an Analysis of Variance (ANOVA) was done to estimate the importance of three factors (initial-condition source, grid length, and PBL scheme) on the hub-height wind-forecast accuracy. An ANOVA is a method of evaluating how much of the overall variance of a quantity (in this case, MAE) can be explained by any number of factors (Wilks, 2011). I use the ANOVA to pinpoint which of these factors has the biggest potential for improving short-term, hub-height wind forecast accuracy over complex terrain for the wind farms studied here.

This chapter does not address aspects of ensemble forecasting or forecast uncertainty, but the authors would like to mention their importance for wind-energy forecasts. Forecast uncertainty is typically derived from an ensemble prediction system, where multiple models are run producing a distribution of possible forecast solutions. The resulting forecast probability distribution provides an indication of forecast uncertainty. This aspect of the study is provided in Chapter 3 and Siuta et al. (2017a).

2.2.4 Observations

Hub-height wind-speed observations used to calculate verification metrics in this dissertation are the confidential property of the wind-farm owners. The authors have taken careful consideration to not reveal wind-farm locations and observed values. I refer to the wind-farm sites as stations numbered 1-4. Hub heights ranged from 80 to 95 m, depending on the wind site. All wind farms were located on mountain ridge-tops with elevations ranging from approximately 500 to 1000 m above sea level. I use wind-farm-averaged observations to compare with model output. Wind-farm-averaged observations are derived from nacelle-based observations. The raw, hub-height wind observations from the wind farms were quality controlled by BC Hydro (a utility company independent of the private wind-farm operators) before being made available to the authors.

2.2.5 Bias Correction

I apply a degree of mass balance (DMB), or multiplicative, bias-correction scheme detailed in Bourdin et al. (2014). This statistical post-processing scheme is suitable for wind forecasts because it does not allow for negative wind speeds. Bias correction effectively removes the systematic component of error, but is only possible by having access to real-time wind-farm observations. Chapter 3 discusses this bias-correction method and the effects of bias-correction on ensemble and probabilistic wind forecasts. I found that a running 30-day training period for bias correction provided the maximum increase in forecast skill over raw forecasts, when compared to other training periods ranging from 1 to 30 days (not shown here).

This bias-correction technique implicitly accounts for average wind-turbine wake effects. However, it is possible that this method does not successfully differentiate between waked and unwaked flows, so some wake-induced biases still exist under specific regimes. I did not use the wind-turbine drag parameterization scheme in WRF, which is not compatible with all the PBL schemes used here.

2.3 Results and Discussion

Given the complexity of evaluating 48 forecasts, my discussion will be split into two subsections. First, I detail the deterministic performance of each configuration for the bias-corrected forecasts. Second, I discuss the sensitivity of the bias-corrected forecast accuracy to each of the three factors (initial-condition source, grid length, and PBL scheme). Appendix A contains the complete MAE dataset used in the following discussion for the bias-corrected forecasts and a summary of forecast skill gained through bias correction.

2.3.1 Deterministic Verification

Annual Taylor diagrams for sites 1-4 (Fig. 2.3) provide a quick summary of the short-term forecast performance of the 48 forecasts. Better forecasts will have lower CRMSE (closer to the center of the gray semi-circles), a linear correlation coefficient (dashed radials) closer to 1.0 (on the abscissa), and match the standard deviation of the observations (be closer to the yellow line). In Fig. 2.3, similar symbols represent similar PBL schemes and similar colors represent similar grids (i.e., grid length and initial-condition source).

While each Taylor diagram contains 48 forecasts, several key patterns can be found by looking at the differences in the plots between wind farm sites. At sites 1 and 4, the Taylor diagrams show *grid clustering*, meaning forecasts from the same grid length perform similarly, regardless of the PBL scheme or initial condition used. In other words, differences in the forecast performance are mainly a function of grid length or of the location of the nearest-neighbor grid point. At site 1,

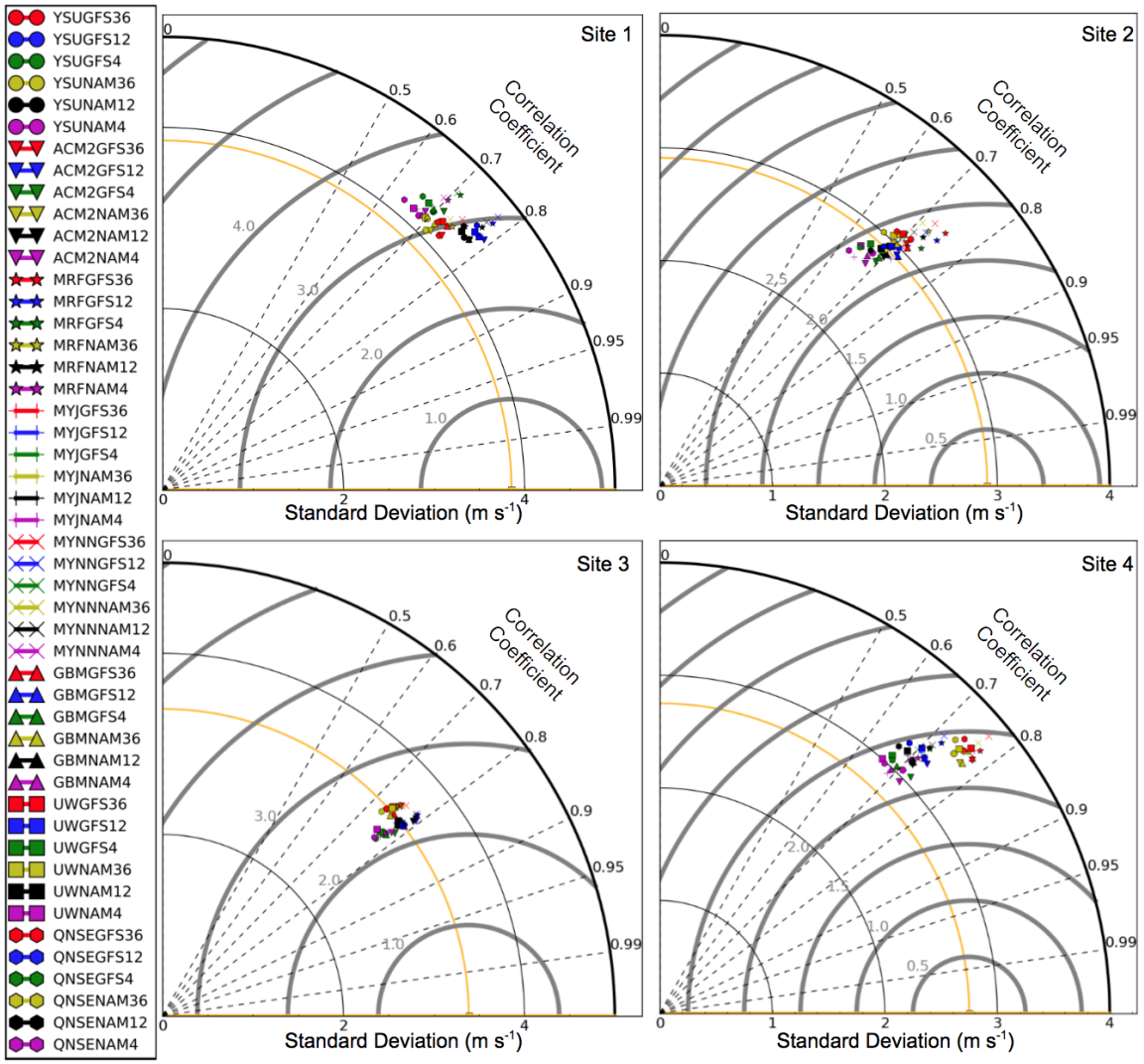

Figure 2.3: Taylor diagrams of bias-corrected wind forecasts for sites 1-4. The symbols in the legend differentiate PBL schemes, where similar colors represent similar grid lengths and initial conditions. Better forecasts will be located close to the yellow curve, which represents the standard deviation of the observations. The best forecasts will be located closest to the bottom of the yellow curve where there is also perfect forecast-observation correlation (r=1.0) and no CRMSE. Correlation values are provided by the dashed radials, CRMSE by the thick gray semi-circles (m s^{-1}), and standard deviation by the thin black quarter-circles. Legend naming convention is [PBL]{IC}(Grid).

the grid clusters are primarily differentiated by their correlation coefficients, with the 12-km grids having the highest correlation and lowest CRMSE. At site 4, the grids are primarily differentiated by their standard deviation, with the 4-km grids able to most closely match the standard deviation of the observations (the yellow quarter-circle curve). At site 4, however, the 36-km grids have the highest correlation, while having similar accuracy to the 12- and 4-km grids.

At sites 2 and 3, the grid-clustering pattern is not so distinct. At site 2, the 12-km grids provide the closest match to the observations, while the 4-km (36-km) grids generally have too little (too much) standard deviation. Forecasts using the MRF PBL scheme, which is an older scheme, have the best correlation with the observations. At site 3, the grid clustering is evident, but the difference in standard deviation between the different grid clusters is much smaller than at sites 1 and 4. At site 3, the 12-km grids also provide the best forecasts because they have higher accuracy and correlation, while better matching the standard deviation of the observations.

While the Taylor diagrams (Fig. 2.3) have hinted at a large importance in the choice of grid length in complex terrain, there is also a clear difference in the performance of forecasts using different PBL schemes for the same grid length. To quantify which of the factors (grid length, PBL scheme, or initial-condition source) is the most important, I later perform an ANOVA on the distribution of MAE scores.

One other finding here is that using either the GFS or NAM as the initial conditions produced similar forecasts for the short-term (see Appendix A). To simplify subsequent figures, the NAM-initialized members are omitted. One potential reason for the similarity in the GFS and NAM produced forecasts for equivalent configurations is that both these models are produced by NCEP and rely on the same underlying data assimilation system, the Gridpoint Statistical Interpolation (GSI, Shao et al., 2016). Thus, their initial states could be more similar than when compared to initial conditions produced from other agencies or data assimilation methods.

Reducing the dataset further, Fig. 2.4 shows the MAE standardized anomaly (z) scores for the bias-corrected WRF forecasts at each site for the summer (June, July, and August) and winter (December, January, February) seasons. Since the standardized anomaly compares the MAE of each scheme to the mean MAE of all schemes, negative values indicate configurations that perform better than the average (lower MAE).

Fig. 2.4 shows two distinct patterns in forecast accuracy. PBL schemes that include non-local mixing (YSU, ACM2, and MRF schemes; boxed, Fig. 2.4) have accuracy higher than those that include only local mixing in summer, although forecasts using the MYJ scheme also generally had better than average accuracy. Non-local schemes treat vertical mixing in a more realistic way by recognizing that turbulent eddies can span and/or move through multiple small layers within the boundary layer. In other words, mixing is attributed to not only eddies present within a small,

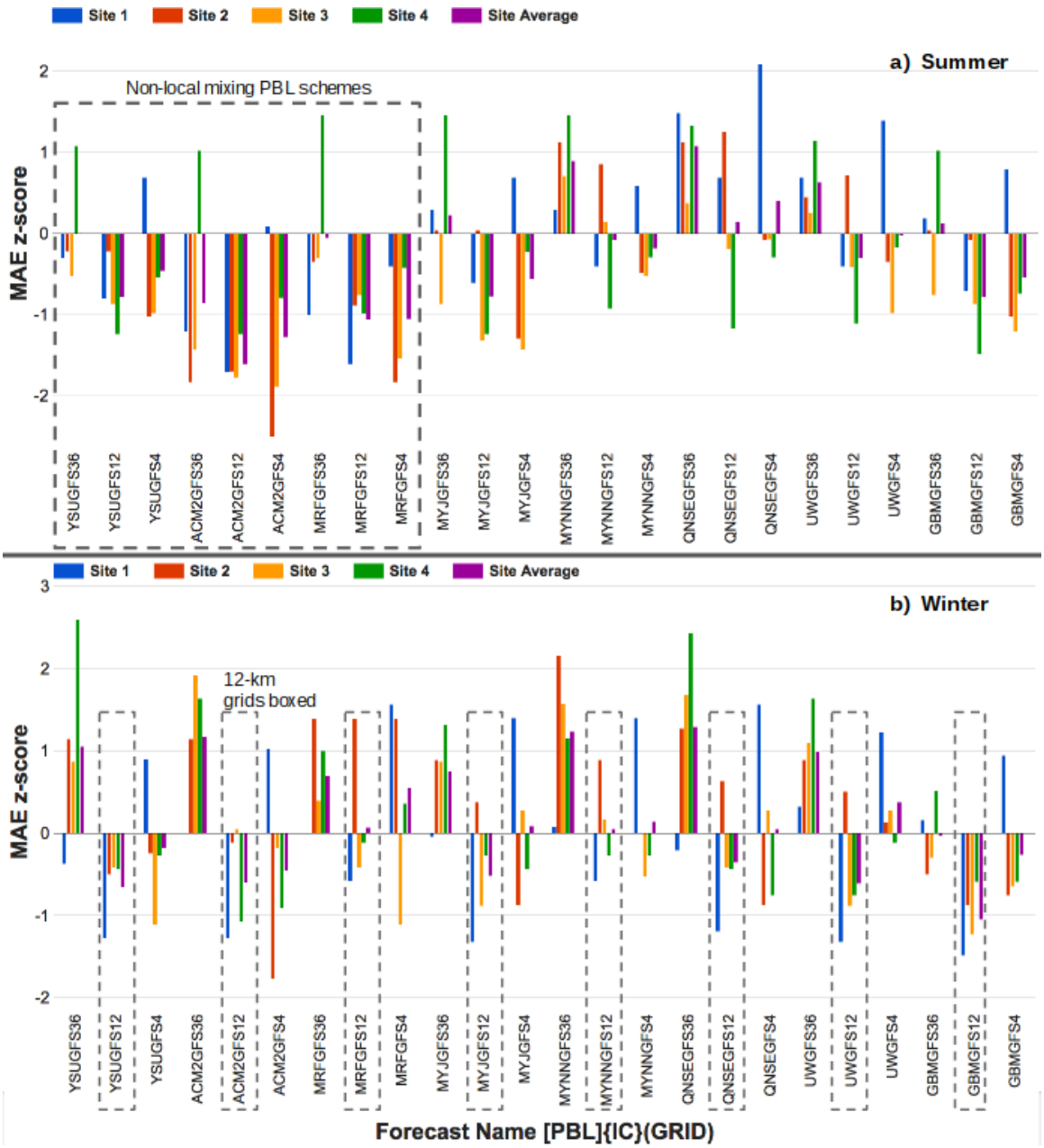

Figure 2.4: Accumulated z-score for each of the 48 bias-corrected wind forecasts for the summer (June, July, and August; top), and winter (December, January, and February; bottom). Colors represent individual wind farm sites, with purple bars indicating the site-averaged performance. A larger-magnitude negative z-score indicates forecasts that are much better than the average. The boxed-areas represent the non-local mixing PBL schemes in the top panel and the 12-km grids in the bottom panel.

local layer of constant static stability, but also those that originate in other layers that move through in transit to their level of neutral buoyancy (Stull, 1993). These large eddies will have the most influence in complex terrain over the summer season where differential heating over sloped terrain can lead to important boundary-layer circulation development.

The 12-km forecast using the ACM2 PBL scheme initialized off the GFS had the best accuracy averaged over all wind farms in summer (purple bars, Fig. 2.4), although other forecasts had better accuracy at individual wind farms. For example, the 4-km forecast using the ACM2 PBL scheme was best at sites 2 and 3, and the 12-km GBM scheme was best at site 4, all initialized off the GFS.

In winter, the MAE z-scores show a different pattern. Here, the 12-km grids for each scheme, regardless of the treatment of vertical mixing, generally perform better (boxed, Fig. 2.4), especially when averaged over all locations. The exceptions are the MYNN and MRF PBL schemes which perform relatively poorly regardless of grid length. I also find several cases where the 4-km grids produce better than average forecasts.

Averaged over all locations, the best forecast for the winter was the 12-km forecast produced using the GBM PBL scheme. Again, individual sites did show slightly better accuracy with other configurations. Site 1 had the best accuracy from the 12-km grid using the GBM PBL scheme. Site 2 had the highest accuracy using the 4-km grid and ACM2 PBL scheme, site 3 with the 4-km grid and MRF PBL scheme, and site 4 with the 12-km grid and ACM2 PBL scheme. In contrast to the summer, the most accurate bias-corrected forecasts in the winter were initialized off the NAM. In winter when static stability is generally higher and wind-shear-generated mechanical turbulence is greater, non-local mixing is likely not as important, especially at these latitudes where sun angle is low. Low-pressure systems frequently approach the coast of British Columbia and move into the interior of North America, and dynamically forced flows have more influence on wind patterns in complex terrain.

While one might expect the 4-km grid to have the best forecast verification due to superior resolution of topography, the results indicate the 12-km grids generally have higher accuracy (with the mentioned exceptions), particularly when averaged over all locations. Mass et al. (2002) suggest that while shorter grid lengths do simulate the structure of atmospheric phenomena better, they are subject to positional error. Mass et al. (2002) and Murphy and Winkler (1987) also mention that skill measures should ultimately be defined by the forecast user. So, while the 4-km grids may produce more realistic-looking feature structures, for the purposes of generation planning the forecast with the least error over the long term at a wind farm location is ultimately preferred. My results show that higher resolution grids (those with shorter grid lengths) may provide a poorer quality forecast than coarser (12-km) grids, while being more computationally expensive to produce (although I do find several cases where the 4-km grids are the most accurate). However, I am not advocating

entirely for coarse grids to be used for wind forecasts in regions of complex terrain. The grid resampling process that occurs with two-way nested domains could result in the 4-km grids have a positive influence on the 12-km grid forecast accuracy. Additionally, it is clear from Fig. 2.4 that the 36-km grids generally produce some of the least accurate forecasts.

To highlight the potential for improvement in forecast accuracy by using particular PBL schemes or grid lengths, I have calculated the MAE skill scores (MAESS) for the best forecast configuration relative to the worst for each site for both the summer and winter seasons. In summer, sites 1, 2, and 3 had MAESS values (i.e., a potential improvement in MAE) of 18, 17, and 19%, respectively. Site 4 had a much larger MAESS of 29%. In winter, MAESS were 24, 17, 15, and 13% for sites 1-4, respectively. Differences of this magnitude are significant to the wind-energy community (Marquis et al., 2011), and highlight the importance of careful grid-length and PBL-physics configuration.

Fig. 2.5 shows the MAE z-score for the entire year. These results show the best forecast averaged over the entire study period and all locations to be the 12-km ACM2 forecast initialized with the GFS initial-condition source (purple bars, Fig. 2.5). Forecasts from the 4-km grid using the ACM2 scheme, and the 12- and 4-km grids using the GBM schemes, all with the GFS initial-condition source, are also some of the more accurate forecasts.

I theorize that the strong performance of the ACM2 and GBM schemes can be attributed to two distinct qualities of these respective schemes. The ACM2 scheme treats vertical mixing in the most realistic manner of all the schemes. It allows for both non-local and local mixing via its formulation (Pleim, 2007). This is similar to transilient turbulence theory, which more accurately captures boundary-layer-spanning eddies that occur in convective boundary layers (Stull, 1993). Such mixing characteristics will not be captured by local-mixing schemes. In terms of the GBM scheme, its physical formulation differs from other schemes in that it incorporates moisture into its thermodynamic calculations, which better represents cloud-topped boundary layers (Grenier and Bretherton, 2001). Due to the active storm track along coastal British Columbia, cloud-topped boundary layers are quite common, and may be a reason why this scheme performs well.

2.3.2 Sensitivity Analysis

So far, this study has shown that proper selection of the PBL scheme and grid length can lead to significantly improved forecast accuracy for hub-height wind forecasts. Also, the best performing configuration varies by location and season. To quantify which factor (PBL scheme, grid length, or initial condition) is the most important to forecast accuracy at these wind farms, a multi-factor (factorial) ANOVA was done on the MAE scores.

Fig. 2.6 shows the contribution to the overall variance in MAE scores for the bias-corrected WRF forecasts attributed to each factor at each wind site. The plot also shows how the contribution

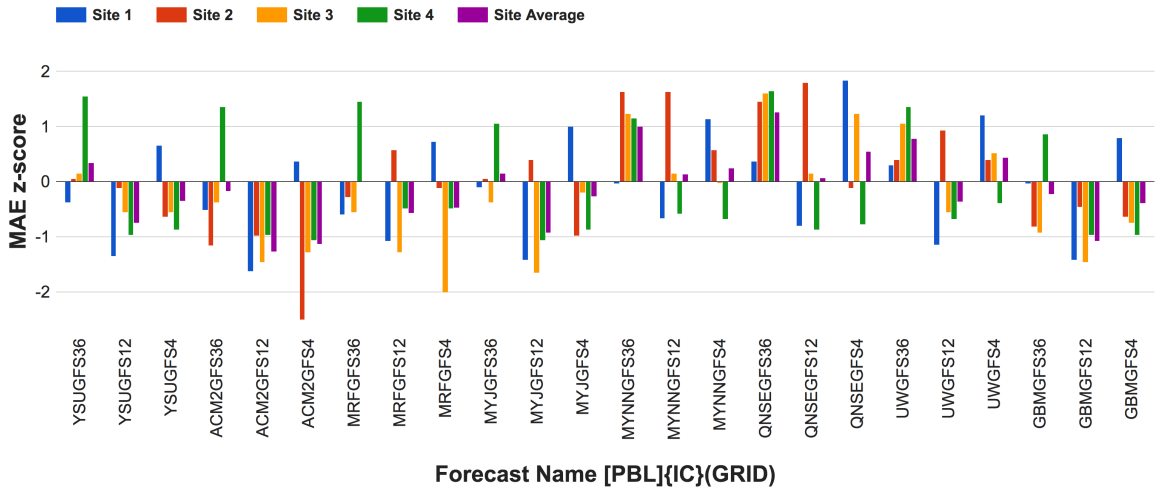

Figure 2.5: Annual MAE z-score for each of the 48 bias-corrected wind forecasts. Colors represent individual wind farm sites, with purple bars representing the site-averaged performance. A larger-magnitude negative z-score indicates forecasts that are much better than the average.

to variance changes by season. In addition, the *residual* column includes the variance that can be attributed to the interaction between two or more factors.

At sites 1 and 4, where the grid clustering was easy to distinguish in Fig. 2.3, the grid length is the dominant factor controlling forecast accuracy for all seasons on the year. Site 4 in particular has the largest influence of the grid length, contributing to 70% or greater of the variance among the MAE scores for all seasons. At site 1, the grid length explains 50% of the variance in MAE, or greater, depending on the season. Both these locations also show a seasonal cycle, where the PBL scheme or the residual increase contribution to the variance in MAE scores. At site 1, the PBL scheme has the largest influence in the spring and summer months, although the grid length is still the dominant factor. At site 4, the influence of the PBL scheme increases in the fall and winter, but explains 12% or less of the variance in forecast accuracy. At site 2, the PBL-scheme is the dominant factor, with also a strong seasonal cycle. The PBL scheme explains a minimum of 36% of the variance in winter and and a maximum of 62-70% in the spring and summer months. At site 3, the PBL scheme is dominant in the fall and spring, the grid length in the winter, and initial-condition source in the summer. Site 3 is the only location where the initial-condition choice explains most of the variance for an individual season (52% in the summer). Site 3 is also the location where the PBL scheme has explained the highest amount of the variance in MAE for an individual season at 80% in the spring.

Like the initial analyses, the ANOVA also indicates that the choice of initial and boundary con-

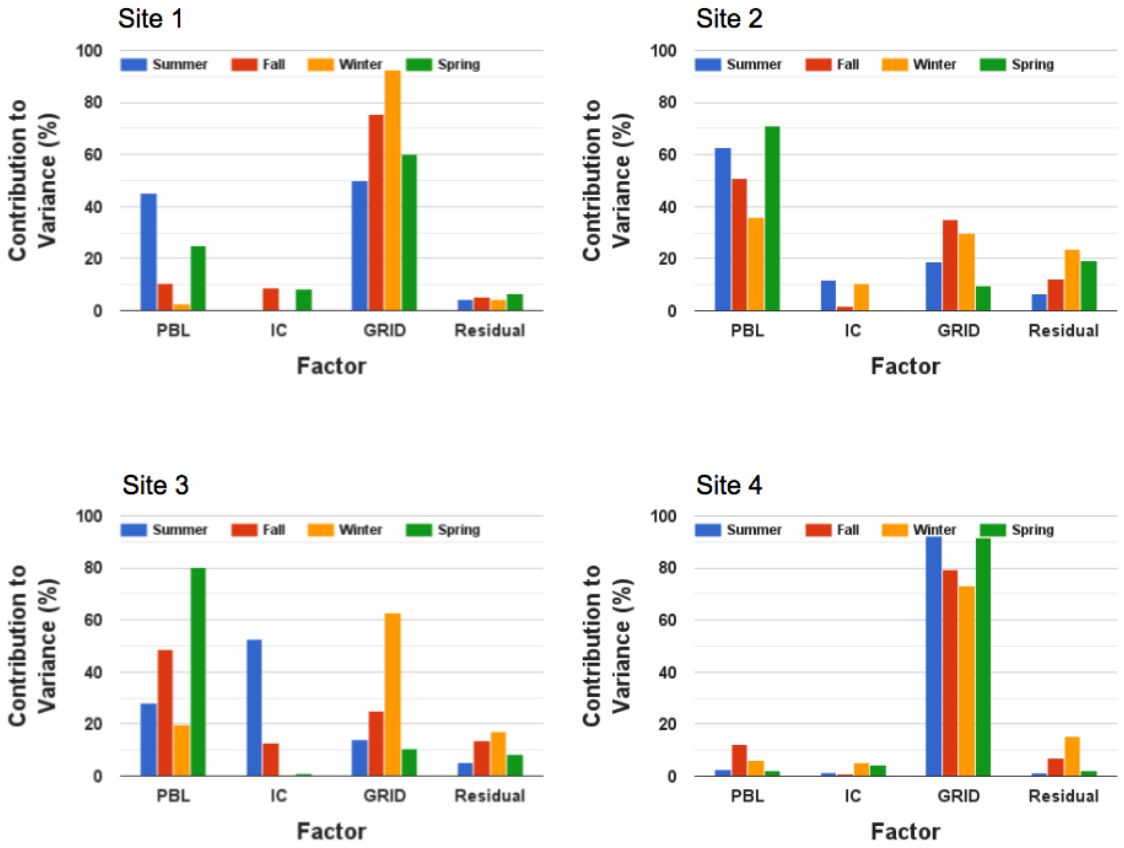

Figure 2.6: Contribution to the variance in MAE scores for the bias-corrected wind forecasts as a percentage contribution by each factor for each site. Different colors represent different seasons, where blue is summer (June, July, and August), red is fall (September, October, and November), orange is winter (December, January, and February), and green is spring (March, April, and May). The residual represents the variance attributed to factor interactions and other unaccounted differences.

ditions explain little of the variance in forecast accuracy for the short term at most of the locations, with the exception being site 3 in the summer. However, there are three potential caveats of this analysis. The first is that both sources are produced by NCEP, with both relying of the GSI data assimilation system. Greater variance may be achieved by using initial-condition sources produced by separate agencies and separate data assimilation methods. The second is that greater variance may be achieved at longer forecast horizons since a 24-hour forecast may not have allowed enough time for the small differences in initial conditions to grow significantly. Finally, while the results of the ANOVA make physical sense, the small sample size does not perfectly meet the primary assumptions made in using an ANOVA. Namely, the factor-level data are not perfectly normal, and

exhibit some heteroscedasticity.

Since the grid length affects the exact location of the nearest-neighbor grid-cell, I posit the physical location of the nearest-neighbor grid point could contribute to the variance in accuracy scores at some locations (especially those that have accuracy largely explained by the grid length). To evaluate these effects, I looked at the distance between the nearest-neighbor grid point and the actual coordinates of the wind-farm site. At site 1 (and site 4) which found the grid length to control MAE variance for individual seasons (Fig. 2.6), I found horizontal distances of 14.5-, 4.1-, and 1.4-km (and 11.8-, 5.5-, 2.1-km), for the 36-, 12-, and 4-km grids, respectively. At site 2 (and site 3), which found the PBL scheme to be the dominant factor for most seasons, the horizontal distances were found to be 15.2-, 6.5-, and 2.1-km (and 19.3-, 2.7-, and 1.1-km), for the 36-, 12-, and 4-km grids, respectively. This indicates that the physical error in location of the nearest-neighbor grid cell may not be the cause for some locations being more largely influenced by the grid length.

Next, I checked to see if differences in the modeled terrain height at the nearest-neighbor grid points are larger between grid lengths at locations that showed higher influence of the grid length. Again, this does not seem to explain why some locations are more effected by the grid length (sites 1 and 4, Fig. 2.6) as all locations had similar height variations between nested grids of approximately 100 to 200 m. I finally checked to see if there was an abrupt change in land-use category and thus roughness between the grids and came to a similar conclusion, although for one location the nearest 36-km grid cell identified as being over water. Thus the reason why hub-height wind forecasts at particular sites are more affected by the PBL scheme, while others by the grid length, remains unclear and is worthy of future study. One theory could be that some locations are more affected by flow channeling and blocking, which may be more largely influenced by the grid length than the PBL scheme. Appendix C provides a summary of the differences between the model-terrain and that of a digital elevation model.

Fig. 2.7 shows the contribution to MAE variance by time of day at each location (over the entire year). The midday (1200-1800 PST) peak in PBL scheme contribution to variance for sites 1, 2, and 3 is consistent with the development of slope flows and boundary-layer growth. Forecast accuracy at site 4 does not exhibit a diurnal cycle, with the grid length explaining the most variance throughout the entire day. Since each forecast run was initialized at the same time of day, the results in Fig. 2.7 also show the relative contribution of each factor by forecast horizon, although the signal of the diurnal cycle appears to be more evident than that of forecast horizon.

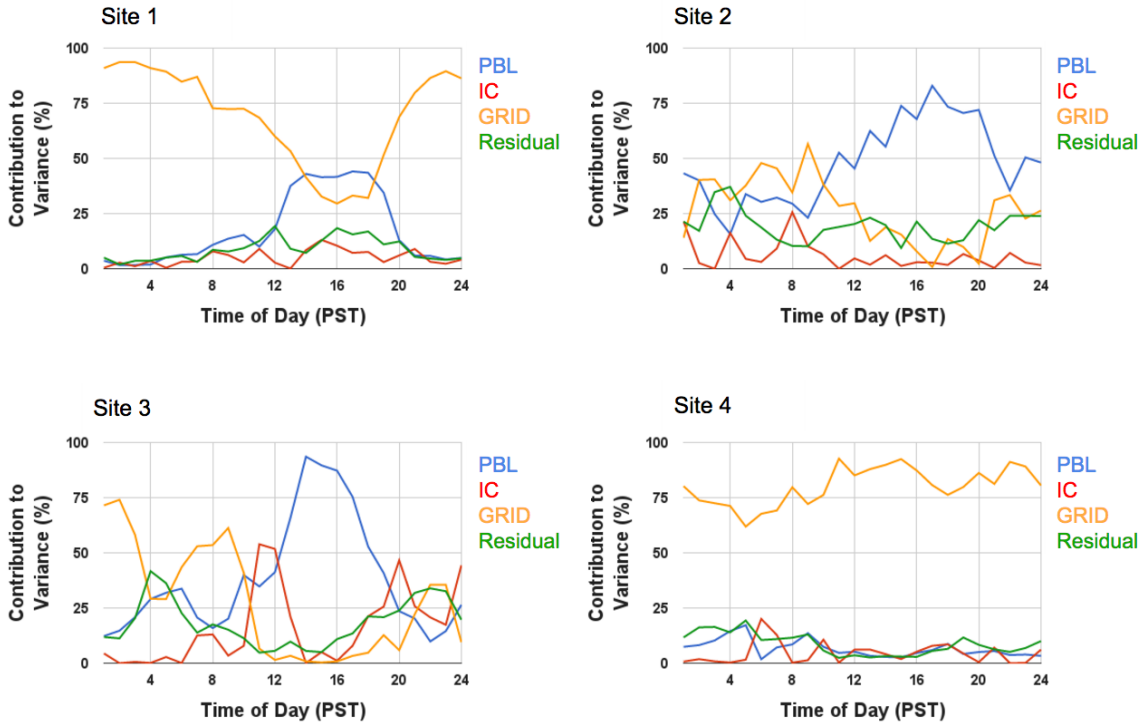

Figure 2.7: Contribution to the variance in annual MAE scores for the bias-corrected wind forecasts by time of day as a percentage contribution by each factor for each site. Different colors represent different factors, where blue is the PBL scheme, red is initial condition, orange is grid length, and green is the residual. The residual represents the variance attributed to factor interactions and other unaccounted sources.

2.4 Conclusions and Future Work

This study compared the sensitivity of hub-height wind-speed forecasts to the PBL scheme, grid length, and initial-condition source for the WRF model over complex terrain. I have provided several important conclusions that forecasters and researchers should consider when using NWP models to produce hub-height wind-speed forecasts. Namely, I found the grid length and PBL scheme were the most important factors to consider for short-term forecasts (24-hour forecast horizon) in complex terrain, for my data sets. For the majority of locations, the effects of PBL scheme and grid length contribute more to differences in the bias-corrected forecast accuracy than initial-condition choice. This means that choosing the best PBL scheme and grid length over the worst could yield the largest accuracy gains. However, one caveat could be a lack of diversity in initial-condition sources, since the processes used during data assimilation for the NAM and GFS are similar, and both are produced by the same agency. Initializing WRF with other initial-condition sources, such

as the with the Canadian Global Deterministic Prediction System (GDPS), the Fleet Numerical Meteorology and Oceanography Center Navy Global Environmental Model (NAVGEM), the European Center for Medium Range Weather Forecast Integrated Forecast System (IFS or ECMWF), or the UK Met Office Unified Model (UKMO) could result in a larger sensitivity to initial-condition choice than shown here, and is worthy of future study.

At sites 2 and 3, the PBL scheme was a large or dominant factor controlling forecast accuracy, regardless of the season. At the other two stations (sites 1 and 4), the grid length had the largest influence on forecast accuracy.

Additionally, the influence of the PBL scheme showed a seasonal cycle at sites 1, 2, and 4. At sites 1 and 2, the influence of the PBL scheme (grid length) peaked in the spring and summer (fall and winter) months. At site 4, the influence of the PBL scheme was small, but peaked in the winter. Site 3 had a different seasonal trend where the PBL scheme and grid length were the largest influence over forecast accuracy for all seasons, except for the summer.

The influence of the PBL scheme was also largest during times of the day where boundary-layer circulations are expected to develop (midday to early afternoon at sites 1-3). Site 4 showed relatively little variation in which factor contributed the most to the variance in accuracy, regardless of the time of day.

I investigated if the reason why sites 1 and 4 were more heavily influenced by the grid length was because of differences in the location of the nearest-neighbor grid point. I concluded that the physical distance of the wind-farm location to the nearest neighbor location was not the primary cause. I also found that the differences in elevation of the nearest-neighbor grid points were similar at all locations, and that land-use types were similar between grids at 3 of the 4 locations. I concluded that the locations that were most strongly influenced by the grid length are likely more impacted by terrain-flow interactions, such as channeling and blocking, than at the other locations, but that this finding is worthy of future study.

When averaged over all locations and seasons, the ACM2 PBL scheme was the most accurate PBL scheme in complex terrain, although for some specific farms and seasons it was slightly outperformed by other schemes. Energy planners will likely benefit the most by using the ACM2 scheme in complex terrain since it had the best overall average accuracy (averaged over all locations). Choosing the best PBL scheme and grid length configuration resulted in reductions of MAE of up to 29% over the poorest configuration, depending on the season, with larger improvements found in the summer.

Improvements of this magnitude can save millions of dollars (Marquis et al., 2011). The sensitivity results provide valuable information to meteorologists, energy planners, and utilities that use wind-energy forecasts. Namely, savings can be realized through careful experimentation with their

WRF configuration, focusing on PBL-scheme and grid-length selection, as well as bias correction.

An interesting study would be to compare the effects of one-way and two-way domain nesting techniques on hub-height wind forecasts over complex terrain. While I find the 12-km grids to generally have higher accuracy, particularly in the winter, it is possible that this is a result of the resampling and averaging process that takes place between nested grids. The 12-km grid forecast at these locations might be improved over the 4-km grid through this feedback process. A useful follow-up to this study would be to rerun the same tests with one-way nesting. This would eliminate fine-grid resampling and may provide different results with respect to the performance of fine and coarse grid lengths. It would also provide insight into the benefits of one-way versus two-way nested grid setups in WRF over complex terrain.

Because I have shown that the best PBL scheme and grid length varies by location and season, it may be beneficial to use an ensemble of multiple PBL schemes, grid lengths, and initial-condition sources to improve accuracy through use of the ensemble mean forecast. This study has only evaluated the deterministic aspect of wind forecasts. Capturing forecast uncertainty is critical, especially for energy planning. The use of an ensemble consisting of all available PBL schemes, grid lengths, and initial conditions may be beneficial, especially over complex terrain where boundary-layer circulations are prevalent. Chapter 3 evaluates the bias-correction method and the use of this dataset for ensemble and probabilistic forecasts. Use of probabilistic information has been shown to provide further economic value over individual deterministic forecasts (Zhu et al., 2002; McCollor and Stull, 2008b).

Chapter 3

Calibrated Probabilistic Hub-Height Wind Forecasts in Complex Terrain

3.1 Introduction

Planners and market traders use wind forecasts based on numerical weather prediction (NWP) models to gauge future wind-energy production. Such NWP models are the largest source of uncertainty in wind-energy generation estimates (Monteiro et al., 2009). Forecast uncertainty arises from incorrect initial-condition (IC) sources, incomplete knowledge of atmospheric physics (e.g., turbulence), and improper assimilation of observational data (Buizza et al., 2005). Better estimation of forecast uncertainty, and higher forecast accuracy lead to more efficient energy planning, optimized market trading opportunities, and significant monetary savings (Pinson et al., 2006; Marquis et al., 2011; Mahoney et al., 2012; Wilczak et al., 2015).

Probabilistic weather forecasts are the primary tools used to gain insight on forecast uncertainty. Earlier forecasts used for wind-energy planning were deterministic, meaning they did not provide information on forecast uncertainty (Pinson et al., 2006). However, in the last 10 years, probabilistic forecasts have begun to be adopted throughout the industry as planners started to recognize the value of quantifying forecast uncertainty. Monteiro et al. (2009), Giebel et al. (2011), and Zhang et al. (2014) provide extensive reviews of the history of probabilistic forecasts for wind energy.

Zhang et al. (2014) detail the two standard approaches for creating probabilistic forecasts for wind energy: parametric and non-parametric methods. Parametric methods prescribe a probability distribution of a particular shape that is typically representative of the past forecast-error distribution. Such distributions are described by location and scale parameters and have the benefit of being computationally cheap (Zhang et al., 2014). Early methods dressed distributions around single deterministic forecast models, with several distributions being tested. Lange (2005) found that the forecast-error distribution calculated from the German Weather Service (DWD) numerical forecasts are well-dressed by a Gaussian distribution. They provided the forecast-error-distribution results for 10-m wind speeds, not hub-height winds, and likely only at a location in flat terrain (the exam-

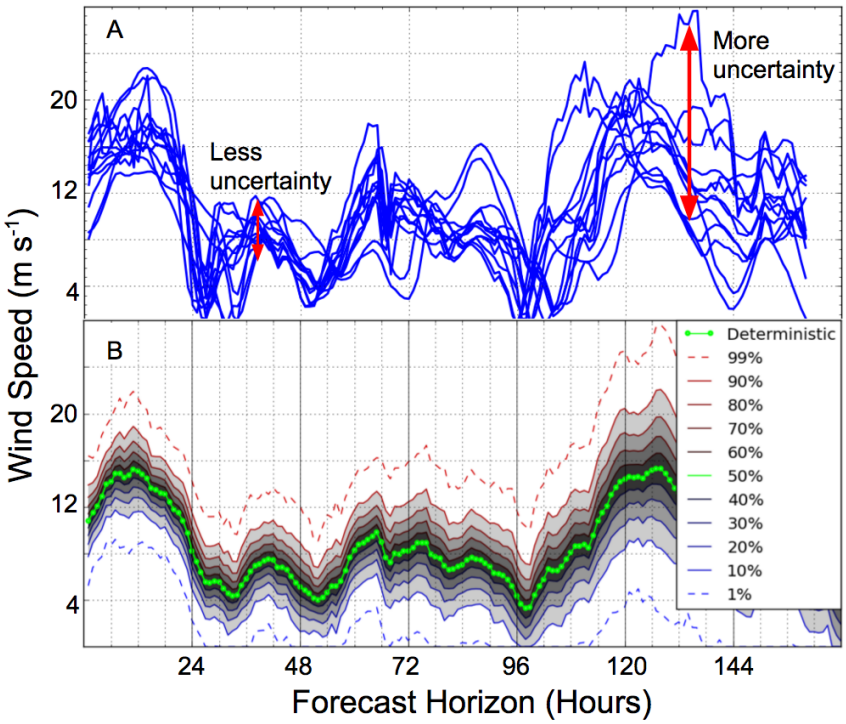

Figure 3.1: Example ensemble forecast (A) and corresponding probabilistic forecast (B).

ple location is confidential and a large quantity of the stations tested were in flat terrain). When converted to wind power, forecast-error distributions have been fit by Gaussian (Lange, 2005), Beta (Bludszuweit et al., 2008), and generalized logit-normal distributions (Pinson, 2012).

Non-parametric approaches do not make a shape assumption. One example is an ensemble of NWP models. Ensemble forecast systems provide a distribution of possible forecast outcomes (Fig. 3.1a), and can be produced by NWP using a variety of methods: 1) using multiple IC sources, 2) stochastically perturbing a single IC, 3) perturbing observations used during data assimilation for observation error, 4) varying model physical parameterization configurations, 5) varying NWP dynamical cores, 6) varying model grid lengths, 7) using stochastic kinetic energy backscatter (SKEBS), or 8) stochastically perturbing parameterization tendencies (SPPT) (Stensrud et al., 2000; Grimit and Mass, 2002; Buizza et al., 2005; Eckel and Mass, 2005; McCollor and Stull, 2008b; Candille, 2009; Berner et al., 2009; Palmer et al., 2009).

An ensemble forecast distribution can be converted into a probability distribution (Fig. 3.1b). This is commonly done by using the empirical distribution of ensemble members (Anderson, 1996). Bayesian Model Averaging has also been successfully applied to wind-forecast systems (Hoeting et al., 1999; Raftery et al., 2005; Sloughter et al., 2010; Courtney et al., 2013). Kernel density

estimation is another non-parametric method that links one or more explanatory variables to forecast probability density functions (Juban et al., 2007). While such methods have been shown to work well, the latter two require large datasets and can become computationally expensive (Juban et al., 2007; Zhang et al., 2014).

A hybrid approach is to parameterize the moments of a prescribed probability distribution (such as Gaussian) as a function of statistical properties of an ensemble (e.g., the ensemble mean, ensemble variance, and ensemble forecast-error distribution; Gneiting et al. (2005); Nipen and Stull (2011)). This method requires running computationally expensive ensembles, but does not require an extensive training dataset. Gneiting et al. (2005) refer to this method as ensemble model output statistics (EMOS).

Probabilistic forecasts are most useful when the forecast probability of an event matches its observed frequency of occurrence. Distributions from raw ensembles are often under-dispersive (Stensrud et al., 2000; Eckel and Mass, 2005; McCollor and Stull, 2008b), and must undergo bias correction and probabilistic calibration (Buizza et al., 2005; Nipen and Stull, 2011). When the forecast probabilities match the relative frequency of occurrence, a probabilistic forecast is *reliable* or *calibrated*. Calibrated probabilistic forecasts can be trusted as accurate estimates of forecast uncertainty. However, as Nipen and Stull (2011) discuss, even forecasts based on climatology can be probabilistically calibrated, so assessing calibration is not enough. One method that can be used to differentiate probabilistically-calibrated forecasts is through evaluating forecast sharpness (Gneiting et al., 2005; Pinson et al., 2006; Juban et al., 2007). Forecast sharpness is a measure of the width of probabilistic spread, with a deterministic forecast being perfectly sharp (Juban et al., 2007). Pinson et al. (2006) suggest forecast calibration is the primary concern of probabilistic forecasts for wind energy, while any improvements in sharpness represent added value.

The use of ensembles in probabilistic forecasts for hub-height wind speeds is growing, but studies evaluating probabilistic skill are still relatively limited (Junk et al., 2015). Deppe et al. (2012) performed an ensemble study over the flat terrain of Iowa using the Weather Research and Forecasting (WRF, Skamarock et al., 2008) model version 3.1.1., but did not assess probabilistic-forecast calibration. Deppe et al. (2012) used an ensemble of different planetary-boundary-layer (PBL) schemes and IC sources. They found a multi-PBL ensemble had a more accurate ensemble-mean forecast when compared to an ensemble formed from perturbed initial conditions, although the latter had larger ensemble variance. Therefore, a multi-PBL ensemble might be a suitable choice when using the EMOS method of generating probabilistic forecasts.

Since none of the existing PBL schemes available in WRF were designed for complex terrain, and because many wind features in complex terrain are dependent on PBL evolution, I posit that an ensemble containing several PBL schemes may be beneficial. The main differences between PBL

schemes are the treatment of vertical mixing and the statistical order of turbulence closure (Stull, 1988; Stensrud, 2007; Deppe et al., 2012). Differences in PBL schemes could be expected to be larger in complex terrain than was found by Deppe et al. (2012) over flat terrain.

For complex terrain, better methods to produce calibrated, sharp probabilistic hub-height wind-speed forecasts are needed. This study evaluates the performance of multi-PBL, multi-IC, multi-grid-length WRF ensembles at four wind farms in the mountainous terrain of British Columbia using the empirical ensemble distribution and EMOS methods. The eight PBL schemes used here are the Yonsei University Scheme (YSU, Hong et al., 2006; Hu et al., 2013), Asymmetric Convective Model version 2 (ACM2, Pleim, 2007), Medium Range Forecast (MRF, Hong and Pan, 1996), Mellor-Yamada-Janjic (MYJ, Janjić, 1994), Mellor-Yamada-Ninno-Nakanishi and Niino Level 2.5 (MYNN, Nakanishi and Niino, 2006), Quasi-Normal Scale Elimination (QNSE, Sukoriansky et al., 2005), University of Washington (UW, Bretherton and Park, 2009), and Grenier-Bretherton-McCaa (GBM, Grenier and Bretherton, 2001) schemes. See Chapter 2 for details.

Further, this study provides insight on probabilistic forecasting techniques for hub-height wind-speed forecasts using the WRF model. It also adds knowledge about the shape of the hub-height wind-speed forecast-error distribution in complex terrain. This work is a follow-up to Siuta et al. (2017b), who evaluated the deterministic-forecast performance of this dataset and the deterministic-forecast sensitivity to the choice of PBL-scheme, grid length, and IC source. The rest of this chapter is organized as follows: Section 3.2 describes the research methodology used. Section 3.3 provides a discussion of the results. Finally, Section 3.4 discusses general conclusions and suggestions for future work.

3.2 Methodology

The forecast methodology detailed in Chapter 2 (Figs. 2.1, 2.2 and Table 2.1) was used to produce ensemble members and to calculate wind-forecast probabilities. Forecasts from each ensemble member, as well as hub-height wind-speed observations were used in the Component-Based Post-Processing System (COMPS, Nipen, 2012) to post-process and evaluate the forecast. Post-processing is anything applied to the raw forecast for improvement, including bias removal, choice of uncertainty model, and any calibration done to adjust the uncertainty model. COMPS is a modular post-processing and verification system, analogous to how WRF is modular for NWP. COMPS uses a series of namelists containing user-defined post-processing options that are applied to the raw forecast input.

The four wind farms used in this study are located on mountain ridge tops, with elevations between 500 and 1000 m MSL. I use wind-farm-averaged nacelle (hub-height) wind-speed obser-

vations for post-processing and model evaluation. These observations from the independent power producers are confidential and were provided to me by the local utility company, BC Hydro. BC Hydro, who purchases the wind power, performed quality control on the data. Wind-farm averaged observations have been shown to better match power generation and output from NWP models over that of individual turbines or meteorological towers. Farm-averaged values remove sub-grid-scale (intra-wind-farm) variability not currently resolved by NWP models (Cutler et al., 2012). Wind-farm names and locations remain anonymous in the figures shown here. However, statistics and aggregated results based on these observations are provided.

I evaluate several factors affecting the probabilistic hub-height wind-speed forecast in complex terrain, including the effect of using multiple PBL schemes, applying bias correction, and the choice of uncertainty model. These tests are summarized by Table 3.1 and are described in detail in the following subsections.

3.2.1 Effect of Number of PBL Schemes in Ensemble

This study is a follow-up to that by Siuta et al. (2017b), who used the same data set to examine deterministic hub-height wind-speed forecast sensitivity to the choice of PBL scheme, grid length, and IC source. The outcome of that study showed that the grid length and PBL scheme had the most influence on deterministic-forecast accuracy, but the most influential factor varied by location, season, and time of day. Appendix A provides the Mean Absolute Error (MAE) scores for the bias-corrected forecasts used in this prior study. Siuta et al. (2017b) found the ACM2 PBL scheme to be the best-performing of all the PBL schemes, and the 12 km to be the best performing grid, when averaged over all four wind farms over the entire year. However, the best PBL scheme (and grid length) differed for individual wind-farms and seasons (see appendix A).

In this study I start with only the ACM2 PBL scheme (six total ensemble members), and then test for short-term probabilistic-forecast improvements by adding PBL schemes one-by-one into the ensemble. I start with the ACM2 PBL scheme because of the prior results mentioned above.

3.2.2 Bias Correction Technique

I bias correct each individual ensemble member prior to forming the ensemble mean. Wind speeds can never be negative. To satisfy this condition, I use a degree-of-mass-balance (DMB) multiplicative bias-correction technique (Grubišić et al., 2005; McCollor and Stull, 2008a; Bourdin et al., 2014) applied to each ensemble member. The current bias correction factor, DMB_{t}, is calculated

Table 3.1: Summary of the tests performed. Bias correction is detailed in section 3.2.2. A description of the uncertainty models is provided in section 3.2.3.

Test ID	Uncertainty Model	Bias Correction		PBL schemes used								Count of
		Yes	No	ACM2	MYJ	YSU	GBM	UW	MRF	QNSE	MYNN	Ensemble Members
R1	Raw		X	X								6
R2	Raw		X	X	X							12
R3	Raw		X	X	X	X						18
R4	Raw		X	X	X	X	X					24
R5	Raw		X	X	X	X	X	X				30
R6	Raw		X	X	X	X	X	X	X			36
R7	Raw		X	X	X	X	X	X	X	X		42
R8	Raw		X	X	X	X	X	X	X	X	X	48
RB1	Raw	X		X								6
RB2	Raw	X		X	X							12
RB3	Raw	X		X	X	X						18
RB4	Raw	X		X	X	X	X					24
RB5	Raw	X		X	X	X	X	X				30
RB6	Raw	X		X	X	X	X	X	X			36
RB7	Raw	X		X	X	X	X	X	X	X		42
RB8	Raw	X		X	X	X	X	X	X	X	X	48
G1	GNS		X	X								6
G2	GNS		X	X	X							12
G3	GNS		X	X	X	X						18
G4	GNS		X	X	X	X	X					24
G5	GNS		X	X	X	X	X	X				30
G6	GNS		X	X	X	X	X	X	X			36
G7	GNS		X	X	X	X	X	X	X	X		42
G8	GNS		X	X	X	X	X	X	X	X	X	48
GB1	GNS	X		X								6
GB2	GNS	X		X	X							12
GB3	GNS	X		X	X	X						18
GB4	GNS	X		X	X	X	X					24
GB5	GNS	X		X	X	X	X	X				30
GB6	GNS	X		X	X	X	X	X	X			36
GB7	GNS	X		X	X	X	X	X	X	X		42
GB8	GNS	X		X	X	X	X	X	X	X	X	48
GSV1	GSEV		X	X								6
GSV2	GSEV		X	X	X							12
GSV3	GSEV		X	X	X	X						18
GSV4	GSEV		X	X	X	X	X					24
GSV5	GSEV		X	X	X	X	X	X				30
GSV6	GSEV		X	X	X	X	X	X	X			36
GSV7	GSEV		X	X	X	X	X	X	X	X		42
GSV8	GSEV		X	X	X	X	X	X	X	X	X	48
GSVB1	GSEV	X		X								6
GSVB2	GSEV	X		X	X							12
GSVB3	GSEV	X		X	X	X						18
GSVB4	GSEV	X		X	X	X	X					24
GSVB5	GSEV	X		X	X	X	X	X				30
GSVB6	GSEV	X		X	X	X	X	X	X			36
GSVB7	GSEV	X		X	X	X	X	X	X	X		42
GSVB8	GSEV	X		X	X	X	X	X	X	X	X	48
GSM1	GSEM		X	X								6
GSM2	GSEM		X	X	X							12
GSM3	GSEM		X	X	X	X						18
GSM4	GSEM		X	X	X	X	X					24
GSM5	GSEM		X	X	X	X	X	X				30
GSM6	GSEM		X	X	X	X	X	X	X			36
GSM7	GSEM		X	X	X	X	X	X	X	X		42
GSM8	GSEM		X	X	X	X	X	X	X	X	X	48
GSMB1	GSEM	X		X								6
GSMB2	GSEM	X		X	X							12
GSMB3	GSEM	X		X	X	X						18
GSMB4	GSEM	X		X	X	X	X					24
GSMB5	GSEM	X		X	X	X	X	X				30
GSMB6	GSEM	X		X	X	X	X	X	X			36
GSMB7	GSEM	X		X	X	X	X	X	X	X		42
GSMB8	GSEM	X		X	X	X	X	X	X	X	X	48

from the ratio of past forecast and observation pairs, weighted by an e-folding time τ:

$$DMB_t = \frac{\tau - 1}{\tau} DMB_{t-1} + \frac{1}{\tau} \frac{F_{t-1}}{O_{t-1}}. \tag{3.1}$$

Here, $DMB_{\text{t-1}}$ is the bias correction factor from the previous day weighted by $(\tau - 1)/\tau$. I used a τ of 30 days because it was found to be the optimum training period when averaged over all locations (Appendix A). The ratio of the mean of the previous day's forecasts ($F_{\text{t-1}}$) and observations ($O_{\text{t-1}}$) is weighted by $1/\tau$. The result is that the influence of the most recent forecast-observation pairs decreases with an e-folding time of τ. This recursive bias correction minimizes the need to store an ever-growing dataset for training.

The factor DMB_{t} is calculated for each individual ensemble member and then applied to the raw ensemble-member forecast F_t to produce a bias-corrected ensemble-member forecast $\hat{F}_t$:

$$\hat{F}_t = \quad \frac{F_t}{DMB_t}. \tag{3.2}$$

Eqs. (3.1) and (3.2) utilize data such as would be available in a true operational setting—current and future observations are not available when the forecast is made.

3.2.3 Choice of Uncertainty Model

The method used to convert the distribution of raw or bias-corrected ensemble members to a probability distribution is called the uncertainty model. The uncertainty model must not assign any probabilities to wind speeds below zero.

Five uncertainty models are evaluated, four of which are shown in Table 3.1. The first method, referred to as Raw (Table 3.1), uses the empirical distribution of ensemble members to assign probability. This is a basic method that could be used by those producing probabilistic forecasts over areas where observations do not exist.

The second uncertainty model, referred to as GNS (Table 3.1) assumes the form of a Gaussian distribution dressed about the ensemble mean. This Gaussian forecast probability distribution ($\mathcal{N}_t$) is described at any time t by

$$\mathcal{N}_t(\bar{F}_t, \sigma_t^2). \tag{3.3}$$

Here, $\overline{F_t}$ is the bias-corrected ensemble mean calculated from (3.1) and (3.2), where each member is equally weighted. The variance, σ_t^2, is adaptively estimated from the square of the past forecast errors calculated from the ensemble mean, weighted by the same 30-day e-folding time

used in (3.1).

The third uncertainty model, GSEV (Table 3.1), scales the variance of the Gaussian distribution based on the ensemble variance. For this method, variance is calculated using a linear regression of the form:

$$\sigma_t^2 = mx + b \tag{3.4}$$

In Eqn. (3.4), *x* is the raw or bias-corrected ensemble variance, and constants *m* and *b* are determined by regression against the square of the past errors (calculated from the bias-corrected ensemble mean). When used in this manner, the GSEV method could capitalize on theorized spread-skill relationships (Wilks, 2011; Grimit and Mass, 2002). Namely, when ensemble variance is smaller (larger), the spread of the Gaussian distribution (Eqn. 3.4) could be smaller (larger), assuming a relationship exists. Generally, linear correlation is used to quantify spread-skill relationships, with values over 0.6 considered strong relationships (Grimit and Mass, 2007).

The fourth uncertainty model, GSEM (Table 3.1), scales the Gaussian distribution by the ensemble mean. For this method, *x* in Eqn. (3.4) represents ensemble mean. This method could allow the uncertainty model to have larger (smaller) spread when the ensemble-mean-forecast wind speed is high (low). For the GSEM and GSEV uncertainty models, the regressed variables in Eqn. (3.4) (e.g., the ensemble mean or the ensemble variance, and the past squared errors) are also adaptively updated with an e-folding time of 30 days.

Fig. 3.2 illustrates how these three Gaussian-based methods could differ for an idealized three-member ensemble. I posit that the GSEV and GSEM methods could allow for sharper probabilistic forecasts than the GNS method. For each of these distributions, I also test for probabilistic forecast improvements resulting from adding more PBL schemes to the ensemble, and through ensemble-member bias correction. An overview of the tests performed is given in Table 3.1.

A fifth uncertainty model, based on a new distribution (which I call the pq distribution), is also tested. It allows for optimization of kurtosis to better match the distribution of past forecast errors (shown later). Probabilistic-forecast results using this model did not improve on those from the Gaussian-based models. The authors felt it important to show the results of this experiment, but details will be relegated to Appendix B.

3.2.4 Verification Metrics

As my focus in this chapter is probabilistic forecasting, I concentrate forecast evaluation on distributions-oriented verification metrics, which use the joint distribution of forecasts and observations (Wilks, 2011).

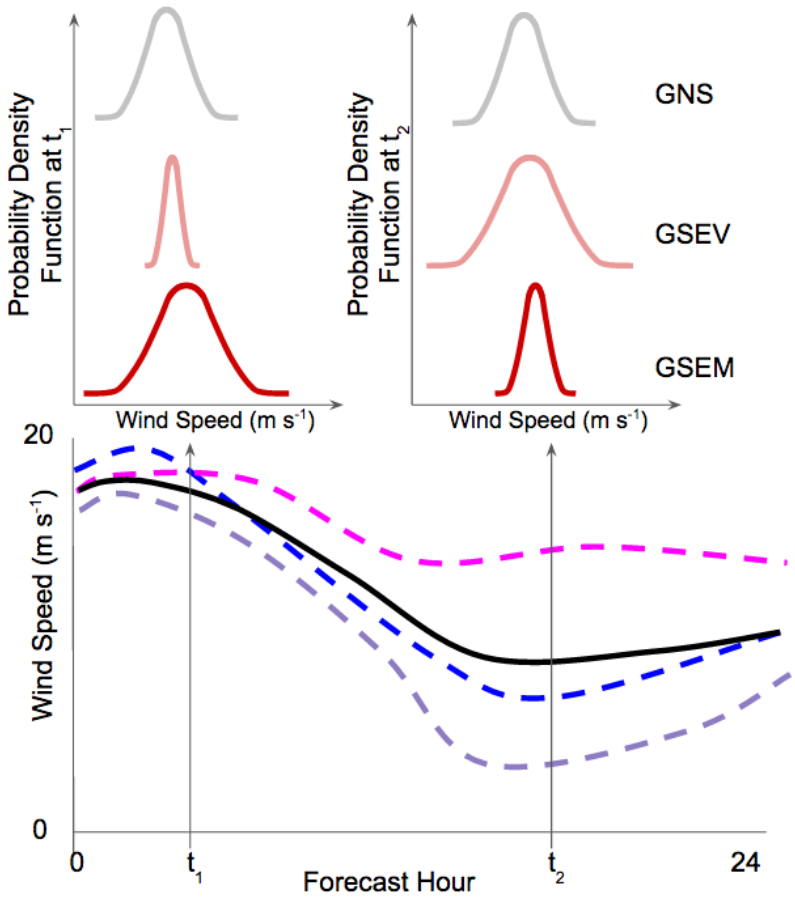

Figure 3.2: Illustration of ensemble-forecast meteogram and Gaussian-based uncertainty models (above the meteogram). Individual ensemble members are indicated by the colored, dashed curves and the ensemble mean by the black curve. Each forecast time has a probability density function based on the Gaussian distribution. The grey curve indicates a Gaussian distribution that does not scale (GNS) during the forecast. The pink curve is a Gaussian distribution that scales with ensemble variance (GSEV), while the red curve scales with the ensemble mean (GSEM).

These types of metrics provide advantages over deterministic measures when assessing probabilistic forecasts. Because they evaluate the full probabilistic-forecast distribution, they provide insight into how well forecast uncertainty is represented by the forecast probability distribution. Accurate representation of forecast uncertainty allows for optimum decision making and cost minimization (Pinson et al., 2006; Juban et al., 2007; McCollor and Stull, 2008c).

As with deterministic verification, the choice of verification metrics should be decided based on the purpose of the forecast and the needs of the end user. Gneiting et al. (2005, 2007) and Juban et al. (2007) propose that ideal probabilistic forecasts achieve calibration while maximizing forecast sharpness (both are described in more detail later). For this I use the Probability Integral Transform (PIT) histogram, reliability diagram, and continuously ranked probability score (CRPS). Skill scores are also used to quantify improvements in verification metrics of a test forecast configuration over that of a reference forecast (Wilks, 2011).

Calibration is a measure of how well forecast probability represents the actual frequency of event occurrence (Wilks, 2011; Nipen and Stull, 2011; Gneiting et al., 2007, 2005). The PIT histogram provides a concise method of evaluating ensemble probabilistic calibration. PIT values are calculated by finding which forecast percentile matches the associated observation for individual forecast-observation pairs. A histogram is generated by counting the frequency of occurrence of observations in each forecast percentile bin. This histogram should be flat for a calibrated forecast, and the deviation from flatness can be used to gauge calibration as was shown in Nipen and Stull (2011).

When the PIT histogram is not flat, it can indicate two things: 1) the forecast is under- or over-dispersive (needs calibration), or 2) the forecast has bias. Both of these situations can be fixed by calibration methods that adjust the probability distribution, resulting in flatter PIT histograms.

Gneiting et al. (2007) highlight that while the use of PIT histograms is prudent to address forecast calibration, it is hardly the only important aspect of a probabilistic forecast. They mention that forecasts that maximize sharpness, while maintaining calibration, are best. In addition, forecast calibration may not hold true over a subset of events. For wind-energy applications, sharper probabilistic forecasts (when also probabilistically calibrated) result in more efficient energy reserve resource planning and market trading opportunities because uncertainty (typical error from the ensemble mean) in the forecast is reduced (Juban et al., 2007). Because even forecasts based on climatology can be probabilistically calibrated, forecasts must be sharper than climatology to be considered skillful.

The observations from a previous 15-day window are used to define climatology at each hour of the day. The choice of a 15-day window is two-fold: (1) limited observations are available with only a year-long dataset, and (2) a shorter definition of climate (over that which averages all days

of the year) could better represent seasonality. I do not use any observations in the future to define climate as this may cause the climate-forecast skill to falsely appear better.

Reliability diagrams allow for the assessment of forecast calibration for a subset of events given a threshold (Wilks, 2011). For this study, wind-speed thresholds of 5, 15, and 20 m s^{-1} are used to represent low winds, turbine-rated winds, and high winds, respectively. Turbine-rated speeds are those at which wind turbines start to generate maximum power output (i.e., increasing winds will no longer increase power output). This speed can differ between turbine models.

A sharpness histogram is located within the reliability diagram. Sharpness is a function of only the forecast, and is a measure of the width of a forecast probability distribution (Gneiting et al., 2007). Given a forecast threshold, it is a count of how many times this threshold value lands in each probability bin. Forecasts that most often place this threshold near either the 0^{th} or 100^{th} percentiles are said to be sharp, while forecasts that frequently place the forecast near the 50^{th} percentile are not sharp. Narrower forecast probability distributions are sharper. While a sharper forecast is useful because it is more decisive, caution must be used to make sure that sharp forecasts are also probabilistically calibrated (i.e., a falsely confident forecast is not useful).

The CRPS is analogous to the MAE, but for probability. It is a measure of how well probability is assigned around each observed value (Hersbach, 2000). A lower CRPS indicates a forecast that better assigns probability with respect to actual event occurrence (i.e., it assigns higher probability for the event). A perfect forecast will have a CRPS of 0. The CRPS reduces to the MAE for deterministic forecasts (the cumulative probability distribution takes the form of a step function), which makes it possible to compare ensemble and deterministic forecasts in a probabilistic sense. In addition, the CRPS provides a strictly proper scoring metric of assessing both forecast calibration and sharpness in a single score (Gneiting et al., 2005). Strictly proper scores promote honesty in the sense that forecasters can not hedge their forecasts to achieve a better verification score (Wilks, 2011). The CRPS skill score is calculated by comparing the CRPS score of any probabilistic forecast to that of a reference probabilistic forecast.

Lastly, I will refer to root mean square error (RMSE) as a deterministic measure of forecast accuracy and also relate it to probabilistic-forecast sharpness, since RMSE is directly related to the spread of the Gaussian uncertainty model.

3.3 Results and Discussion

3.3.1 Raw Ensemble Distribution as a Probability Distribution

Tests R1-R8 (Table 3.1) are designed to quantify improvements in probabilistic calibration due to adding more PBL schemes to a short-range ensemble forecast in complex terrain. Starting with the best PBL scheme (ACM2, test R1), additional PBL schemes were added to the ensemble in tests R2-R8. Each additional PBL scheme added six total members to the ensemble, comprised of three grid sizes and two ICs. R1-R8 used the binned, raw-ensemble distribution as the forecast-probability distribution. Tests RB1-RB8 followed the same method, but used ensemble-member bias correction.

Fig. 3.3 shows the improvement in probabilistic calibration resulting from adding up to eight PBL schemes (i.e., improvement over R1). For the raw-ensemble distribution, forming a multi-PBL scheme, multi-IC, and multi-grid ensemble leads to large improvements in probabilistic calibration over an ensemble consisting of just a single PBL scheme (six members). For example, using 2 PBL schemes (12 members; R2, Table 3.1) results in 9% calibration improvement over 1 PBL scheme (R1, Table 3.1). A statistically significant 19% improvement in calibration (relative to R1, Table 3.1) results from using all eight PBL schemes (R8, Table 3.1). Throughout this chapter, I refer to statistical significance as exceeding the $p = 0.05$ significance level in a Student's t-test.

Once bias correction is applied, using all eight PBL schemes (RB8, Table 3.1) provides a 27% improvement in calibration (relative to RB1, Table 3.1). A 14% improvement is found when using only two PBL schemes (RB2, Table 3.1). Differences in forecast calibration between RB1 and RB 8 are statistically significant. Additionally, improvements in calibration through the application of bias-correction also pass statistical significance testing (e.g., comparing tests R1-R8 to RB1-RB8).

Even though R8 and RB8 were the best within their respective group of experiments, their PIT histograms (Fig. 3.4) show that the binned, raw-ensemble distribution is under-dispersive, even with a 48-member ensemble. The probability forecast is too confident and events too-often fall at the extremes of the distribution. To fix this, a better uncertainty model is needed, and I address this through the next set of tests.

3.3.2 Prescribed Probability Distributions Dressed on the Ensemble Mean

Tests G1-G8 (Table 3.1) use the GNS uncertainty model dressed about the raw ensemble mean with spread (variance) based on the past squared errors of the ensemble mean. Using this method with the best overall PBL scheme (G1), or all eight PBL schemes (G8), results in a biased forecast (Fig. 3.5). To remove the bias, I performed bias correction on each ensemble member in tests GB1-GB8.

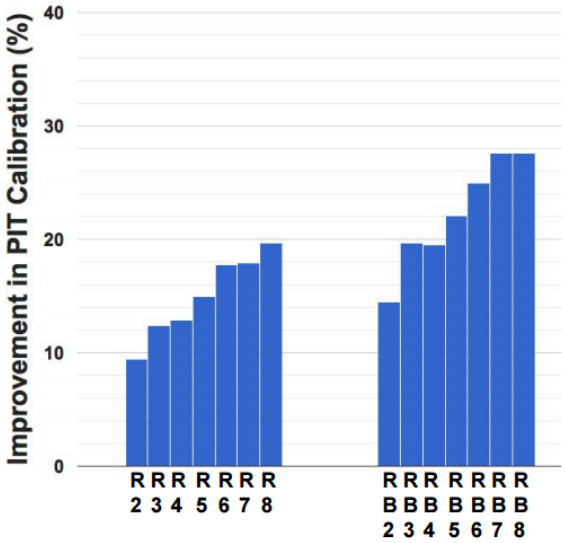

Figure 3.3: Improvement in PIT histogram calibration for one year of hourly wind forecasts by using more PBL schemes in the ensemble for the raw ensemble distribution (R2-R8, Table 3.1) and bias-corrected ensemble distribution (RB2-RB8) uncertainty models. Improvement is based on reduction in deviation between bins of the PIT histogram (such as for the PIT histogram in Fig. 3.4). Larger improvement is better.

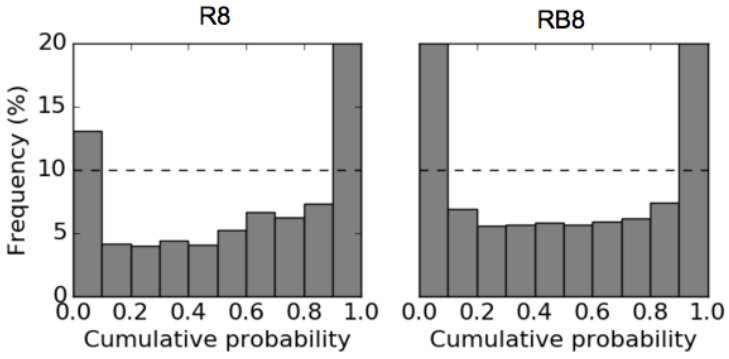

Figure 3.4: PIT histograms indicating an under-dispersive ensemble for tests R8 and RB8. Under-dispersion occurs when observed events fall too often at or outside the extremes of the ensemble forecast distribution. Flatter PIT histograms are better.

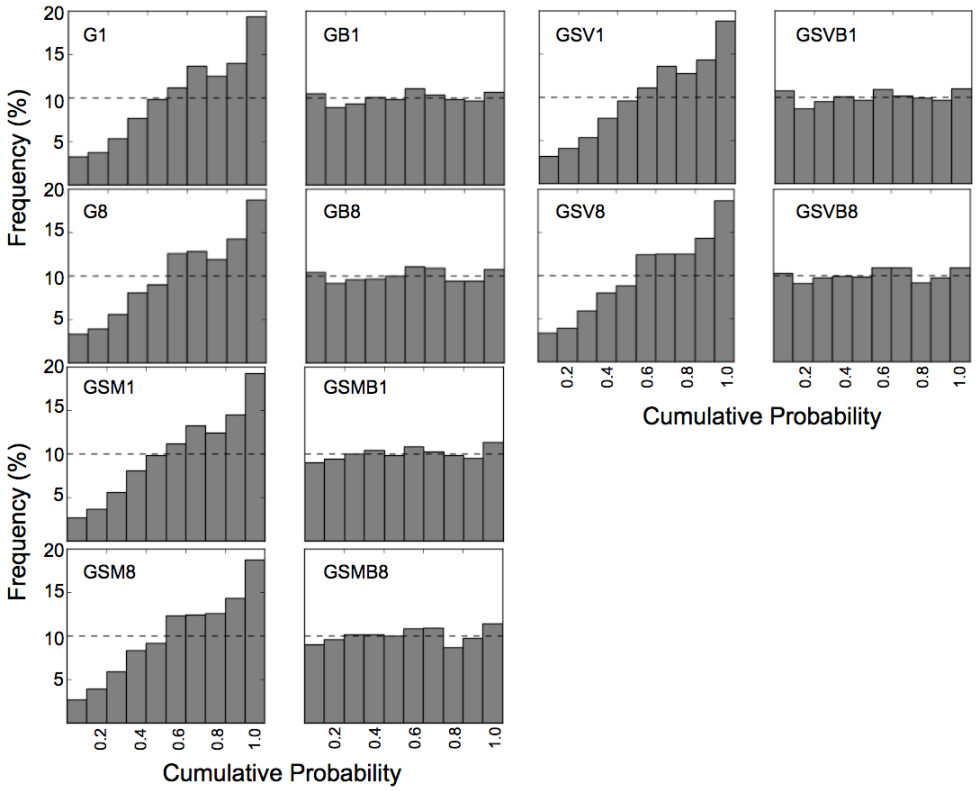

Figure 3.5: PIT histograms comparing the results of the three Gaussian-based distributions for the six- and 48-member ensembles, prior to and after bias correction. Labels are the test name (Table 3.1). Flatter PIT histograms are better (closer to the horizontal dashed line).

These also used the GNS uncertainty model, but centered the forecast probability distribution on the bias-corrected ensemble mean. The removal of the bias resulted in a nearly flat PIT histogram, indicating a calibrated forecast (GB1 and GB8 in Fig. 3.5). A separate calibration step is therefore not needed.

Next, I tested two additional Gaussian-based uncertainty models that allow the distribution to scale by either the ensemble variance or the ensemble mean. First, I scaled the distribution by the ensemble variance in tests GSV1-GSV8 (no ensemble-member bias-correction, Table 3.1) and GSVB1-GSVB8 (with ensemble-member bias correction, Table 3.1). Tests GSV1-GSV8 and GSVB1-GSVB8 show nearly identical results to using a non-scaling Gaussian uncertainty model (GNS). PIT histograms for GSV1 and GSV8 show a biased probabilistic forecast (GSV1 and GSV8; Fig. 3.5), that result in a well-calibrated forecast once ensemble-member bias correction is applied (GSVB1 and GSVB8; Fig. 3.5). Tests GSM1-GSM8 and GSMB1-GSMB8 yielded similar re-

sults (Fig. 3.5) for scaling the Gaussian distribution by the ensemble mean. When using any of the Gaussian-based uncertainty models, improvements in calibration through the use of the bias-corrected ensemble members passed significance testing.

Fig. 3.6 shows the actual distributions of wind-speed forecast errors with respect to the bias-corrected ensemble mean. Each histogram is for a different wind farm. By inspection, these distributions do not have a Gaussian shape. These distributions have more probability near the center (kurtosis = 3.5 to 6.6) than a Gaussian (kurtosis = 3), but they are nearly symmetric (skewness of less than 0.5, close to the Gaussian zero skewness). This motivates two questions: a) what aspects of the Gaussian distribution allow it to work so well to produce calibrated probabilistic forecasts, and b) can I find a different theoretical distribution that is a better fit to the observed error distributions? I address question one next, and address question two in section 3.3.6 and in Appendix B.

Perhaps all that is needed to produce calibrated forecasts is to have almost any distribution shape that has central tendency, spread about the mean, and is symmetric. The fact that the Gaussian distribution has tails that unphysically extend to $\pm\infty$ appears to be irrelevant to its ability to satisfactorily dress the ensemble mean. So the next question is how much spread in the distribution is needed?

First, I investigate whether larger forecast errors correspond to a) greater spread among the raw-ensemble members, or b) larger ensemble-mean-forecast winds. To do this, I calculated the coefficient of determination (r^2) for the linear relationship between the squared error of the bias-corrected ensemble mean, and either the bias-corrected ensemble variance or bias-corrected ensemble mean. I found the relationship between the bias-corrected ensemble variance and the squared error had r^2 values of 0.027 or less. This means that only 2.7% of the error variance can be explained by the linear relationship between the ensemble variance and squared-error magnitude (e.g., the spread-skill relationship does not apply at these four locations in complex terrain).

The corresponding analysis comparing the bias-corrected ensemble mean with the squared error magnitude resulted in r^2 values no larger than 0.053. The implications of this are that the Gaussian distributions used in GB1-GB8, GSVB1-GSVB8, and GSMB1-GSMB8 are similar, since the scaling relationships are weak. At some of the four locations, the differences between the GNS, GSEV, and GSEM distributions are statistically insignificant. Because of this, the next analyses will focus on the GNS distribution used in tests GB1-GB8.

An r^2 value of 1 is not expected or feasible, as several studies have noted (Whitaker and Loughe, 1998; Grimit and Mass, 2007; Hopson, 2014). Some of those studies, and Wang and Bishop (2003), propose potential methods for extracting a stronger spread-skill relationship. Exploring those methods is beyond the scope of this study. Hopson (2014), however, concludes that in the case of a weak-spread relationship, an invariant spread, derived from ensemble-mean error, may be used in-

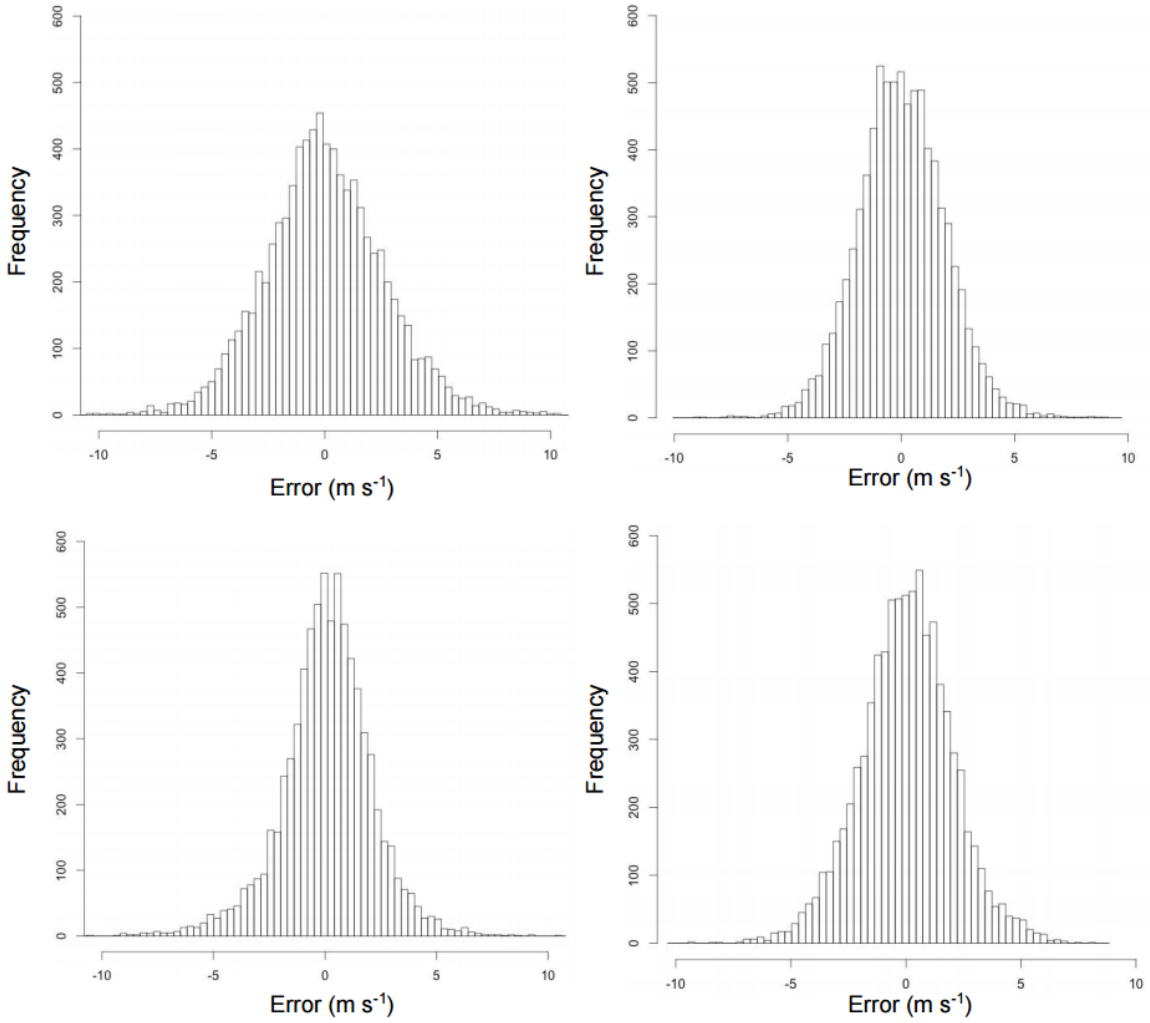

Figure 3.6: Actual wind-speed forecast-error distribution about the ensemble mean in test GB8 for each of the four wind farms.

stead. I reach the same conclusion here.

Having addressed calibration, I now look at sharpness. For the GNS uncertainty model, spread is only a function of the past error of the bias-corrected ensemble mean (3). The smaller the error (which can be measured by RMSE), the narrower the prescribed forecast probability distribution, and the sharper the forecast. I address this in the next section.

3.3.3 Use of Ensembles to Improve Forecast Sharpness

The benefit in using ensembles in the short-term forecast horizons has been debated, especially for near surface variables like hub-height winds (Mass et al., 2002; Eckel and Mass, 2005; Warner, 2011). I approach this question by looking at the effects of an ensemble on forecast sharpness. When using the GNS uncertainty model, forecast systems that have lower RMSE must have smaller probabilistic spread (improved sharpness).

Fig. 3.7 shows the average RMSE for each ensemble member, as well as the six-member bias-corrected ensemble from the best PBL scheme (GB1, Table 3.1) and the full 48-member bias-corrected ensemble (GB8, Table 3.1). While probabilistic calibration could be achieved by dressing a Gaussian distribution around any of these *individual* ensemble members, the sharpest forecasts are produced by dressing the distribution on the ensemble means from GB1 or GB8, which have the lowest RMSE. While I show that the best-PBL ensemble (GB1) and full-PBL ensemble (GB8) have equivalent sharpness, it is not possible to know which PBL scheme is best apriori, unless it has been shown to consistently perform best over an extended period of time. Nonetheless, these results indicate that larger ensembles do not necessarily perform better in short-term forecasts than smaller, selective ensembles. This is an important finding for short-term wind planning as it reduces computational cost, while still providing sharp, probabilistically calibrated forecasts.

3.3.4 Probabilistic Forecasts for Wind Events

While the bias-corrected ensemble forecasts show statistical calibration for the data used in the training period, they are not necessarily calibrated over subsets of this training data. Forecasts of rare events, which in this case are high-wind events, can be uncalibrated. To assess calibration over different wind-speed subsets, I provide the reliability diagrams using event thresholds of 5, 15, and 20 m s^{-1} for tests RB1, RB8, GB1, GB8, and also for a GNS distribution dressed on the single-best bias-corrected ensemble member (ACM2 12-km GNS; not in Table 3.1) (Fig. 3.8).

RB1 and RB8 are under-dispersive for an event threshold of 5 m s^{-1}, while GB1, GB8, and the ACM2 12-km GNS are calibrated. However, for the 15 and 20 m s^{-1} thresholds, all forecasts exhibit poor calibration, typically overforecasting probabilities. The horizontal dashed line in Fig. 3.8 represents the climatological frequency of events above the given threshold. Out of a total sample of 32,442 observations, 20,455 (63%) are above 5 m s^{-1}, 603 (1.8%) are above 15 m s^{-1}, and 35 (0.1%) are above 20 m s^{-1}. For the 15 m s^{-1} threshold, forecasts have near-zero skill (close to red dashed line, Fig. 3.8). For the 20 m s^{-1} threshold, the probabilistic forecast has no skill. The lack of probabilistic-forecast calibration for high-wind events is likely because these events comprise a small percentage of the training data, and forecast biases for these events may differ from that of the

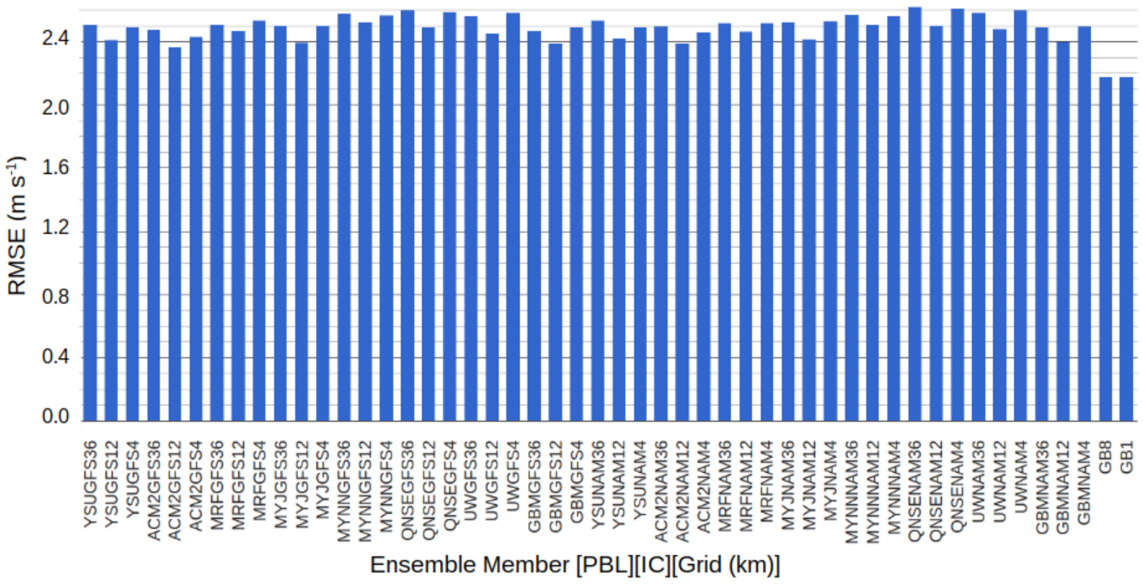

Figure 3.7: Annual RMSE for each bias-corrected ensemble member, the six-member ensemble in test GB1, and the 48-member ensemble in test GB8, averaged over all four wind sites. Ensemble members are named under the convention [PBL][IC][Grid (km)]. Smaller RMSE is better.

bulk of the training data. Separate calibration for different thresholds may alleviate this issue.

Insets within the panels of Fig. 3.8 are sharpness diagrams for the event thresholds. These diagrams indicate how often the event threshold falls in each forecast probability bin. Sharper forecasts will more frequently give event probabilities closer to 0% or 100%, while less sharp forecasts give event probabilities closer to 50%. For all thresholds, I see that the raw bias-corrected-ensemble distribution (RB1 and RB8) produces the sharpest forecasts. However, sharpness without calibration is worthless. The high sharpness is a result of the under-dispersive, over-confident nature of raw-ensemble distributions. Each of the GNS distributions (GB1, GB8, ACM2 12-km GNS in Fig. 3.8) produce sharp forecasts for each of these event thresholds.

3.3.5 Summary of Skill vs. Computation Cost

As a summary of the forecast improvements obtained by using probabilistic-based forecasts, I plot the CRPS skill score, which allows a comparison deterministic and ensemble forecasts in a probabilistic sense (Fig. 3.9). I use the worst-performing deterministic forecast (the bias-corrected 36-km QNSE forecast) as the reference forecast to gauge short-term probabilistic-forecast improvements. Using the best deterministic forecast (12-km ACM2 forecast) provides an 11% improvement in the

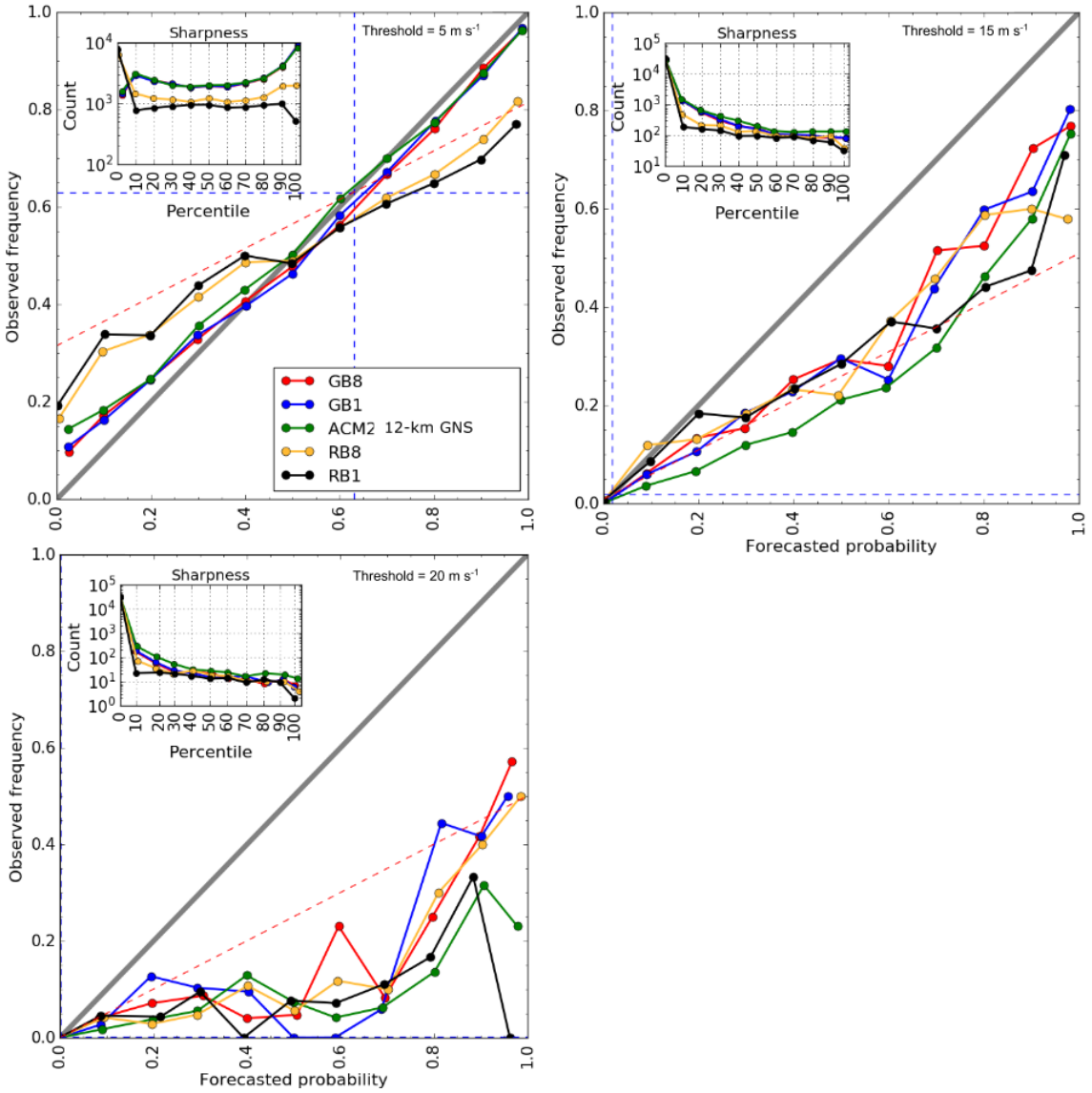

Figure 3.8: Reliability diagrams for tests RB1, RB8, GB1, GB8, and the GNS distribution dressed over the bias-corrected 12-km ACM2 deterministic forecast initialized off the GFS (ACM2 12km GNS) for event thresholds of 5, 15, and 20 m s^{-1}. Reliability curves closer to the diagonal 1:1 line (thick grey) are better (i.e., are more calibrated). The blue dashed line represents the climatological event threshold frequency. The red dashed line represents the no-skill line. The inset figure is the sharpness histogram and is a count of the number of occurrences the event threshold is forecast in each probability bin (from 0^{th} to 100^{th} percentile in 10% increments.) Sharper forecasts assign mostly 0^{th} and 100^{th} percentiles while non-sharp forecasts issue mainly 50^{th} percentile forecasts.

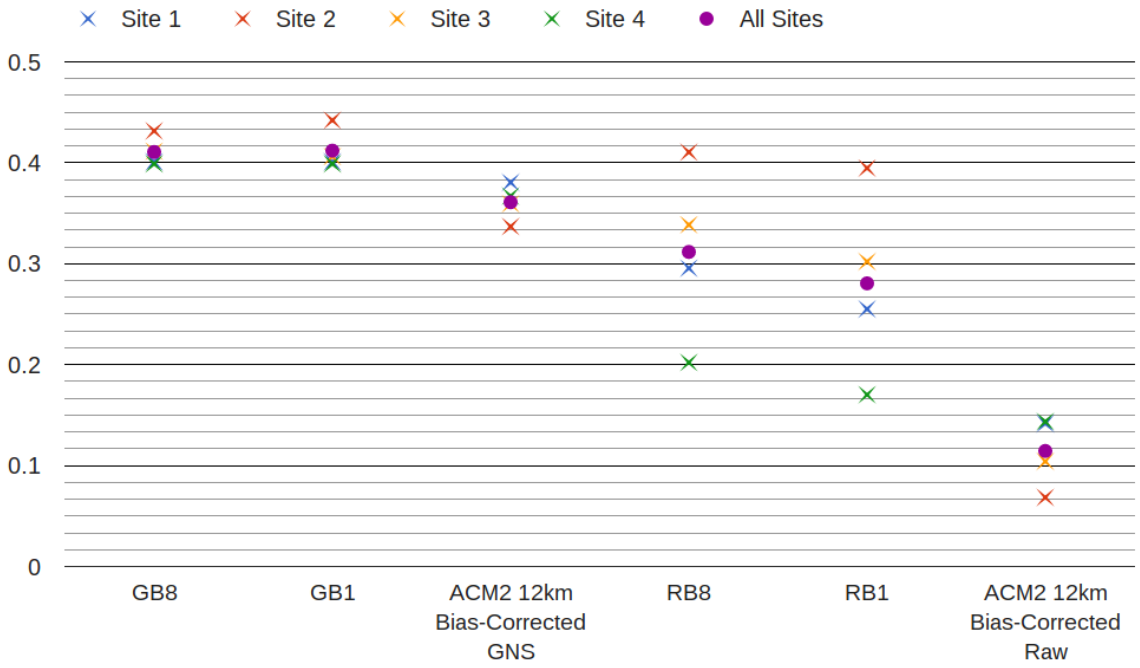

Figure 3.9: CRPS skill scores for test RB1, RB8, GB1, GB8, the ACM2 12-km GNS, and the bias-corrected deterministic 12-km ACM2 forecast initialized off the GFS. Skill is calculated relative to the worst performing deterministic forecast, the 36-km QNSE scheme initialized with the NAM. Larger CRPS Skill Score is better.

CRPS, averaged over all stations. A six-member ensemble forecast based on the best PBL scheme using the raw-ensemble distribution as the uncertainty model (RB1, Table 3.1) improves the CRPS by an additional 17% (or 28% over the worst deterministic forecast). Using the raw distribution with the full 48-member ensemble (RB8, Table 3.1), provides an additional improvement in CRPS of 3%, or a 31% improvement over the worst deterministic forecast.

Fig. 3.9 shows the large improvement that is gained by dressing the ensemble mean with an uncertainty model. The bias-corrected ACM2 12-km GNS provides a better probabilistic forecast than that of RB8 (on average), at a fraction of the computational cost. CRPS is improved over RB8 by an additional 5%. Test GB1, which uses all the ACM2-based forecasts, provides the most CRPS skill, with a 41% improvement over the worst deterministic forecast, and an additional 5% improvement over the best deterministic forecast dressed by the GNS distribution. No further improvement is gained with a larger ensemble for this short-term forecast (GB8, Fig. 3.9), while computational cost increases. All ensemble forecasts in Fig. 3.9 (e.g., tests RB1, RB8, ACM2 12-km GNS, GB1, and GB8) have probabilistic skill (increased sharpness) over a climatology forecast while the two

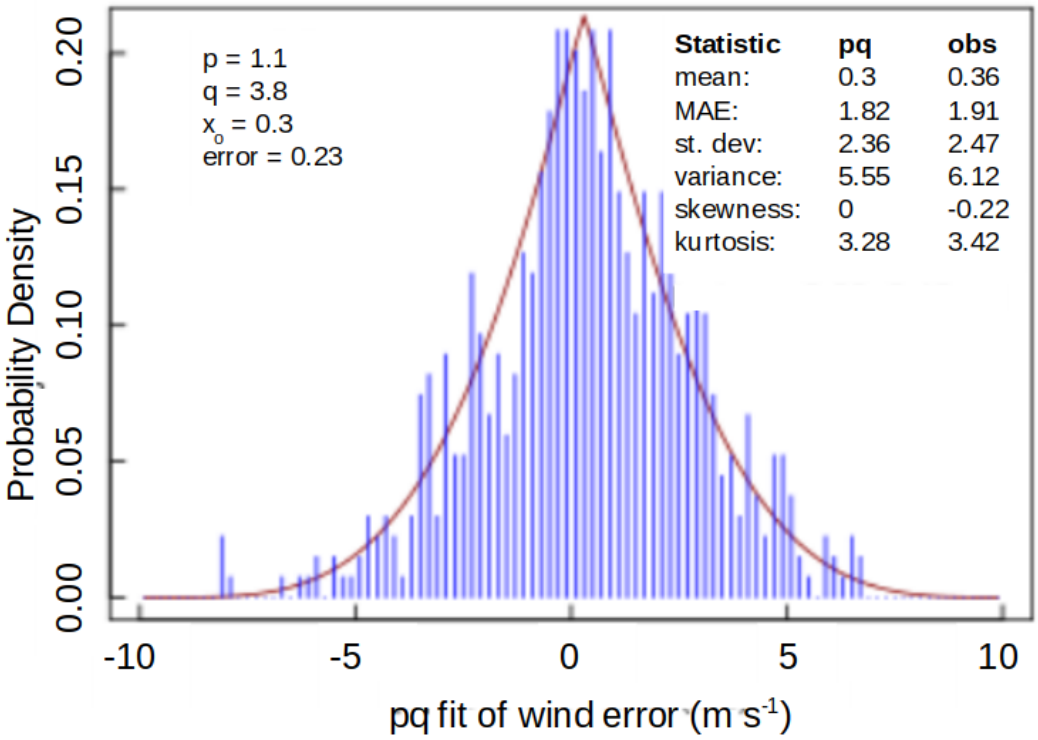

Figure 3.10: Typical 30-day training period past error distribution (vertical bars) fit by the pq probability distribution (smooth curve). Bins are every 0.2 m s^{-1}.

deterministic forecasts do not (the ACM2 12-km deterministic and the 36-km QNSE deterministic reference forecasts; not explicitly shown).

3.3.6 Shape of the Prescribed Distribution

In section 3.3.2, I suggested that the exact shape of the prescribed distribution used to dress the ensemble mean is less important than the distribution's gross attributes: central tendency, spread, and symmetry. As a preliminary test of this hypothesis, I use a new distribution, called the pq distribution. This distribution can better fit the shape of the observed wind-forecast-error distribution (Fig. 3.10). The pq distribution is symmetric and bounded; see Appendix B for details.

When Fig. 3.9 was re-calculated using CRPS from the best-fitting pq distributions, the results (not shown) were nearly identical to those from the GNS distribution. The lack of improvement in CRPS through the use of a better-fitted distribution (pq) is preliminary evidence that the distribution's gross attributes are more important, rather than the exact shape. Although the wind-forecast errors were nearly symmetric for the four wind farms studied here, there might be other situations

where the forecast-error distributions may be asymmetric or even multi-modal (i.e., not Gaussian; not pq), and other distributions would better describe the errors.

3.4 Conclusions and Future Work

I detailed methods to produce calibrated hub-height wind-speed forecasts over complex terrain from a multi-PBL-scheme, multi-IC, multi-grid-length WRF ensemble. Tests evaluated the effects of having multiple PBL schemes in the ensemble, bias correction, and uncertainty model choice. Probabilistic forecast performance was evaluated based on improvements in the PIT histogram, reliability diagram, sharpness, CRPS, and RMSE.

For the binned raw-ensemble distribution, increasing the number of PBL schemes available improved forecast calibration. However, this distribution was under-dispersive, and remained probabilistically uncalibrated even after removing the bias from each individual ensemble member.

To improve probabilistic calibration, I tested if a prescribed uncertainty model could be used to better represent forecast uncertainty. Three Gaussian-based uncertainty models and one pq-model were evaluated. The first had no scaling; the distribution was based on the past error. The second and third Gaussian-based uncertainty models attempted to scale the Gaussian distribution based on a linear regression of either the ensemble variance or the ensemble mean against the square of the past errors from the ensemble mean. Using any Gaussian uncertainty model without individual member bias correction yielded biased PIT histograms. However, bias correcting each individual ensemble member resulted in probabilistically-calibrated forecasts for all three Gaussian-based uncertainty models.

When using the prescribed uncertainty models, using additional PBL schemes did not result in improved probabilistic calibration. Therefore, large ensembles are not necessary to produce probabilistically calibrated forecasts. Instead, the addition of multiple PBL schemes improved sharpness. Although the RMSE (and thus sharpness) was similar when using the full 48-member ensemble and the reduced 6-member ensemble using only the ACM2 PBL scheme, the single best PBL scheme may not be known apriori. If the best PBL scheme is known, a selective ensemble can be used to save computation costs.

Additionally, the ensemble-mean RMSE, and thus sharpness, may be improved through the use of ICs produced by other agencies. Examples include the Canadian Global Deterministic Prediction System, the Fleet Numerical Meteorology and Oceanography Center Navy Global Environmental Model, the UK Metoffice Unified Model, and the European Centre for Medium Range Weather Forecasts Integrated Forecast System. The reason I suggest testing the use of these other sources is because the underlying data assimilation system used for both the GFS and the NAM is the same:

the Gridpoint Statistical Interpolation (Shao et al., 2016). Differences between ICs may be larger when the sources come from distinctly different agencies (those using different data-assimilation techniques).

My error analysis shows that the Gaussian distribution describes the forecast error well, but that forecast error has little relationship to either the ensemble mean or the ensemble variance. I also found that gross attributes (central tendency, spread, symmetry) of a prescribed distribution are more important than its exact shape.

Chapter 4

A Flux-Profile Relationship for Winds over Mountain Ridges

4.1 Introduction

The atmospheric surface layer (ASL) typically occupies the lowest 10% of planetary boundary layer (Stull, 1988). The ASL has a depth that ranges on the order of several 10s to over 100s of meters. The effects of the Coriolis force are negligible within this layer such that wind direction does not vary significantly with height. Additionally, turbulent fluxes of heat and momentum in this layer are approximately constant with height and play a large factor in the shape of vertical profiles of wind and temperature. The shapes of wind and temperatures profiles in the ASL are largely influenced by land-atmosphere interactions (Jiménez et al., 2012).

Wind-turbine hub heights (50- and 120-m AGL) are often located within the ASL. Therefore, accurate parameterization of turbulent fluxes in the ASL, and thus representation of vertical profiles of wind, are critical to the wind-energy community. Such parameterizations are commonly used for wind-resource estimates, which extrapolate near-surface (10-m) wind observations up to hub height. ASL schemes are also used within numerical models like the Weather Research and Forecasting model (WRF, Skamarock et al., 2008) to parameterize surface fluxes. However, most methods of parameterizing fluxes in the ASL (and thus the shape of vertical wind profiles), have been derived from observations over flat, homogeneous terrain.

Perhaps the most-widely used flux parameterization is Monin-Obukhov Similarity Theory (MOST, Monin and Obukhov, 1954; Panofsky, 1963). MOST parameterizes surface fluxes of wind and temperature, and thus ASL wind and temperature profiles, through the use of two dimensionless quantities: the non-dimensional wind shear ϕ_m, and the non-dimensional temperature gradient ϕ_h. They take the form:

$$\phi_m = \frac{kz}{u_*}\frac{\partial u}{\partial z}, \tag{4.1}$$

$$\phi_h = \frac{kz}{\theta_*}\frac{\partial\theta}{\partial z}, \tag{4.2}$$

where k is the von Karman constant, which has been shown to have a value of approximately 0.4 (Stull, 1988), z is height above the ground, u is wind speed, and θ is potential temperature. Variables u_* and θ_* are the friction velocity and temperature scales, respectively, and are related to the surface momentum (τ) and sensible heat (H) fluxes as follows:

$$\tau = \rho u_*^2 = \rho C_D u^2, \tag{4.3}$$

$$H = -\rho c_p u_* \theta_* = -\rho c_p C_H u(\theta - \theta_g), \tag{4.4}$$

where, ρ is air density, c_p is specific heat at constant pressure, and θ_g is potential temperature at the ground. Variables C_D and C_H represent bulk transfer coefficients of momentum and heat, respectively.

Eqns. (4.1) and (4.2) can be integrated with respect to height to diagnose the wind and temperature in the ASL when the variables u_* and θ_* are known:

$$u(z) = \frac{u_*}{k}[ln(\frac{z}{z_o}) - \psi_m], \tag{4.5}$$

$$\theta(z) - \theta_g = \frac{\theta_*}{k}[ln(\frac{z}{z_o}) - \psi_h]. \tag{4.6}$$

Eqn. (4.5) gives the logarithmic wind profile, where z_o is roughness length and ψ_m is the integrated similarity function for momentum. In Eqn. (4.6), ψ_h is the integrated similarity function for heat. The integrated similarity functions provide an adjustment to the standard logarithmic profile shape, with values dependent on the static stability. ϕ_m and ϕ_h, and their corresponding integrated functions ψ_m and ψ_h are empirically estimated from field data, as described next.

4.1.1 Empirical Formula for ϕ_m and ϕ_h

Several empirical functions for ϕ_m and ϕ_h have been derived from observations over homogeneous terrain, with perhaps the most well-known being the Businger-Dyer (BD, Businger et al., 1971; Dyer, 1974) relationships derived from observations at a 32-m tower in western Kansas. Other forms have been suggested by Webb (1970), Hicks (1976), Zhang and Anthes (1982), Beljaars and Holtslag (1991), Santoso and Stull (2001), and Arya (2001), among others. Optis et al. (2016) provide a summary of several existing similarity relationships and show that the largest profile differences at the Cabauw meteorological tower in Netherlands occur during statically-stable conditions.

Empirical functions for ϕ_m and ϕ_h generally take similar log-linear forms, such as:

$$\phi_{m,h} = [\alpha + \beta\zeta]^{\gamma}. \tag{4.7}$$

Here, ζ is the Monin-Obukhov stability parameter, z/L, where L is the Obukhov length. The α, β, and γ parameters are empirical constants derived from experimental data. Inclusion of the ζ parameter allows for a dependence on boundary-layer static stability.

For ϕ_m, most studies find a linear relationship during statically-stable regimes ($\zeta > 0$) with $\gamma = 1$. For example, the BD relationships (Businger et al., 1971) are as follows:

$$\phi_m = 1 + 4.7\zeta, \quad \text{for} \quad \zeta > 0, \tag{4.8}$$

$$\phi_m = 1, \quad \text{for} \quad \zeta = 0, \tag{4.9}$$

$$\phi_m = [1 - 15\zeta]^{-0.25}, \quad \text{for} \quad \zeta < 0, \tag{4.10}$$

and

$$\phi_h = 0.74 + 4.7\zeta, \quad \text{for} \quad \zeta > 0, \tag{4.11}$$

$$\phi_h = 0.74, \quad \text{for} \quad \zeta = 0, \tag{4.12}$$

$$\phi_h = 0.74[1 - 9\zeta]^{-0.5}, \quad \text{for} \quad \zeta < 0. \tag{4.13}$$

Once empirical constants have been derived from experimental data, the integrated similarity functions shown in (4.5) and (4.6) are found by

$$\psi_{m,h} = \int_0^{\zeta} [1 - \phi_{m,h}] \frac{\partial \zeta}{\zeta}. \tag{4.14}$$

For the BD relationships, integrated similarity functions for momentum and heat are

$$\psi_m = \psi_h = -4.7\zeta, \quad \text{for} \quad \zeta > 0, \tag{4.15}$$

$$\psi_m = \psi_h = 0, \quad \text{for} \quad \zeta = 0, \tag{4.16}$$

$$\psi_m = ln[(\frac{1+x^2}{2}) + (\frac{1+x}{2})^2] - 2tan^{-1}x + \frac{\pi}{2}, \quad \text{for} \quad \zeta < 0, \tag{4.17}$$

$$\psi_h = 2ln(\frac{1+x^2}{2}), \quad \text{for} \quad \zeta < 0, \tag{4.18}$$

where

$$x = (1 - 15\zeta)^{0.25}. \tag{4.19}$$

The most commonly used ASL scheme in the WRF model, the *MM5 surface layer scheme*, uses a modified form of the BD relationships and is detailed in Zhang and Anthes (1982) and Jimenez and Dudhia (2012). Hereafter, I refer to these as the ZA relationships. The ZA relationships divide stability into four regimes based on values of the bulk Richardson number (Ri_b). The regimes are (a) stable, laminar flow ($Ri_b > 0.2$), (b) damped mechanical turbulence ($0.2 > Ri_b > 0$), (c) neutral conditions ($Ri_b = 0$), (d) free convection and unstable conditions ($Ri_b < 0$), and have the following integrated similarity relationships:

$$a : \psi_m = \psi_h = -10ln(\frac{z}{z_o}), \tag{4.20}$$

$$b : \psi_m = \psi_h = -5Ri_b\frac{ln\frac{z}{z_o}}{1.1 - 5Ri_b}, \tag{4.21}$$

$$c : \psi_m = \psi_h = 0, \tag{4.22}$$

$$d : \psi_m = ln[(\frac{1+x^2}{2}) + (\frac{1+x}{2})^2] - 2tan^{-1}x + \frac{\pi}{2}, \;\; \psi_h = 2ln(\frac{1+x^2}{2}). \tag{4.23}$$

In Eqn. (4.23), x is the same as that in Eqn. (4.19).

For NWP models like WRF, Eqns (4.5) and (4.6) are typically rearranged to diagnose the values of u_* and θ_* by using the values of u and θ at the first model level and the ground, and using the values of the integrated similarity functions provided in Eqns. (4.20)-(4.23). Thus, flux-profile relationships play a key role in representation of surface stress and heat flux within NWP models.

4.1.2 Issues for Current Flux-Profile Relationships

Businger et al. (1971) and others note some scatter in their observational datasets during statically-stable regimes. Optis et al. (2016) attribute this scatter to the intermittent nature of turbulence during statically-stable conditions. Additionally, the large number of published similarity relationships suggest there is no unique solution, particularly for statically-stable surface layers.

While several studies have found good performance of MOST over flat, homogeneous terrain (e.g., those cited in section 4.1.1), the assumptions involved and added complexity of flows in mountainous terrain may limit its applicability there. Complex topography introduces terrain inhomogeneity that can induce internal boundary layers and mountain waves (Optis et al., 2016). Internal boundary layers occur when turbulent eddies generated over rough surfaces mix vertically and are advected over large distances by the mean horizontal wind (Stull, 1988). Wind and temperature profiles downstream of the roughness feature can be significantly affected.

Under stably stratified conditions, internal boundary layers in coastal regions have been shown to advect downstream by several 10s of km, leading to periods of enhanced mixing and rapid fluc-

tuations in wind speed (Dörenkämper et al., 2015). Over mountainous regions, one could expect similar internal boundary layers to be present due to surface heterogeneity. Additionally, dynamic and thermally-induced circulations like flow channeling, anabatic/katabatic winds, mountain/valley breezes, and drainage flows can affect wind and temperature profiles in complex terrain. Several of these features occur in statically-stable regimes.

One study by Rotach and Zardi (2007) found that boundary-layer features over valley locations in the Alps were consistently reproducible, suggesting that parameterization of turbulent structures is possible in regions of complex terrain. However, Nadeau et al. (2013), who apply MOST over a steep alpine slope in the Swiss Alps, find significant departures between observed wind and temperature profiles and those predicted by the BD relationships. The Nadeau et al. (2013) study was in a valley, where they found $\phi_m < 0$ during statically-stable regimes where slope flows caused winds to be faster near the surface. They also found ϕ_m was between 0 and 0.2 for unstable regimes. ϕ_h was found to be unrelated to stability. While Nadeau et al. (2013) did not have much success in applying current MOST along an alpine slope, the application of MOST to mountain ridge tops represents an existing knowledge gap. Such locations are becoming increasingly common for wind farms. Better similarity relationships for ridge tops will result in better wind resource estimates as well as hub-height wind-speed predictions from NWP models.

Further, tree canopies significantly affect wind and temperature profiles in regions of tall vegetation (Stull, 1988). Namely, vegetation canopies act like displaced roughness surfaces such that height z in Eqns. (4.1-4.23) must be adjusted by a zero-plane displacement distance d to account for a shifted reference plane (Stull, 1988; Arya, 2001). This results in Eqns. (4.1-4.2) and Eqns. (4.5-4.6) being rewritten, respectively, as:

$$\phi_m = \frac{k(z-d)}{u_*}\frac{\partial u}{\partial z}, \tag{4.24}$$

$$\phi_h = \frac{k(z-d)}{\theta_*}\frac{\partial \theta}{\partial z}, \tag{4.25}$$

and

$$u(z) = \frac{u_*}{k}[ln(\frac{z-d}{z_o}) - \psi_m], \tag{4.26}$$

$$\theta(z) - \theta_g = \frac{\theta_*}{k}[ln(\frac{z-d}{z_o}) - \psi_h]. \tag{4.27}$$

Also, in Eqns. (4.7-4.23), $\zeta = (z-d)/L$. The value of d is is related to the canopy height h, with a value of $d \approx 0.75h$ (Kaimal and Finnigan, 1994). The exact value of d is difficult to obtain, and requires direct flux measurements to do so. Primarily, d is not usually a constant and varies with canopy porosity (Garratt, 1992). Kaimal and Finnigan (1994) note that investigators who do not

have access to direct flux measurements (e.g., who must estimate u_* and θ_* from mean quantities of wind and potential temperature in vertical profiles) can minimize error by fixing the value of d based on estimates of h.

Several studies have found observed values of ϕ_m and ϕ_h to significantly deviate from those predicted by standard MOST (e.g. BD, ZA) in a region of the ASL known as the roughness sublayer, or transition layer (Raupach, 1979; Garratt, 1983; Högström et al., 1989; Garratt, 1992; Arya, 2001; Lee and Mahrt, 2005). The transition layer is a region of the ASL that spans from the ground up to approximately 3 times the canopy height (e.g., $3h$), and is characterized by enhanced turbulent mixing caused by flow impinging on the canopy top (Kaimal and Finnigan, 1994; Arya, 2001). Thus, wind turbines are sometimes within the transition layer. Values for ϕ_m and ϕ_h have been found to be reduced by 65-75% in the transition layer (Garratt, 1992; Arya, 2001), with the magnitude of reduction found to be independent of boundary-layer static stability. Canopy-corrected flux profiles are described by

$$\phi_m \phi_* = \frac{k(z-d)}{u_*}\frac{\partial u}{\partial z}, \tag{4.28}$$

$$\phi_h \phi_* = \frac{k(z-d)}{\theta_*}\frac{\partial \theta}{\partial z}, \tag{4.29}$$

where ϕ_* is the empirical reduction factor

$$\phi_* = \exp[-0.7(1 - \frac{z-d}{z_*})]. \tag{4.30}$$

In Eqn. (4.30), z_* is the height of the transition layer ($z_* \approx 3h$).

4.1.3 Goals of this Research

This research addresses the applicability of currently accepted MOST theory over mountain ridge-top locations with wind-energy applications in mind. Namely, the goal is to improve flux-profile parameterization over mountain ridge tops for hub-height winds. Observed values of ϕ_m and ϕ_h will be compared to those predicted by the BD and ZA relationships. Canopy effects (Eqns. 4.28-4.30) will be evaluated, and a new flux-profile relationship will be formulated for mountain ridge tops. The new relationship is evaluated at an independent test location. Section 4.2 provides an overview of the one-year-long field campaign used to obtain the dataset, and describes the methodology. Section 4.3 provides a discussion of results and limitations, followed by conclusions and suggestions for future work in section 4.4.

Table 4.1: Meteorological sensor description, accuracy, and time averaging for field campaign.

	Sensor description	Accuracy	Sampling Frequency	Averaging	Height AGL (m)
All sites	Onset HOBO Temperature and RH sensor (U23-001)	T: +/- $0.21°C$ RH: +/- 3.5%	15 min	None	1, 1.5, 2, 3, 4
Site 1	Thies Clima First Class Wind Velocity Sensor (cup anemometer) (4.3350.00.000)	+/- 0.2 m s^{-1}	1 Hz	15 min	38, 59, 78
	Campbell Scientific Temperature sensor (107)	+/- $0.4°C$	1 Hz	15 min	75
	R.M. Young Barometric Pressure (61205V)	+/- 0.5 hPa	1 Hz	15 min	75
Site 2	Vaisala (cup) Anemometer (WAA252)	+/- 0.17 m s^{-1}	1 Hz	10 min	72, 95
	Vaisala Temperature Probe (HMP155)	+/- $0.3°C$	1 Hz	10 min	91
	Vaisala Barometer (PTB210)	+/- 0.25 hPa	1 Hz	10 min	2

4.2 Field-Work Description and Data

4.2.1 Overview of Field Campaign

A one-year-long field campaign from 1 June 2014 through 31 May 2015 was done to measure temperature and wind profiles at two anonymous wind-farm locations in British Columbia (Fig. 4.1). Observations from these wind farms are confidential and the property of the wind-farm owners. Hence, I do not reveal exact locations and names and hereafter refer to these wind farms as sites 1 and 2. Sites 1 and 2 in this chapter do not correspond to sites 1 and 2 in Chapters 2 and 3.

Sites 1 and 2 are located at elevations between 500 and 1000 m ASL along mountain ridge tops or plateaus (Appendix C). The straight-line horizontal distance between sites 1 and 2 is approximately 60 km. Both wind farms have existing meteorological towers with anemometers on at least two levels, temperature near wind-turbine hub height, and a barometer. At each site I deployed five temperature/RH data loggers (Onset HOBO model U23-001) aligned in the vertical near the bottom of the meteorological tower to complement the existing sensors (Fig. 4.1). Such HOBO sensors have been used in a few previous field studies with much success (Whiteman et al., 2000; Lundquist and Cayan, 2007; Whiteman and Hoch, 2014). Table 4.1 provides the full sensor details at sites 1 and 2, including their elevations, accuracy, sampling frequency, and time-averaging scheme. I use these sensors to observe vertical profiles of wind and temperature, and also static stability.

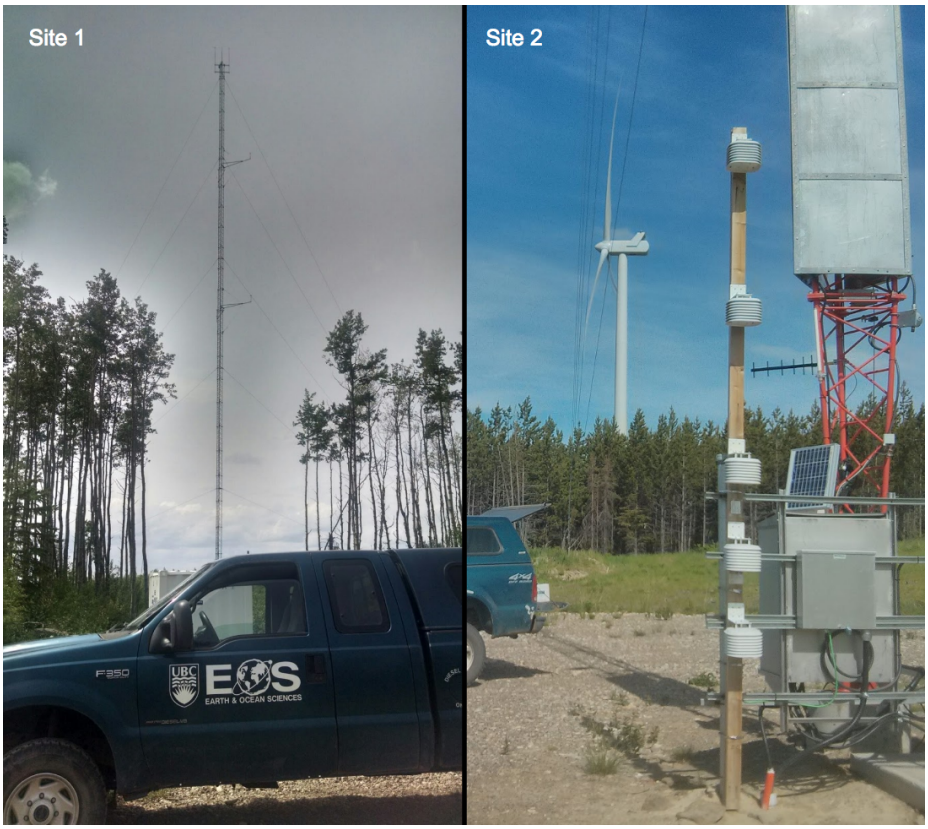

Figure 4.1: Photographs of site 1 (left) and site 2 (right). The photo of site 2 shows the five Onset HOBO temperature/RH sensors in their white radiation shields aligned in the vertical at 1-, 1.5-, 2-, 3-, and 4-m AGL. Trees at both locations are predominantly of pine-forest type.

Temperature measurements were converted to potential temperature by using a reference height of mean sea level. Potential temperatures were calibrated by removing the mean bias found during statically-neutral conditions relative to the sensor at 1 m. This method is used in lieu of laboratory measurements, which were not available. Measurements of RH allowed for the calculation of virtual potential temperature to include the thermodynamic effects of water vapor on air density. Since calibration results of the wind measurements were not available, the effects of anemometer overspeeding and flow modification due to the meteorological tower have not been explicitly accounted for. Overspeeding is only a small effect in modern cup anemometers (Pindado et al., 2014). The artifacts of flow modification may have been reduced by averaging wind speeds from the two sensors located on opposite booms at the same elevation.

Sites 1 and 2 are located in coniferous-forested regions in a continental-polar climate type.

Clear-cuts are present in locations immediately around wind-turbines, and immediately adjacent to roads. Subjective estimates of the total amount of area at each location covered by forest, based on satellite imagery, is 70 to 90%. Fig. 4.1 provides an estimate of canopy density. At site 2, dead tree stands are also present from a recent forest fire. Measurements taken at this location are less than a 1-km distance from the dead tree stands. In the immediate vicinity of the measurements, a clear-cut area around the meteorological tower ranged from approximately 10 (site 1) to 50 (site 2) meters, with patchy grass and bare soil present (Fig. 4.1). Additionally, tree-top elevations (canopy heights) are estimated to be between 10 and 15 m. The Hobo sensors were also located in the vicinity of metal boxes, or the meteorological tower itself. A check of the data did not reveal any interference of these objects at sites 1 or 2.

4.2.2 Methodology for Deriving Observed ϕ_m, ϕ_h, and New Flux-Profile Relationships

Time-averaged measurements (Table 4.1) valid on the hour were used to calculate the observed values of ϕ_m, ϕ_h, and ζ at site 1 for comparison to the BD and ZA relationships (Eqns. 4.8-4.23). I included the ZA relationships because of their common use within WRF [see Jiménez et al. (2012)], which has been used in two related components of this research (Siuta et al., 2017a,b). The observations were also used to evaluate canopy effects (Eqns. 4.28-4.30), and to derive a new flux-profile relationship for mountain ridge tops. Site 2 is used as an independent dataset to evaluate the new flux-profile relationship for wind. While sites 1 and 2 are located approximately 60-km apart, they are subject to the same synoptic conditions. Therefore, there is some correlation between the observations at each location. The linear correlation coefficient for the wind speeds near hub-height between sites 1 and 2 was 0.63 over the entire year.

I start by calculating the values of u_* and θ_* with Eqns. (4.3) and (4.4) at a height of $z = 48.5$ m. This corresponds to the midpoint between the lowest two anemometers at site 1, and is the middle of the layer that values of ϕ_m and ϕ_h will be calculated. The average of the observed winds at 38 and 59 m is used in Eqn. (4.3) to estimate u_*. A log-linear interpolation of the θ values observed at 1, 1.5, 2, 3, 4, and 75 m plotted against the logarithm of height is used to obtain the value of θ at $z = 48.5$ m. The value of θ at 48.5 m is used in Eqn. (4.4) to obtain θ_*. I use the potential temperature at 1 m as a substitute for potential temperature at the physical ground surface (θ_g), which was not available.

Because I do not have direct flux measurements, estimates of the bulk drag coefficients C_D and C_H are needed to close Eqns. (4.3) and (4.4). The values of C_D and C_H are often assumed constant for simplicity, and due to the lack of sufficient observational data (Arya, 2001). I assume values for

these coefficients based on Stull (1988),

$$C_{D,H} = \frac{k^2}{ln(\frac{z-d}{z_o})^2}. \tag{4.31}$$

Estimates of the canopy height at sites 1 and 2 (Fig 4.1) were used to derive values for the zero-plane displacement distance. I estimate $h = 15$ m ($h = 10$ m) and $d = 11.25m$ ($d = 7.5$ m) at site 1 (site 2), with both sites covered predominantly in a pine-forest vegetation type. These values for h and d are consistent with those provided in (Graefe, 2004), who summarize values of h and d for several studies. Using these approximations for h, the transition layer depth (z_*) extends approximately 45 m above the canopy top at site 1 and 30 m above the canopy top at site 2.

Measurements *above* the transition layer are required to determine z_o (Kaimal and Finnigan, 1994), because z_o is an aerodynamic surface-layer parameter. I cannot use observed wind profiles to estimate z_o because there are insufficient observation heights above the transition layer. Instead, I use a value of 0.93 m for z_o provided in Graefe (2004) for a coniferous forested region. Graefe (2004) list values for z_o between 0.4 and 0.93 m for coniferous-forested regions with similar canopy heights to sites 1 and 2. I use the upper end of the range because of the increased complexity of the terrain in my locations. An analysis of the sensitivity of my results on the assumptions in values of h, d, and z_o is provided in section 4.3.5.

Using these assumptions, the values for u_* and θ_* were used to calculate the observed values of ϕ_m and ϕ_h with Eqns. (4.24) and (4.25).

Values of ϕ_m and ϕ_h were used to calculate ζ as (Arya, 2001):

$$\zeta = \frac{Ri_b}{(\frac{\phi_h}{\phi_m^2})}, \tag{4.32}$$

where Ri_b is

$$Ri_b = \frac{g}{\theta} z \frac{\theta_v - \theta_{vg}}{u^2}. \tag{4.33}$$

Following the above process allows me to compare observed flux-profiles to those of the BD and ZA relationships, and to fit a new stability-dependent relationship to the data. When comparing to the BD or ZA relationships, theoretical values of ϕ_m and ϕ_h are calculated by using the observed value for ζ in Eqns (4.8)-(4.23).

I eliminate low wind speeds (less than 3 m s^{-1}) from the data since light winds often do not follow MOST, nor are low wind speeds of concern to the wind-energy community (Optis et al., 2016). Fig. 4.2 shows a histogram of observed wind directions at site 1. I discretize this histogram into three wind regimes to examine the role of wind direction on observed values of ϕ_m. The three

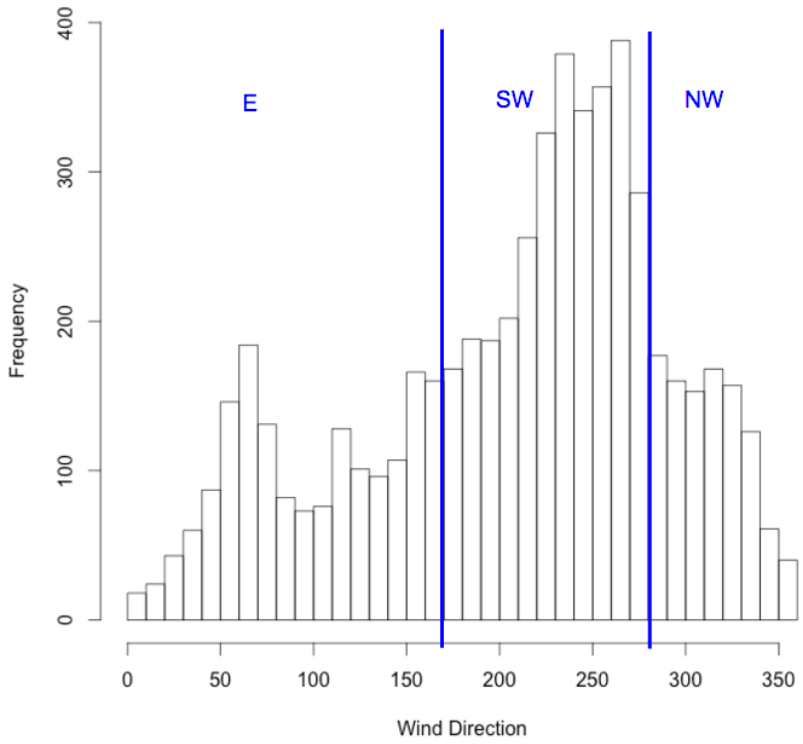

Figure 4.2: Histogram of wind directions at site 1. Three wind regimes labeled: southwesterly (SW), northwesterly (NW), and easterly (E).

regimes are southwest (SW, 180° to 270°), northwest (NW, 270° to 360°), and east (E, 0° to 180°).

Winds with a dominant westerly or southwesterly component occur when upper-level troughing is present over the eastern Pacific, bringing southwesterly flow across British Columbia. This setup is typical of the Pacific Northwest and is the dominant wind direction at site 1. A secondary peak is evident through winds having an easterly component. Winds with an easterly component are observed during periods of lee cyclogenesis in the Canadian Rocky Mountains, or during arctic outflow events. The northerly or northwesterly flow regime is observed when ridging is present over Alaska, Yukon, and British Columbia.

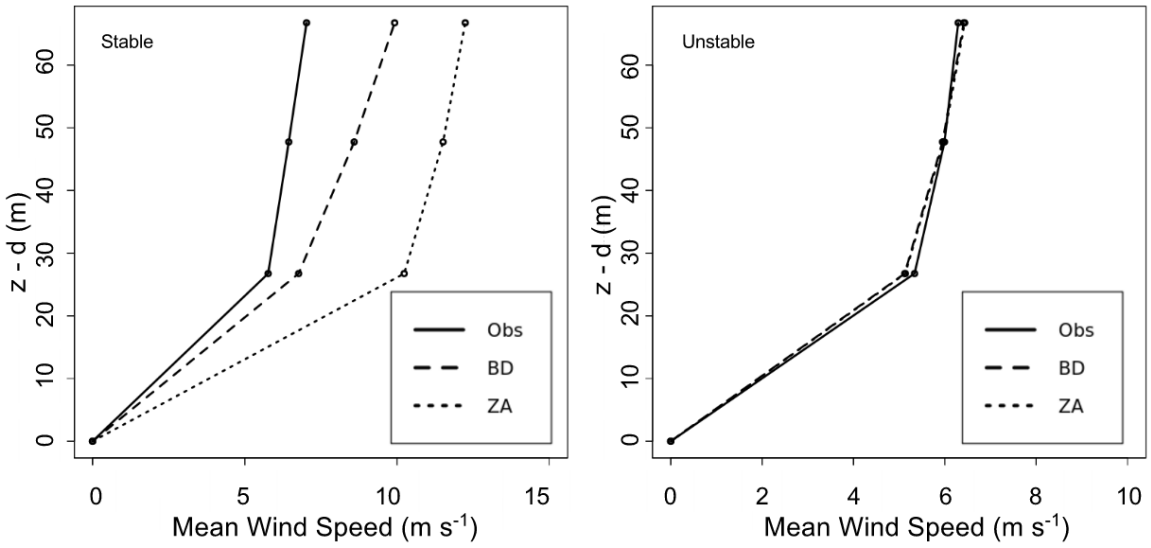

Figure 4.3: Comparison of the observed mean-wind profiles at site 1 (solid) to the profiles produced from the BD (dashed) and ZA (dotted) relationships. The left panel is for statically-stable ASL conditions ($\zeta > 0$) while the right is for statically-unstable conditions ($\zeta < 0$).

4.3 Results and Discussion

4.3.1 Comparison of Observed Wind Profiles with those of the BD and ZA Relationships

I calculate the average observed wind profiles at site 1 by static stability regime to compare with those produced using the BD and ZA relationships. Statically-stable conditions occur when $\zeta > 0$ and statically-unstable conditions when $\zeta < 0$. For the BD and ZA relationships, wind profiles are first calculated based on the observed values of u_* and ζ prior to averaging. A value of 11.25 m was used for d at site 1 when producing the profiles.

Fig 4.3 compares the observed, BD, and ZA wind profiles at site 1 during statically-stable and statically-unstable conditions. For statically-unstable conditions, the BD and ZA relationships are identical (i.e., Eqns. (4.17) and (4.23) are the same). Thus, their profiles (Fig. 4.3) are the same. The average BD and ZA profiles for statically-unstable conditions are a good fit to the average observed wind profile.

For statically-stable conditions, however, the BD and ZA profiles provide a poor fit to the observations (Fig. 4.3). Both profiles substantially overestimate the winds at site 1. For winds closest to hub height (e.g., the 78-m AGL observations), the BD and ZA profiles are too fast by factors of 1.4

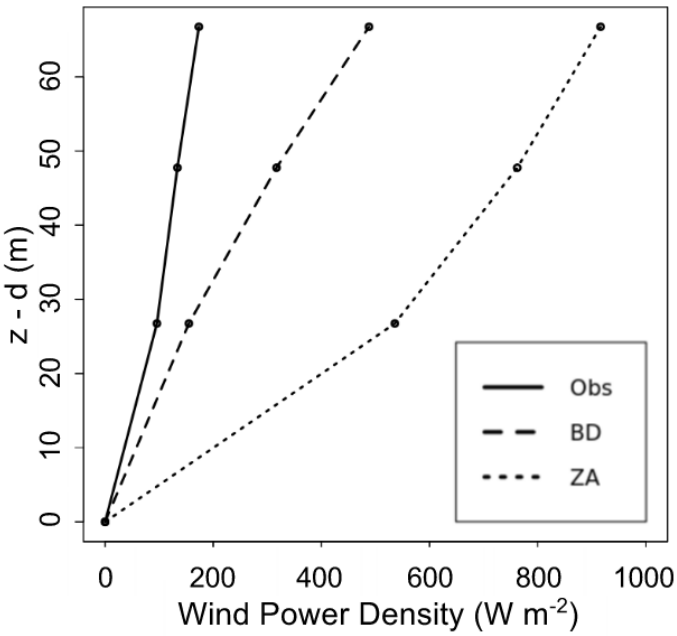

Figure 4.4: Wind-power density (WPD) at site 1 under statically-stable ($\zeta > 0$) conditions for the observed (solid), BD (dashed), and ZA (dotted) mean-wind profiles.

and 1.7, respectively. Wind power is related to the cube of wind speed for the speeds shown in the average profiles in Fig. 4.3 for statically-stable regimes. Therefore, the BD and ZA profiles provide overestimates of wind power by factors of approximately 2.7 and 4.9, respectively, for statically-stable conditions. To highlight this point, I plot the wind-power density (WPD) at site 1 in Fig. 4.4 for the observations and the BD and ZA relationships for statically-stable conditions. WPD is calculated as

$$WPD = \frac{1}{2}\rho u^3. \tag{4.34}$$

The remainder of my discussion will be focused on statically-stable conditions because of the good fit by the BD and ZA relationships to the observations in statically-unstable conditions. I also focus the comparison to the BD profiles because of a better fit to the observed wind profile under statically-stable regimes (Fig. 4.3).

4.3.2 Observations of ϕ_m, Theoretical ϕ_m, and Canopy Effects

Observed and theoretical (BD) values of ϕ_m at site 1 are plotted in Fig. 4.5. The observed ϕ_m values (black dots, Fig. 4.5) are consistently less than the BD relationship (red curve, Fig. 4.5)

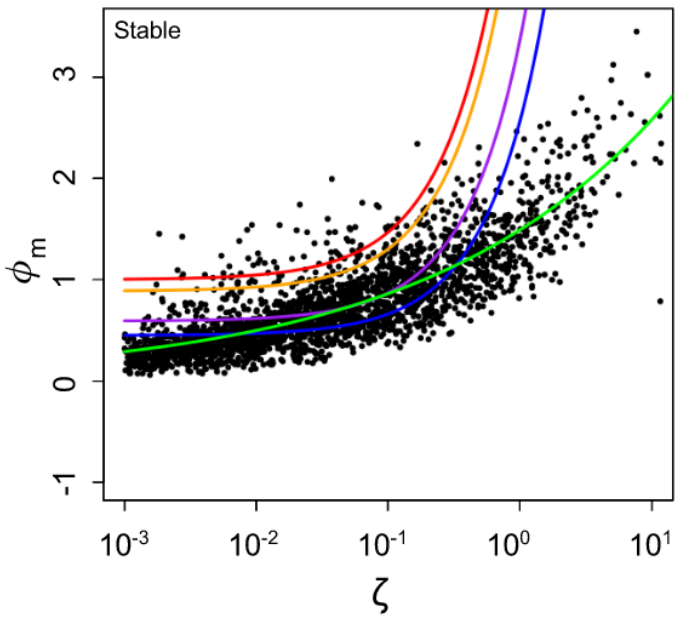

Figure 4.5: Plot of ϕ_m at site 1 as a function of ζ. Observed values are represented by black dots. The red curve is the standard BD relationship. The orange and purple curves are canopy-corrected BD relationships with a $z_* = 45$ and $z_* = 150$ m, respectively. The blue curve is a constant reduction factor of 55% appled to the BD curve. The green curve is the least-squares best-fit to the observed dataset.

allows, with only a few observations larger than theoretical. Fig. 4.6 shows the ratio of ϕ_m to the theoretical values produced using the BD relationship $\phi_{m,BD}$. On average, ϕ_m is 55% smaller than $\phi_{m,BD}$ for statically-stable conditions. This is evidence that surface stress (u_*) is likely being underestimated under statically-stable conditions, providing wind profiles that increase speed too quickly with height (e.g., allow winds to be to fast near hub heights).

For numerical modeling, the implications of the results in Figs. 4.5 and 4.6 are that parameterized values of u_* using the BD relationships will be underestimated for statically-stable conditions along mountain ridges. Because ϕ_m values provided by BD are larger than observed, values for u_* are smaller than observed when Eqn. (4.8) is rearranged to solve for u_*. Since values of u_* are used to update momentum tendencies for subsequent model time-steps, forecast winds under statically-stable conditions are likely to be overestimated because of the underestimated surface stress.

I theorize the reduced magnitude of ϕ_m could be the result of canopy effects since the measurements are likely to be within the transition layer at site 1. I use Eqns. (4.28)-(4.30) with $z_* = 45$ m to include canopy effects on the theoretical BD relationships (orange curve, Fig. 4.5). The result is a 12% reduction in the values predicted by the BD relationships (red curve, Fig. 4.3). The bulk of

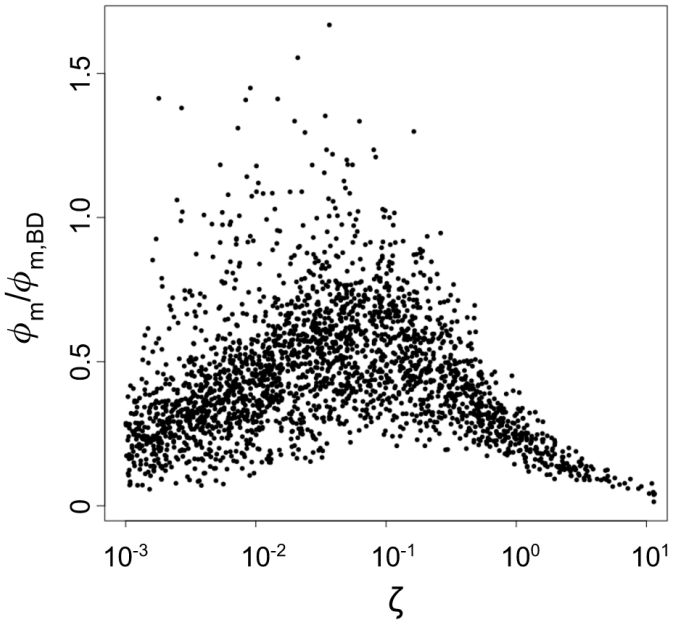

Figure 4.6: Ratio of observed non-dimensional wind shear ϕ_m to the theoretical values diagnosed from the BD relationships $\phi_{m,BD}$.

the observed ϕ_m are still smaller than theoretical after including canopy effects. This finding leads me to hypothesize that a) the transition-layer depth may be larger over forested mountain crests, or b) the assumption of a stability-independent correction factor does not apply over mountain ridges.

To evaluate a), I increased the transition-layer-depth magnitude until the resultant curve (purple, Fig. 4.5) intersected the least-squares best-fit to the observations (green, Fig. 4.5). The curves intersected when z_* was 150 m. This is preliminary evidence that the transition layer may extend to higher altitudes over mountain ridges (by a factor of 3). However, this assumes that the application of a stability-independent reduction factor is correct. The blue curve in Fig. 4.5 represents a constant reduction of the BD relationships by 55%, found by taking the average ratio in Fig. 4.6.

It is clear that a stability-independent correction factor is not sufficient at my location on a mountain ridge [e.g, b) and Eqns. (4.28)-(4.30) do not apply]. The increase of ϕ_m with increasing static stability is much less for the observations (green, Fig. 4.5) than the theoretical BD curves (red, Fig. 4.5), including when accounting for canopy effects (orange, purple, Fig. 4.5). It is possible that increased mountain-wave activity and/or advection of internal boundary layers from upstream roughness features may be contributing to enhanced turbulent mixing in statically stable ASLs, thus reducing observed values of ϕ_m at site 1. However, without vertical soundings, or the availability of lidar and/or sodar equipment, I cannot confirm the roles of mountain waves or internal boundary

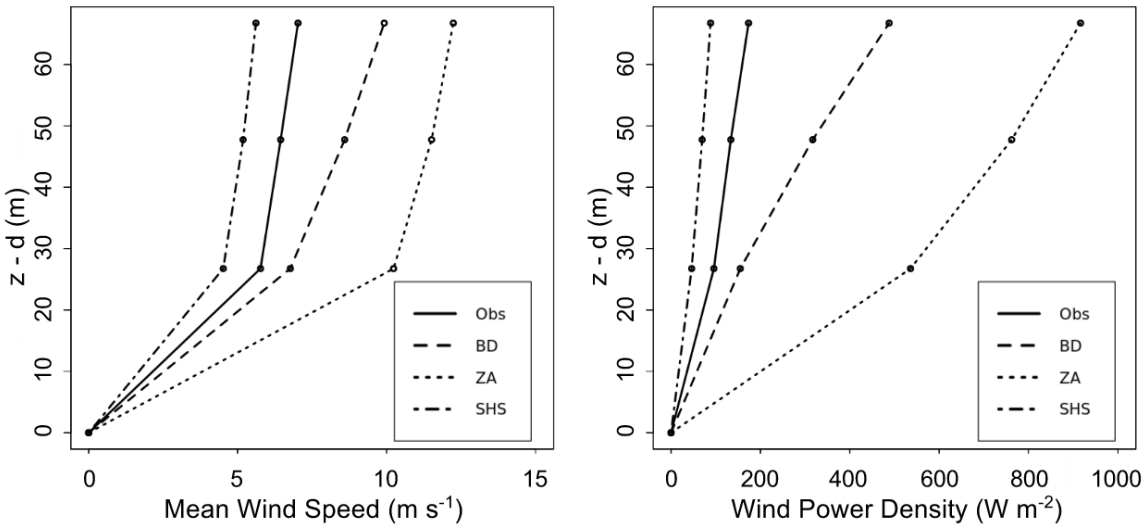

Figure 4.7: Comparison of diagnosed wind profiles using the SHS (dot-dashed), BD (dashed), and ZA (dotted) relations to that of the observations (solid) at site 1.

layers. Such a study is recommended for future work based on my findings. Appendix C provides an overview of the terrain in the vicinity of the wind farms used in this chapter.

I propose that the best-fit curve (green, Fig. 4.5) is a new similarity relationship applicable to mountain ridge tops for statically-stable conditions. The curve is described by:

$$\phi_m = 0.001 + 1.49\zeta^{0.238}, \tag{4.35}$$

with integrated form

$$\psi_m = 0.999ln(\frac{z}{z_o}) + 6.26\zeta^{0.238}. \tag{4.36}$$

Hereafter, I refer to this new flux-profile relationship as Siuta-Howard-Stull (SHS). I plot the fit of the mean wind and wind-power density profiles produced with the SHS relationship in Fig. 4.7. The SHS profile is a much better fit to the observed wind-speed and wind-power density profiles at site 1. However, the SHS fit does have a tendency to underestimate the wind speeds and wind-power densities near hub heights. Nonetheless, it is a much better fit than the BD and ZA relationships. I apply the SHS profile at site 2 for independent verification in section 4.3.4.

4.3.3 Sensitivity of ϕ_m to Season and Wind Direction

Scatter around the SHS curve in Fig. 4.5 is observed. To try to explain the scatter, I divide the observed ϕ_m values into seasonal and wind-flow regime datasets. I use June, July, August as summer,

Table 4.2: Flow-regime-dependent summary statistics of fit to SHS relationship for statically-stable conditions at site 1. Residual calculated as $\phi_{m,obs} - \phi_{m,SHS}$, where $\phi_{m,SHS}$ is from the SHS relationship.

Wind Regime	Mean Residual	Residual Variance
SW	-0.03	0.04
NW	0.04	0.09
E	0.04	0.11

September, October, November as fall, December, January, February as winter, and March, April, May as spring for the seasonal analysis shown in Fig. 4.8. The wind-direction regimes in Fig. 4.2 are used to isolate ϕ_m by flow regime in Fig. 4.9.

The seasonal plots (Fig. 4.8) do not show any substantial change in scatter as a function of season. However, when analyzing the dataset by wind direction (Fig. 4.9), more scatter is observed in the E and NW regimes than with the SW regime. Table 4.2 shows residual error statistics to the fit of the observed ϕ_m values by the SHS curve. Residuals are small for all three regimes. However, residual variance is larger for the NW and E regimes. Using a Kolmogorov-Smirnov (K-S) statistical significance test, the differences in the residual distributions for NW and E flow regimes passes significance testing at the 5% level when compared to the SW regime. Thus, the increased variance in the data for the NW and E regimes is statistically significant. It is possible that flow from the NW or E regimes is interacting with roughness elements of different characteristics and may be causing the scatter. Looking at Figs. 4.8 and 4.9, it appears that the proposed new similarity curve is not a function of wind direction or season for this site.

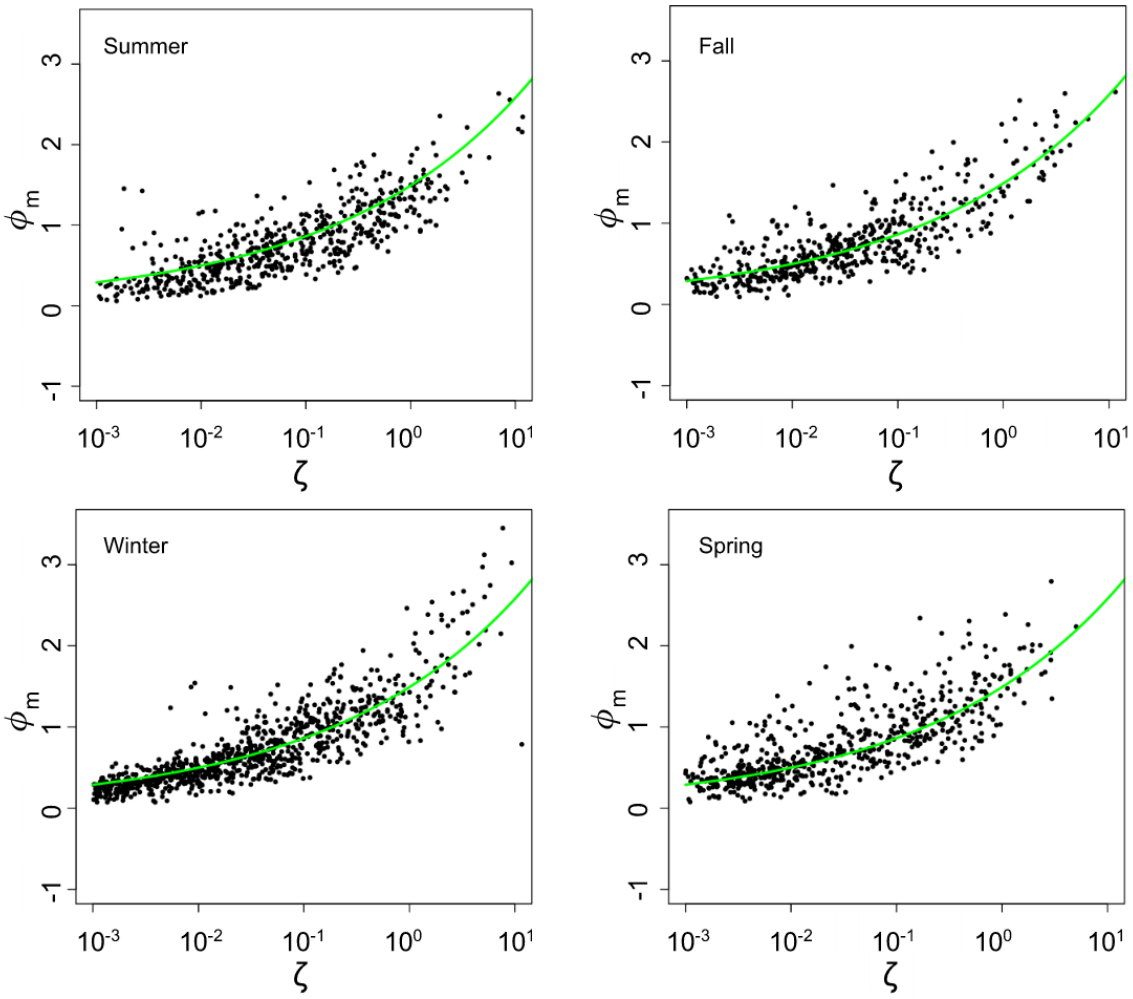

Figure 4.8: ϕ_m observed at site 1 during statically-stable conditions (black dots) for the summer (top left), fall (top right), winter (bottom left), and spring (bottom right). The green curve is the SHS theoretical fit derived using the data inclusive of all seasons.

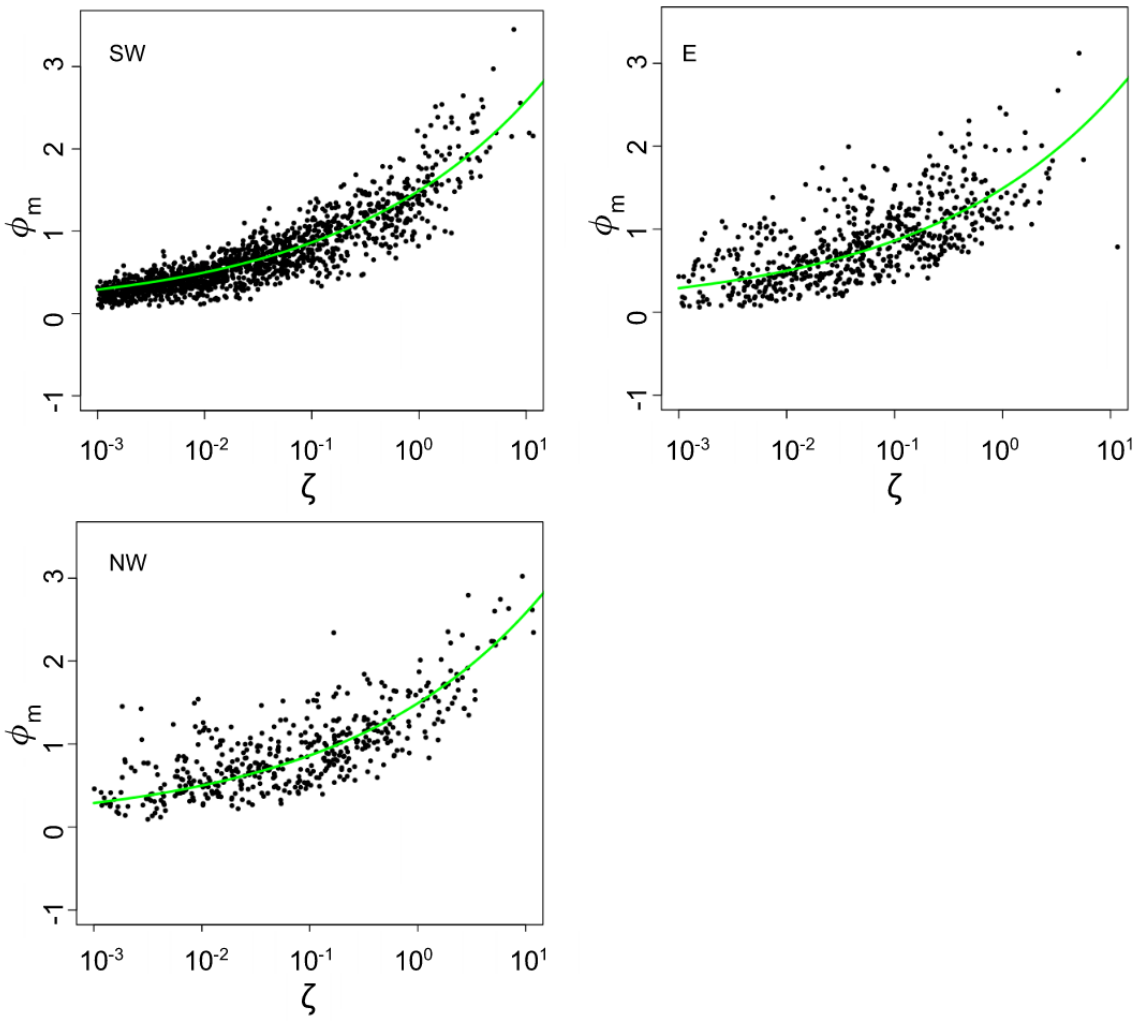

Figure 4.9: ϕ_m observed at site 1 during statically-stable conditions (black dots) for the southwest (top left), east (top right), and northwest (bottom) wind regimes. The green curve is the SHS theoretical fit derived using the data inclusive of all wind directions.

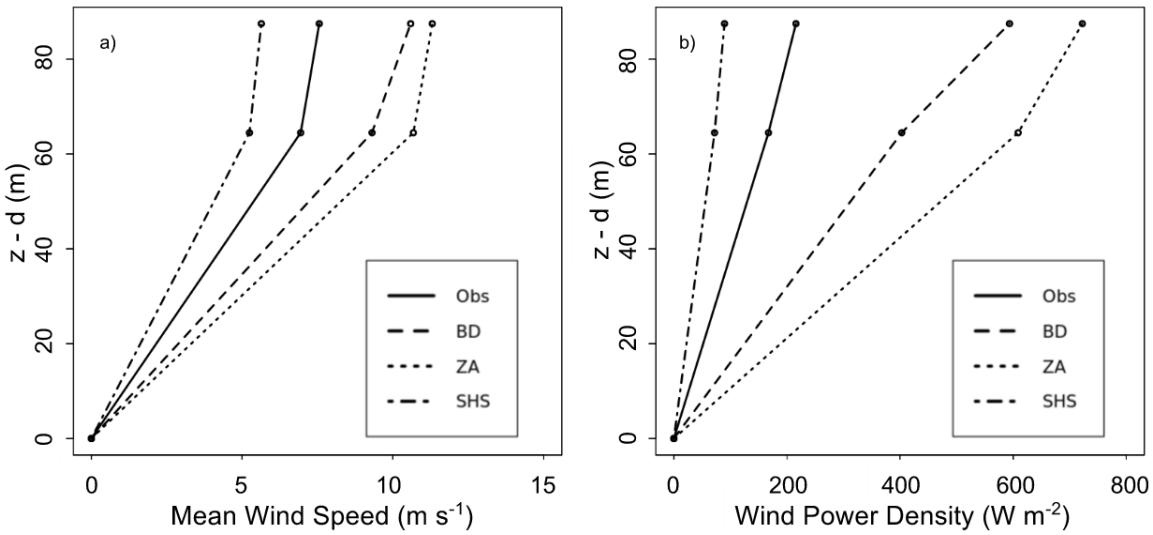

Figure 4.10: Comparison of diagnosed wind profiles using the SHS (dot-dashed), BD (dashed), and ZA (dotted) relations to that of the observations (solid) at site 2.

4.3.4 Independent Verification of SHS Profile at Site 2

I compare the wind profiles produced by the SHS, BD, and ZA flux-profile relationships during statically-stable conditions to the independent observations at site 2 in Fig. 4.10. Observed values of u_*, θ_*, and ζ are used as input in each respective relationship to produce the profiles. A summary of the mean-profile fits is given in Table 4.3 where mean absolute error skill scores (MAESSs) are provided to show relative improvement in the representation of wind speeds by using the SHS relationship over the BD and ZA relationships.

The BD and ZA relationships produce winds speeds that are too fast while the SHS relationship is too slow. Profiles produced using the SHS relationship, however, are large improvements over those produced using the BD or ZA relationships. The SHS profile reduces MAE over that of the BD relationships by 28 and 36% at 72 and 95 m, respectively. When compared to the ZA relationships, MAE is reduced by 54 and 48%, respectively.

The improvement is even more evident once converting to WPD, where the BD and ZA relationships largely overestimate the wind resource potential (Fig. 4.10). As Siuta (2013) discusses, WPD values of 500 W m^{-2} are prime wind resources. My results show that use of the BD and ZA relationships would significantly overestimate the wind resource potential under statically-stable conditions. The SHS profile on the other hand results in an underestimation. The underestimation, however, is of much less magnitude then the overestimation produced by the BD and ZA profiles.

Table 4.3: Verification summary statistics for BD, ZA, and SHS profiles at site 2 for statically-stable conditions. $\bar{U}_{SHS}$ is the mean wind speed using the SHS flux-profile. $\bar{U}_{BD}$ is the mean wind speed using the BD flux-profile. $\bar{U}_{ZA}$ is the mean wind speed using the ZA flux-profile. $\bar{O}$ is the mean observed wind speed. $MAESS_{BD}$ is the mean absolute error skill score of the SHS wind profile relative to that of the BD wind profile. $MAESS_{ZA}$ is the mean absolute error skill score of the SHS wind profile relative to that of the ZA wind profile. Larger skill scores are better.

Height	$\bar{U}_{SHS}$	$\bar{U}_{BD}$	$\bar{U}_{ZA}$	$\bar{O}$	$MAESS_{BD}$	$MAESS_{ZA}$
m	m s^{-1}	m s^{-1}	m s^{-1}	m s^{-1}	%	%
72	5.24	9.30	10.68	6.94	28	54
95	5.63	10.58	11.30	7.55	36	48

4.3.5 Sensitivity of Observed ϕ_m to Zero-Plane Displacement and Roughness Length

To solve for the observed values of ϕ_m, I had to approximate the values of d and z_o based on estimates of the canopy height, or prior results over similar vegetation in the literature (see section 4.1.2). To quantify the effects of these assumptions, I recalculated the values for ϕ_m at site 1 through several additional tests summarized by Fig. 4.11.

First, I varied the choice of d. I found that using the zero-plane displacement distance within the calculation had a large effect. Without its use (e.g., $d = 0$ m), ϕ_m values were 29% larger (Fig. 4.11a). Much less uncertainty arises through the exact value of d (Fig. 4.11b). Values for ϕ_m were effected by 7% when using $d = 9$ m instead of $d = 11.25$ m. This represents the difference between a canopy height estimate of 12 vs. 15 m. Since $d \approx 0.75h$ (Garratt, 1992), I conclude that uncertainty in the estimate of h results in only a small uncertainty in ϕ_m. Namely, the effects of differences in the canopy height over a forested region are not as large as neglecting to use a zero-plane displacement.

Second, I varied z_o based on the values quoted in Graefe (2004). ϕ_m was recalculated using $z_o = 0.4$ to compare with the results shown throughout this chapter using $z_o = 0.93$ (Fig. 4.11c). Using the larger roughness value over the smaller roughness resulted in a 19% difference in ϕ_m. Thus, I conclude that my assumption of the roughness-length value is the largest source of uncertainty in my study. The 19% difference, however, is marginal compared to the much larger reduction found in Fig. 4.6 over that of the BD relationships. I therefore conclude that my assumptions have little effect on the results that show that the BD and ZA relationships have generally poorer fits than the SHS profile for statically-stable conditions.

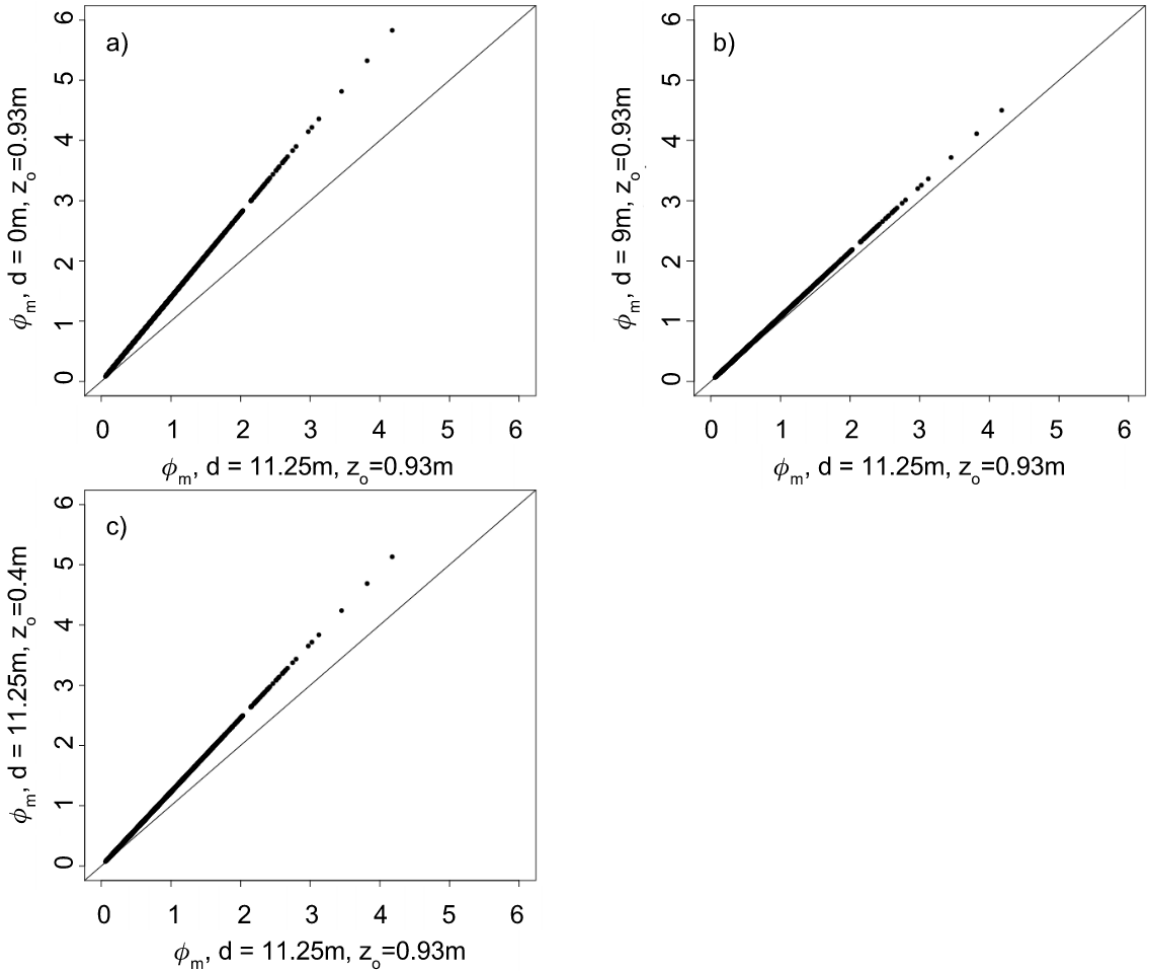

Figure 4.11: Sensitivity of the calculated ϕ_m values to the use of $d = 0$ m vs. $d = 11.25$ m (a), $d = 9$ m vs. $d = 11.25$ m (b), and $z_o = 0.93$ m vs. $z_o = 0.4$ m (c).

4.3.6 Observations of ϕ_h

While this study has been focused on improvements in the relationship for ϕ_m in statically-stable conditions because of the direct applications to wind energy, I felt it important to show the results for ϕ_h at site 1 for the same static-stability conditions (Fig. 4.12). I find observed values of ϕ_h are relatively constant with static-stability magnitude. There is also less scatter in the observations of ϕ_h (Fig. 4.12) than in ϕ_m (Fig. 4.5). Fig. 4.12 confirms a finding by Nadeau et al. (2013) who saw similar behavior in the observed values of ϕ_h over a steep alpine slope. I propose that sinking motions, which develop along mountain ridges in statically-stable conditions as a result of katabatic flows, may cause this behavior. Such sinking motions would limit the strength of static stability in

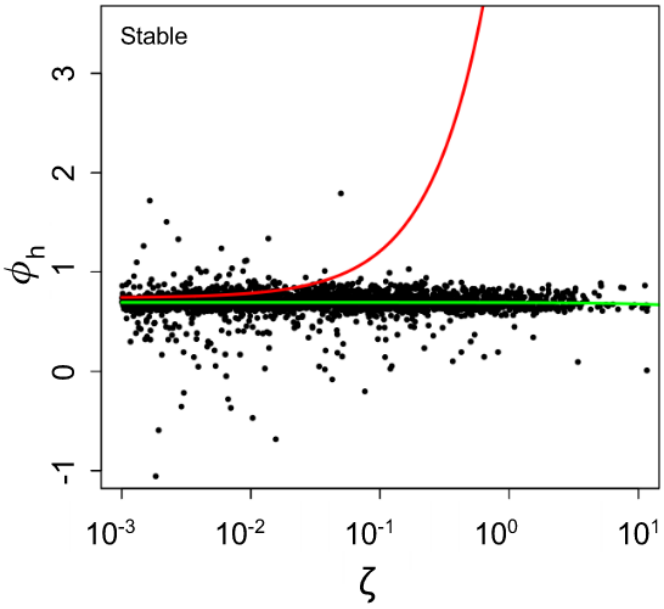

Figure 4.12: Observed ϕ_h (black dots) and theoretical BD flux-profile curve (red) under statically-stable conditions ($\zeta < 0$). The green curve is the best-fit line to the data.

the ASL through adiabatic warming.

4.3.7 Limitations and Avenues for Future Work

There are several limitations to the work presented here, which I must highlight. First, the data does exhibit scatter. One potential cause of this could be the lack of precise flux measurements, or lack of vertical resolution. Direct flux measurements would eliminate the need to estimate u_* and θ_* from mean-wind profiles.

Second, I have made several necessary assumptions that could be affecting the results: constant bulk drag and heat coefficients, non-varying roughness length, and assumptions about the relationship between canopy height and zero-plane displacement distance. The effects of these assumptions were evaluated in the previous section and concluded to be small factors in my results. New studies over mountain ridges should use direct flux measurements to confirm my results. Additionally, since the locations used here were only separated by 60 km, it is possible the relatively high degree of correlation (0.63) of the observations at sites 1 and 2 has resulted in a better fit by the SHS profile at site 2 than may apply at other ridge-top locations. Future work should test stations that are more independent.

More work must be done before the SHS flux-profile can be implemented in a model like WRF.

One of my primary arguments for deriving a new flux-profile relationship was because the BD and ZA relationships were derived from data over homogeneous terrain and thus may not apply over ridge tops. However, applying the SHS relationship across entire WRF domains containing uniform and complex terrain would violate my own motivations for doing this study. Such a method might actually worsen wind forecasts. Future work should involve testing methods of selectively applying the SHS flux-profile relationships in WRF over only regions classified as mountain ridge tops.

4.4 Conclusions

The research provided here showed that widely-accepted forms of flux-profile relationships, such as BD and ZA, fail over the mountain ridge tops for this case study at two wind farms in western North America. Observed values for ϕ_m and ϕ_h depart significantly from their theoretical counterparts. Namely, both were found to be much smaller at my locations. This deviation from theoretical values has important consequences for those wanting to estimate wind resources at similar locations, and also those using NWP models over mountain tops. Since ϕ_m values during statically-stable conditions are much smaller (on average 55%) than those of the BD and ZA relationships, surface-stress magnitudes are underestimated leading to overestimates of wind speeds near wind-turbine hub heights.

I evaluated if canopy effects could be the result of the observed reduction in ϕ_m values at my locations. Inclusion of canopy effects did not reduce theoretical ϕ_m values enough, nor did the canopy reduction factor capture the dependence on static stability. Therefore, a new flux-profile relationship (SHS) was derived from the observations and evaluated at an independent test location. The SHS relationship yielded wind profiles with significant reductions in MAE when compared to that of the BD and ZA wind profiles. The SHS profile did, however, underestimate the wind speeds at the test location. Nonetheless, I find MOST can be applied successfully along mountain ridge tops.

I also confirm the finding from Nadeau et al. (2013) about the lack of slope in ϕ_h during statically stable regimes.

The deviation of ϕ_m and ϕ_h from theoretical may be attributed to the nature of flows in complex terrain. Internal boundary layers advected from rough terrain upstream, mountain-wave breaking, and the development of slope flows could result in intermittent periods of enhanced mixing during statically-stable conditions. Additionally, canopy effects are also likely to be influencing vertical profiles of wind and temperature at my locations.

While I have been successful in deriving a new flux-profile relationship for the two wind farms studied here, future work should be focused in two areas. The first focus should confirm these

results at locations with higher independence; namely, locations that were not subject to the same synoptic conditions as the derivation source. Second, future work should determine how to apply these new flux-profile relationships to numerical weather prediction models. Care should be taken such that the new relationships are not applied across flatter regions of model domains. Care also must be taken if applying the SHS relationship below canopy height since their applicability to this region has not been tested. The relationships found by Businger et al. (1971) and Zhang and Anthes (1982), among others, have been found to strongly correlate to observations over flatter locations. The best approach may be to apply my relationships at ridge-top locations and the existing BD or ZA relationships over other locations. The SHS scheme may also be useful as a diagnostic tool to produce more accurate forecasts of winds at hub-heights using existing NWP-model output. However, such a method would be sensitive to NWP estimates of u_* and θ_*.

Chapter 5

Conclusions

The goal of this dissertation was to improve wind forecasts at wind-turbine hub heights in regions of complex terrain using NWP models. Namely, knowledge gaps have been filled with respect to understanding the sensitivity of deterministic hub-height wind-forecast accuracy to WRF-model configuration, the creation of calibrated probabilistic hub-height wind forecasts, and the limitations of current surface-layer similarity theory over mountain ridge tops.

5.1 Summary of Research Goals and Methods

Energy planners rely on forecasts from NWP models to schedule energy-reserve resources and for market trading. Such forecasts are subject to error due to incorrect initial-condition sources, incomplete knowledge of atmospheric physics, and grid discretization. Improvements in forecast accuracy result in large cost savings as wind penetration climbs to significant levels (Marquis et al., 2011; Wilczak et al., 2015). Research presented in this dissertation focused in three primary areas related to these forecast uncertainties at four wind-farms located on mountain ridge tops in British Columbia:

- In Chapter 2, the sensitivity of WRF-model hub-height wind forecast accuracy to PBL scheme, initial condition, and grid length choice in mountainous terrain was addressed. A year-long case study of 0-24 hour forecasts was produced using two initial-condition sources, three grid lengths, and eight PBL schemes yielding a total of 48 individual forecasts. An ANOVA was performed on the accuracy scores to explain the forecast-accuracy variance amongst the 48 total forecasts. Deterministic verification measures were also used to identify the best performing WRF configurations.

- Since all model forecasts are subject to uncertainty, probabilistic forecasts are used to provide a measure of forecast confidence. The use of probabilistic forecasts for wind energy is relatively new and Chapter 3 evaluated two methods of creating probabilistic hub-height wind forecasts in regions of complex terrain. The first was binning the raw-ensemble distribution. The second was through prescribing a probability distribution based on the error statistics of

the ensemble mean. The effect of ensemble size on short-term probabilistic wind forecasts was also addressed.

- NWP models depend on surface-layer parameterization schemes to represent the effects of the ground surface on the lowest portion of the planetary boundary layer. Such parameterizations rely on the use of MOST derived from data over homogeneous terrain. Chapter 4 evaluated the applicability of traditional MOST relationships over mountain ridge tops through a field campaign and proposed a new relationship.

5.2 Summary of Findings

There were several key findings that will help NWP modelers tasked with producing hub-height wind forecasts in regions of complex terrain:

- Deterministic WRF-forecast accuracy in regions of complex terrain is most influenced by the grid length or PBL scheme, with the dominant factor varying by location, season, and time of day.

- Medium-resolution WRF domains (12-km grid length) produced the highest-accuracy scores on average, although several 4-km grids were best at individual locations after the application of bias correction.

- The ACM2 PBL scheme was generally the most accurate of the eight schemes tested, but the best PBL scheme is unique to the location and/or season. An ensemble of all tested PBL schemes matched the forecast accuracy of the single-best PBL scheme, which may not be known apriori.

- Use of empirical ensemble distributions to represent forecast uncertainty of hub-height winds in complex terrain did not produce calibrated probabilistic forecasts. Instead, prescribing a Gaussian probability model around a bias-corrected ensemble mean resulted in a calibrated probabilistic forecast.

- The benefit of increased ensemble size is through improved forecast sharpness when using a prescribed Gaussian probability model. This is due to the difficulty in knowing the best ensemble members *a priori*. However, probabilistic-forecast sharpness from a small, selective ensemble of only the best PBL scheme was shown to match that of the full, 48-member ensemble. Smaller, selective ensembles have the potential to save substantial computational costs when the best model configurations are known.

- There was no evidence of a spread-skill relationship for short-term hub-height wind forecasts. Nor was there evidence of a relationship between ensemble-mean forecast magnitude and the typical forecast error.

The field campaign highlighted deficiencies in current flux-profile relationships used within NWP models over mountain ridges, which may be used by model developers to improve NWP models:

- Current empirical formulas that diagnose the non-dimensional wind shear and temperature gradient values were demonstrated to poorly fit the observed values for statically-stable surface layers.
- The non-dimensional wind shear was observed to be much less than predicted by existing flux-profile relationships. I theorize this is due to enhanced turbulent mixing over mountain tops in statically-stable regimes due to slope flows, mountain-wave activity, internal boundary layers, and canopy effects. These features may effect wind speeds at other mountain-ridge locations than those shown here because of the common physical mechanisms required for their formation. However, the magnitude of turbulent transport caused by such features may vary by location, thus the reduction in non-dimensional wind shear values over current theory may vary.
- Existing correction factors for canopy effects alone were not found to be sufficient.
- An improved flux-profile relationship for the mountain ridge-top locations studied under statically-stable conditions was given as Eqns. (4.35) or (4.36).

5.3 Potential Applications

Several aspects of this research are currently used in real-time by the Weather Forecast Research Team (WFRT) at the University of British Columbia (UBC). UBC runs an EPS to produce tailored weather forecasts for clients, including BC Hydro, one of the primary funding sources for this research. Based on the results shown here, additional WRF models were added to the UBC EPS to increase ensemble diversity. The best four PBL schemes evaluated in Chapter 2 are now used daily. In addition, since grid length was found to be one of the most important factors influencing forecast accuracy, I further diversified our EPS by adding new grid lengths. Similar tactics could be used by forecast agencies setting up WRF forecast systems for wind energy based on my results.

Prior to this work, the WFRT provided only deterministic hub-height wind forecasts to BC Hydro. The results in Chapter 3 allowed the WFRT to provide calibrated probabilistic forecasts in

real-time. The results of Chapter 3 could be used by energy planners in other regions to also obtain calibrated probabilistic forecasts with computational efficiency.

Evidence of the breakdown of widely-accepted flux-profile relationships for the ridge tops studied should pinpoint areas for future research in complex terrain. My refined flux-profile relationships could be used by model developers to improve forecasts over complex terrain with similar ridge topography and land use. Namely, the new flux-profile relationship should be tested at locations along mountain ridges. Ridge-top locations are likely subject to the same forcing mechanisms causing thermally-induced circulations and dynamic effects which enhance mixing during statically-stable conditions.

5.4 Limitations and Recommendations for Future Work

While the work presented in this dissertation has demonstrated several ways to improve hub-height wind forecasts over complex terrain for the four locations studied, it has also resulted in new questions for future investigation, and is subject to some limitations.

For the results in Chapters 2 and 3, I used forecasts with a maximum time horizon of only 24 hours. These forecasts are useful to same-day reserve scheduling and market trading, but forecasts of longer duration would be beneficial. My primary limitation was availability of spare computer cycles to produce such an extensive dataset. I hypothesize that the sensitivity analysis in Chapter 2 would hold true for longer forecast horizons, but this should be confirmed through additional testing. Fortunately, the advent of cloud-computing has started to remove some of these constraints through availability of vast computing resources to the academic community (Siuta et al., 2016).

The results in Chapter 2 found the initial-condition source to be the smallest factor influencing forecast accuracy. However, a potential caveat to this analysis is that I use initial-condition sources produced only by NCEP (e.g., the GFS and NAM), which use the same underlying data-assimilation system (Shao et al., 2016). It is possible that using initial-condition sources from other national meteorological agencies that use different data-assimilation methods could result in an increase in the influence of initial-condition choice. To this end, I have worked (along with others in the WFRT) to add new initial-condition sources to the UBC EPS by using the Canadian Meteorological Centre's Global Deterministic Prediction System (GDPS or GEM) and the Fleet Numerical Meteorology and Oceanography Center's Navy Global Environmental Model (NAVGEM) to initialize WRF.

In Chapter 3, I use a simple, computationally efficient method to produce calibrated probabilistic hub-height wind forecasts from an EPS. However, I find subsets of the dataset remain uncalibrated. It is possible that other methods not evaluated here (detailed in Chapter 3) may provide better probabilistic calibration than the method shown, particularly for extreme events. Provided the extensive

list of possible methods to create probabilistic forecasts, a comparative study of all these methods over a single dataset in complex terrain would be a welcome contribution to the wind-energy community.

Chapter 4 detailed a field campaign at two wind farms in mountainous areas of British Columbia. Challenges with this field campaign were site access, particularly in the winter, and lack of remote communication abilities. I was therefore limited to deploying self-containing temperature/RH data loggers to complement existing sensors located along meteorological towers. Unfortunately, direct flux measurements were not available, which lessens the accuracy of the observed non-dimensional wind shear and temperature gradient values. Measurements at more levels would have also been ideal, particularly anemometers below the canopy height, and temperature/RH sensors between 4 m and hub height. Future work should directly measure fluxes to confirm the work in Chapter 4, and provide insight into the cause of the reduced non-dimensional wind shear values. Such future work should also use locations that are more independent than the two locations used in Chapter 4. More detailed measurements should be used to confirm if the reduced wind shear under statically-stable conditions is the result of thermal effects, dynamic effects, or a combination of the two. Future work should also measure the changes between vertical wind and temperature profiles in adjacent mountain/valley or mountain/prairie systems. I have also demonstrated a method using simple, inexpensive data loggers that other researchers and energy forecasters might find useful.

Bibliography

Ahlstrom, M., and Coauthors, 2013: Knowledge Is Power: Efficiently Integrating Wind Energy and Wind Forecasts. *IEEE Power and Energy Magazine*, **11 (6)**, 45–52, doi:10.1109/MPE.2013.2277999. → pages 2

Anderson, J. L., 1996: A Method for Producing and Evaluating Probabilistic Forecasts from Ensemble Model Integrations. *Journal of Climate*, **9 (7)**, 1518–1530, doi:10.1175/1520-0442(1996)009⟨1518:AMFPAE⟩2.0.CO;2. → pages 30

Arya, S. P., 2001: *Introduction to Micrometeorology*. 2nd ed., Academic Press, San Diego, 420 pp. → pages 3, 6, 8, 53, 56, 57, 60, 61

Bakhshaii, A., and R. Stull, 2009: Deterministic Ensemble Forecasts Using Gene-Expression Programming. *Weather and Forecasting*, **24 (5)**, 1431–1451, doi:10.1175/2009WAF2222192.1. → pages 5

Beljaars, A. C. M., 1995: The Parametrization of Surface Fluxes in Large-Scale Models Under Free Convection. *Quarterly Journal of the Royal Meteorological Society*, **121 (522)**, 255–270, doi:10.1002/qj.49712152203. → pages 6

Beljaars, A. C. M., and A. A. M. Holtslag, 1991: Flux Parameterization over Land Surfaces for Atmospheric Models. *Journal of Applied Meteorology*, **30 (3)**, 327–341, doi:10.1175/1520-0450(1991)030⟨0327:FPOLSF⟩2.0.CO;2. → pages 53

Berner, J., G. J. Shutts, M. Leutbecher, and T. N. Palmer, 2009: 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 (3)**, 603–626, doi:10.1175/2008JAS2677.1. → pages 30

Bludszuweit, H., J. A. Dominguez-Navarro, and A. Llombart, 2008: Statistical Analysis of Wind Power Forecast Error. *IEEE Transactions on Power Systems*, **23 (3)**, 983–991, doi:10.1109/TPWRS.2008.922526. → pages 30

Bourdin, D. R., T. N. Nipen, and R. B. Stull, 2014: Reliable Probabilistic Forecasts from an Ensemble Reservoir Inflow Forecasting System. *Water Resources Research*, **50 (4)**, 3108–3130, doi:10.1002/2014WR015462. → pages 17, 33

Bretherton, C. S., and S. Park, 2009: A New Moist Turbulence Parameterization in the Community Atmosphere Model. *Journal of Climate*, **22 (12)**, 3422–3448, doi:10.1175/2008JCLI2556.1. → pages 15, 32

Buizza, R., P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Monthly Weather Review*, **133 (5)**, 1076–1097, doi:10.1175/MWR2905.1. → pages 29, 30, 31

Businger, J. A., J. C. Wyngaard, Y. Izumi, and E. F. Bradley, 1971: Flux-Profile Relationships in the Atmospheric Surface Layer. *Journal of the Atmospheric Sciences*, **28 (2)**, 181–189, doi:10.1175/1520-0469(1971)028⟨0181:FPRITA⟩2.0.CO;2. → pages 6, 53, 54, 55, 76

Canadian Wind Energy Association, 2016: Pan-Canadian Wind Integration Study. Tech. rep. → pages 2

Candille, G., 2009: The Multiensemble Approach: The NAEFS Example. *Monthly Weather Review*, **137 (5)**, 1655–1665, doi:10.1175/2008MWR2682.1. → pages 2, 30

Cheng, W. Y., Y. Liu, Y. Liu, Y. Zhang, W. P. Mahoney, and T. T. Warner, 2013: The Impact of Model Physics on Numerical Wind Forecasts. *Renewable Energy*, **55**, 347–356, doi:10.1016/j.renene.2012.12.041. → pages 12

Chow, F. K., B. J. Snyder, and S. F. J. De Wekker, 2013: *Mountain Weather Research and Forecasting: Recent Progress and Current Challenges*. Springer, Dordrecht, London, 750 pp. → pages 4, 10

Clement, M. A., 2012: A Methodology to Assess the Value of Integrated Hydropower and Wind Generation. Ph.D. thesis, University of Colorado, 183 pp. → pages 2

Corbus, D., D. Lew, G. Jordan, W. Winters, F. Van Hull, J. Manobianco, and B. Zavadil, 2009: Up With Wind. *IEEE Power and Energy Magazine*, **7 (6)**, 36–46, doi:10.1109/MPE.2009.934260. → pages 1, 2

Courtney, J. F., P. Lynch, and C. Sweeney, 2013: High Resolution Forecasting for Wind Energy Applications Using Bayesian Model Averaging. *Tellus A*, **65 (0)**, doi:10.3402/tellusa.v65i0.19669. → pages 5, 30

Cutler, N. J., H. R. Outhred, and I. F. MacGill, 2012: Using Nacelle-based Wind Speed Observations to Improve Power Curve Modeling for Wind Power Forecasting. *Wind Energy*, **15 (2)**, 245–258, doi:10.1002/we.465. → pages 33

De Wekker, S. F. J., and M. Kossmann, 2015: Convective Boundary Layer Heights Over Mountainous TerrainA Review of Concepts. *Frontiers in Earth Science*, **3**, 77. → pages 10

De Wekker, S. F. J., D. G. Steyn, and S. Nyeki, 2004: A Comparison Of Aerosol-Layer And Convective Boundary-Layer Structure Over A Mountain Range During Staaarte '97. *Boundary-Layer Meteorology*, **113 (2)**, 249–271, doi:10.1023/B:BOUN.0000039371.41823.37. → pages 10

DeMeo, E. A., G. A. Jordan, C. Kalich, J. King, M. R. Milligan, C. Murley, B. Oakleaf, and M. J. Schuerger, 2007: Accommodating Wind's Natural Behavior. *Power and Energy Magazine, IEEE*, **5 (6)**, 59–67, doi:10.1109/MPE.2007.906562. → pages 1

Deppe, A. J., W. A. Gallus, and E. S. Takle, 2012: A WRF Ensemble for Improved Wind Speed Forecasts at Turbine Height. *Weather and Forecasting*, **28 (1)**, 212–228, doi:10.1175/WAF-D-11-00112.1. → pages 4, 9, 14, 15, 31, 32

Dörenkämper, M., M. Optis, A. Monahan, and G. Steinfeld, 2015: On the Offshore Advection of Boundary-Layer Structures and the Influence on Offshore Wind Conditions. *Boundary-Layer Meteorology*, **155 (3)**, 459–482, doi:10.1007/s10546-015-0008-x. → pages 56, 103

Draxl, C., A. N. Hahmann, A. Peña, and G. Giebel, 2014: Evaluating Winds and Vertical Wind Shear from Weather Research and Forecasting Model Forecasts Using Seven Planetary Boundary Layer Schemes. *Wind Energy*, **17 (1)**, 39–55, doi:10.1002/we.1555. → pages 1, 2, 4, 9

Dyer, A. J., 1974: A Review of Flux-Profile Relationships. *Boundary-Layer Meteorology*, **7 (3)**, 363–372, doi:10.1007/BF00240838. → pages 6, 53

Eckel, F. A., and C. F. Mass, 2005: Aspects of Effective Mesoscale, Short-Range Ensemble Forecasting. *Weather and Forecasting*, **20 (3)**, 328–350, doi:10.1175/WAF843.1. → pages 5, 30, 31, 45

Garratt, J., 1983: Surface Influence Upon Vertical Profiles in the Nocturnal Boundary Layer. *Boundary-Layer Meteorology*, **26 (1)**, 69–80, doi:10.1007/BF00164331. → pages 57

Garratt, J., 1992: *The Atmospheric Boundary Layer*. Cambridge University Press, 316 pp. → pages 56, 57, 72

Giebel, G., R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, 2011: The State-Of-The-Art in Short-Term Prediction of Wind Power - A Literature Overview, 2nd edition. Tech. rep., Riso DTU, 109 pp. → pages 29

Gneiting, T., F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic Forecasts, Calibration and Sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69 (2)**, 243–268, doi:10.1111/j.1467-9868.2007.00587.x. → pages 38, 39

Gneiting, T., A. E. Raftery, A. H. Westveld, and T. Goldman, 2005: Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Monthly Weather Review*, **133 (5)**, 1098–1118, doi:10.1175/MWR2904.1. → pages 4, 5, 31, 38, 39

Graefe, J., 2004: Roughness Layer Corrections with Emphasis on SVAT Model Applications. *Agricultural and Forest Meteorology*, **124 (3)**, 237–251, doi:10.1016/j.agrformet.2004.01.003. → pages 61, 72

Grenier, H., and C. S. Bretherton, 2001: A Moist PBL Parameterization for Large-Scale Models and its Application to Subtropical Cloud-Topped Marine Boundary Layers. *Monthly Weather Review*, **129 (3)**, 357–377, doi:10.1175/1520-0493(2001)129⟨0357:AMPPFL⟩2.0.CO;2. → pages 14, 22, 32

Grimit, E. P., and C. F. Mass, 2002: Initial Results of a Mesoscale Short-Range Ensemble Forecasting System over the Pacific Northwest. *Weather and Forecasting*, **17 (2)**, 192–205, doi:10.1175/1520-0434(2002)017⟨0192:IROAMS⟩2.0.CO;2. → pages 30, 36

Grimit, E. P., and C. F. Mass, 2007: Measuring the Ensemble Spread-Error Relationship with a Probabilistic Approach: Stochastic Ensemble Results. *Monthly Weather Review*, **135 (1)**, 203–221, doi:10.1175/MWR3262.1. → pages 36, 43

Grubišić, V., R. K. Vellore, and A. W. Huggins, 2005: Quantitative Precipitation Forecasting of Wintertime Storms in the Sierra Nevada: Sensitivity to the Microphysical Parameterization and Horizontal Resolution. *Monthly Weather Review*, **133 (10)**, 2834–2859, doi:10.1175/MWR3004.1. → pages 33

Hahmann, A. N., C. L. Vincent, A. Peña, J. Lange, and C. B. Hasager, 2015: Wind Climate Estimation Using WRF Model Output: Method and Model Sensitivities Over the Sea. *International Journal of Climatology*, **35 (12)**, 3422–3439, doi:10.1002/joc.4217. → pages 9

Hersbach, H., 2000: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting*, **15 (5)**, 559–570, doi:10.1175/1520-0434(2000)015⟨0559:DOTCRP⟩2.0.CO;2. → pages 39

Hicks, B. B., 1976: Wind Profile Relationships from the Wangara Experiment. *Quarterly Journal of the Royal Meteorological Society*, **102 (433)**, 535–551, doi:10.1002/qj.49710243304. → pages 53

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky, 1999: Bayesian Model Averaging: A Tutorial. *Statistical Science*, **14 (4)**, 382–417. → pages 30

Högström, U., H. Bergström, A. S. Smedman, S. Halldin, and A. Lindroth, 1989: Turbulent Exchange Above a Pine Forest, I: Fluxes and Gradients. *Boundary-Layer Meteorology*, **49 (1-2)**, 197–217, doi:10.1007/BF00116411. → pages 57

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes. *Monthly Weather Review*, **134 (9)**, 2318–2341, doi:10.1175/MWR3199.1. → pages 14, 32

Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal Boundary Layer Vertical Diffusion in a Medium-Range Forecast Model. *Monthly Weather Review*, **124 (10)**, 2322–2339, doi:10.1175/1520-0493(1996)124⟨2322:NBLVDI⟩2.0.CO;2. → pages 13, 32

Hopson, T. M., 2014: Assessing the Ensemble Spread-Error Relationship. *Monthly Weather Review*, **142 (3)**, 1125–1142, doi:10.1175/MWR-D-12-00111.1. → pages 43

Hu, X.-M., P. M. Klein, and M. Xue, 2013: Evaluation of the Updated YSU Planetary Boundary Layer Scheme within WRF for Wind Resource and Air Quality Assessments. *Journal of Geophysical Research: Atmospheres*, **118 (18)**, 410–490, doi:10.1002/jgrd.50823. → pages 4, 14, 32

Jambunathan, M. V., 1954: Some Properties of Beta and Gamma Distributions. *The Annals of Mathematical Statistics*, **25 (2)**, 401–405, doi:10.1214/aoms/1177728800. → pages 98

Janjić, Z. I., 1994: The Step-Mountain Eta Coordinate Model: Further Developments of the Convection, Viscous Sublayer, and Turbulence Closure Schemes. *Monthly Weather Review*, **122 (5)**, 927–945, doi:10.1175/1520-0493(1994)122⟨0927:TSMECM⟩2.0.CO;2. → pages 14, 32

Jiménez, P. A., and J. Dudhia, 2011: Improving the Representation of Resolved and Unresolved Topographic Effects on Surface Wind in the WRF Model. *Journal of Applied Meteorology and Climatology*, **51 (2)**, 300–316, doi:10.1175/JAMC-D-11-084.1. → pages 4, 9

Jiménez, P. A., J. Dudhia, J. F. González-Rouco, J. Navarro, J. P. Montávez, and E. García-Bustamante, 2012: A Revised Scheme for the WRF Surface Layer Formulation. *Monthly Weather Review*, **140 (3)**, 898–918, doi:10.1175/MWR-D-11-00056.1. → pages 6, 52, 60

Juban, J., N. Siebert, and G. N. Kariniotakis, 2007: Probabilistic Short-Term Wind Power Forecasting for the Optimal Management of Wind Generation. *2007 IEEE Lausanne Power Tech*, IEEE, 683–688, doi:10.1109/PCT.2007.4538398. → pages 31, 38

Junk, C., S. Späth, L. von Bremen, and L. Delle Monache, 2015: Comparison and Combination of Regional and Global Ensemble Prediction Systems for Probabilistic Predictions of Hub-Height Wind Speed. *Weather and Forecasting*, **30 (5)**, 1234–1253, doi:10.1175/WAF-D-15-0021.1. → pages 5, 31

Kaimal, J. C., and J. J. Finnigan, 1994: *Atmospheric Boundary Layer Flows: Their Structure and Measurement*, Vol. 72. 289 pp., doi:10.1016/0021-9169(95)90002-0. → pages 56, 57, 61

Kumaraswamy, P., 1980: A Generalized Probability Density Function for Double-Bounded Random Processes. *Journal of Hydrology*, **46 (1-2)**, 79–88, doi:10.1016/0022-1694(80)90036-0. → pages 98

Lange, M., 2005: On the Uncertainty of Wind Power Predictions—Analysis of the Forecast Accuracy and Statistical Distribution of Errors. *Journal of Solar Energy Engineering*, **127 (2)**, 177–184, doi:10.1115/1.1862266. → pages 29, 30

Lee, Y. H., and L. Mahrt, 2005: Effect of Stability on Mixing in Open Canopies. *Agricultural and Forest Meteorology*, **135 (1-4)**, 169–179, doi:10.1016/j.agrformet.2005.11.013. → pages 57

Lundquist, J. D., and D. R. Cayan, 2007: Surface Temperature Patterns in Complex Terrain: Daily Variations and Long-Term Change in the Central Sierra Nevada, California. *Journal of Geophysical Research*, **112 (D11)**, D11 124, doi:10.1029/2006JD007561. → pages 58

Mahoney, W. P., and Coauthors, 2012: A Wind Power Forecasting System to Optimize Grid Integration. *IEEE Transactions on Sustainable Energy*, **3 (4)**, 670–682, doi:10.1109/TSTE.2012.2201758. → pages 29

Markowski, P., and Y. Richardson, 2010: *Mesoscale meteorology in midlatitudes*. Wiley-Blackwell, 407 pp. → pages 10

Marquis, M., J. Wilczak, M. Ahlstrom, J. Sharp, A. Stern, J. C. Smith, and S. Calvert, 2011: Forecasting the Wind to Reach Significant Penetration Levels of Wind Energy. *Bulletin of the American Meteorological Society*, **92 (9)**, 1159–1171, doi:10.1175/2011BAMS3033.1. → pages 2, 10, 22, 27, 29, 77, 93

Martner, B. E., and J. D. Marwitz, 1982: Wind Characteristics in Southern Wyoming. *Journal of Applied Meteorology*, **21 (12)**, 1815–1827, doi:10.1175/1520-0450(1982)021⟨1815:WCISW⟩2.0.CO;2. → pages 8

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does Increasing Horizontal Resolution Produce More Skillful Forecasts? *Bulletin of the American Meteorological Society*, **83 (3)**, 407–430, doi:10.1175/1520-0477(2002)083⟨0407:DIHRPM⟩2.3.CO;2. → pages 2, 21, 45

McCollor, D., and R. Stull, 2008a: Hydrometeorological Accuracy Enhancement via Postprocessing of Numerical Weather Forecasts in Complex Terrain. *Weather and Forecasting*, **23 (1)**, 131–144, doi:10.1175/2007WAF2006107.1. → pages 33

McCollor, D., and R. Stull, 2008b: Hydrometeorological Short-Range Ensemble Forecasts in Complex Terrain. Part I: Meteorological Evaluation. *Weather and Forecasting*, **23 (4)**, 533–556, doi:10.1175/2008WAF2007063.1. → pages 5, 28, 30, 31

McCollor, D., and R. Stull, 2008c: Hydrometeorological Short-Range Ensemble Forecasts in Complex Terrain. Part II: Economic Evaluation. *Weather and Forecasting*, **23 (4)**, 533, doi:10.1175/2007WAF2007064.1. → pages 38

Monin, A. S., and A. M. Obukhov, 1954: Basic Laws of Turbulent Mixing in the Surface Layer of the Atmosphere. *Contrib. Geophys. Inst. Acad. Sci. USSR*, **24 (151)**, 163–187. → pages 52

Monteiro, C., R. Bessa, V. Miranda, A. Botterud, J. Wang, and G. Conzelmann, 2009: Wind Power Forecasting: State-of-the-Art 2009. *Argonne National Laboratory*, 1–216, doi:10.2172/968212. → pages 2, 4, 29

Murphy, A. H., and R. L. Winkler, 1987: A General Framework for Forecast Verification. *Monthly Weather Review*, **115 (7)**, 1330–1338, doi:10.1175/1520-0493(1987)115⟨1330:AGFFFV⟩2.0.CO;2. → pages 15, 21

Nadeau, D. F., E. R. Pardyjak, C. W. Higgins, and M. B. Parlange, 2013: Similarity Scaling over a Steep Alpine Slope. *Boundary-Layer Meteorology*, **147 (3)**, 401–419, doi:10.1007/s10546-012-9787-5. → pages 56, 73, 75

Nakanishi, M., and H. Niino, 2006: An Improved MellorYamada Level-3 Model: Its Numerical Stability and Application to a Regional Prediction of Advection Fog. *Boundary-Layer Meteorology*, **119 (2)**, 397–407, doi:10.1007/s10546-005-9030-8. → pages 32

Nipen, T., and R. Stull, 2011: Calibrating Probabilistic Forecasts from an NWP Ensemble. *Tellus A*, **63 (5)**, 858–875, doi:10.1111/j.1600-0870.2011.00535.x. → pages 5, 31, 38

Nipen, T. N., 2012: A Component-Based Probabilistic Weather Forecasting System for Operational Usage. Ph.D. thesis, University of British Columbia, 101 pp., doi:10.14288/1.0053570. → pages 32

Optis, M., A. Monahan, and F. C. Bosveld, 2016: Limitations and Breakdown of Monin-Obukhov Similarity Theory for Wind Profile Extrapolation under Stable Stratification. *Wind Energy*, **19 (6)**, 1053–1072, doi:10.1002/we.1883. → pages 6, 53, 55, 61

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic Parametrization and Model Uncertainty. Tech. rep., European Centre for Medium-Range Weather Forecasts, 42 pp. → pages 30

Panofsky, H. A., 1963: Determination of Stress from Wind and Temperature Measurements. *Quarterly Journal of the Royal Meteorological Society*, **89 (379)**, 85–94, doi:10.1002/qj.49708937906. → pages 52

Paulson, C. A., 1970: The Mathematical Representation of Wind Speed and Temperature Profiles in the Unstable Atmospheric Surface Layer. *Journal of Applied Meteorology*, **9 (6)**, 857–861, doi:10.1175/1520-0450(1970)009⟨0857:TMROWS⟩2.0.CO;2. → pages 6

Peterson, E. W., and J. P. Hennessey, 1978: On the Use of Power Laws for Estimates of Wind Power Potential. *Journal of Applied Meteorology*, **17 (3)**, 390–394, doi:10.1175/1520-0450(1978)017⟨0390:OTUOPL⟩2.0.CO;2. → pages 3, 8

Pindado, S., J. Cubas, and F. Sorribes-Palmer, 2014: The Cup Anemometer, a Fundamental Meteorological Instrument for the Wind Energy Industry. Research at the IDR/UPM Institute. *Sensors (Basel, Switzerland)*, **14 (11)**, 21 418–21 452, doi:10.3390/s141121418. → pages 59

Pinson, P., 2012: Very-short-term Probabilistic Forecasting of Wind Power with Generalized Logit-normal Distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **61 (4)**, 555–576, doi:10.1111/j.1467-9876.2011.01026.x. → pages 30

Pinson, P., J. Juban, and G. N. Kariniotakis, 2006: On the Quality and Value of Probabilistic Forecasts of Wind Generation. *2006 International Conference on Probabilistic Methods Applied to Power Systems*, IEEE, 1–7, doi:10.1109/PMAPS.2006.360290. → pages 4, 29, 31, 38

Pleim, J. E., 2007: A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary Layer. Part I: Model Description and Testing. *Journal of Applied Meteorology and Climatology*, **46 (9)**, 1383–1395, doi:10.1175/JAM2539.1. → pages 14, 22, 32

Poulos, G. S., and Coauthors, 2002: CASES-99: A Comprehensive Investigation of the Stable Nocturnal Boundary Layer. *Bulletin of the American Meteorological Society*, **83 (4)**, 555–581, doi:10.1175/1520-0477(2002)083⟨0555:CACIOT⟩2.3.CO;2. → pages 8

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Monthly Weather Review*, **133 (5)**, 1155–1174, doi:10.1175/MWR2906.1. → pages 5, 30

Raupach, M. R., 1979: Anomalies in Flux-Gradient Relationships Over Forest. *Boundary-Layer Meteorology*, **16 (3)**, 467–486, doi:10.1007/BF03163564. → pages 57

Rotach, M. W., and D. Zardi, 2007: On the Boundary-Layer Structure over Highly Complex Terrain: Key Findings from MAP. *Quarterly Journal of the Royal Meteorological Society*, **133 (625)**, 937–948, doi:10.1002/qj.71. → pages 56

Rucker, M., R. M. Banta, and D. G. Steyn, 2008: Along-Valley Structure of Daytime Thermally Driven Flows in the Wipp Valley. *Journal of Applied Meteorology and Climatology*, **47 (3)**, 733–751, doi:10.1175/2007JAMC1319.1. → pages 10

Santoso, E., and R. Stull, 1998: Wind and Temperature Profiles in the Radix Layer: The Bottom Fifth of the Convective Boundary Layer. *Journal of Applied Meteorology*, **37 (6)**, 545–558, doi:10.1175/1520-0450(1998)037⟨0545:WATPIT⟩2.0.CO;2. → pages 8

Santoso, E., and R. Stull, 2001: Similarity Equations for Wind and Temperature Profiles in the Radix Layer, at the Bottom of the Convective Boundary Layer. *Journal of the Atmospheric Sciences*, **58 (11)**, 1446–1464, doi:10.1175/1520-0469(2001)058⟨1446:SEFWAT⟩2.0.CO;2. → pages 6, 8, 53

Shao, H., and Coauthors, 2016: Bridging Research to Operations Transitions: Status and Plans of Community GSI. *Bulletin of the American Meteorological Society*, **97 (8)**, 1427–1440, doi:10.1175/BAMS-D-13-00245.1. → pages 19, 51, 80

Shin, H., and S.-Y. Hong, 2011: Intercomparison of Planetary Boundary-Layer Parametrizations in the WRF Model for a Single Day from CASES-99. *Boundary-Layer Meteorology*, **139 (2)**, 261–281, doi:10.1007/s10546-010-9583-z. → pages 4, 6, 9, 12

Siuta, D., 2013: An Analysis of the Weather Research and Forecasting Model for Wind Energy Applications in Wyoming. M.S. thesis, University of Wyoming. 124 pp. → pages 3, 8, 71

Siuta, D., G. West, H. Modzelewski, R. Schigas, and R. Stull, 2016: Viability of Cloud Computing for Real-Time Numerical Weather Prediction. *Weather and Forecasting*, **31 (6)**, 1985–1996, doi:10.1175/WAF-D-16-0075.1. → pages 80

Siuta, D., G. West, T. N. Nipen, and R. Stull, 2017a: Calibrated Probabilistic Hub-Height Wind Forecasts in Complex Terrain. *Weather and Forecasting*, doi:10.1175/WAF-D-16-0137.1. → pages 7, 16, 60, 93

Siuta, D., G. West, and R. Stull, 2017b: WRF Hub-Height Wind-Forecast Sensitivity to PBL Scheme, Grid Length, and Initial-Condition Choice in Complex Terrain. *Weather and Forecasting*, doi:10.1175/WAF-D-16-0120.1. → pages 7, 32, 33, 60

Skamarock, W., J. Klemp, J. Dudhia, D. Gill, D. Barker, W. Wang, X.-y. Huang, and M. Duda, 2008: A Description of the Advanced Research WRF Version 3. NCAR Tech. Note NCAR/TN-475+STR. 113 pp., doi:10.5065/D68S4MVH. → pages 8, 13, 31, 52

Skamarock, W. C., 2004: Evaluating Mesoscale NWP Models using Kinetic Energy Spectra. *Monthly Weather Review*, **132 (12)**, 3019–3032, doi:10.1175/MWR2830.1. → pages 11

Sloughter, J. M., T. Gneiting, and A. E. Raftery, 2010: Probabilistic Wind Speed Forecasting using Ensembles and Bayesian Model Averaging. *Journal of the American Statistical Association*, **105 (489)**, 25–35, doi:10.1198/jasa.2009.ap08615. → pages 5, 30

Stensrud, D. J., 2007: *Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models*. Cambridge University Press, 478 pp. → pages 32

Stensrud, D. J., J.-W. Bao, and T. T. Warner, 2000: Using Initial Condition and Model Physics Perturbations in Short-Range Ensemble Simulations of Mesoscale Convective Systems. *Monthly Weather Review*, **128 (7)**, 2077–2107, doi:10.1175/1520-0493(2000)128⟨2077:UICAMP⟩2.0.CO;2. → pages 5, 30, 31

Storm, B., J. Dudhia, S. Basu, A. Swift, and I. Giammanco, 2009: Evaluation of the Weather Research and Forecasting Model on Forecasting Low-Level Jets: Implications for Wind Energy. *Wind Energy*, **12 (1)**, 81–90, doi:10.1002/we.288. → pages 9

Stull, R. B., 1988: *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers, 666 pp. → pages 3, 8, 10, 32, 52, 53, 55, 56, 61

Stull, R. B., 1993: Review of Non-Local Mixing in Turbulent Atmospheres: Transilient Turbulence Theory. *Boundary-Layer Meteorology*, **62 (1)**, 21–96, doi:10.1007/BF00705546. → pages 14, 21, 22

Sukoriansky, S., B. Galperin, and I. Staroselsky, 2005: A Quasinormal Scale Elimination Model of Turbulent Flows with Stable Stratification. *Physics of Fluids*, **17 (8)**, 085 107, doi:10.1063/1.2009010. → pages 14, 32

Taylor, K. E., 2001: Summarizing Multiple Aspects of Model Performance in a Single Diagram. *Journal of Geophysical Research*, **106 (D7)**, 7183, doi:10.1029/2000JD900719. → pages 16

United States Department of Energy, 2015: Wind Vision: A New Era for Wind Power in the United States. Tech. rep. → pages 2

Wang, X., and C. H. Bishop, 2003: A Comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes. *Journal of the Atmospheric Sciences*, **60 (9)**, 1140–1158, doi:10.1175/1520-0469(2003)060⟨1140:ACOBAE⟩2.0.CO;2. → pages 43

Warner, T. T., 2011: *Numerical Weather and Climate Prediction*. Cambridge University Press, 550 pp. → pages 2, 5, 11, 45

Webb, E. K., 1970: Profile Relationships: The Log-Linear Range, and Extension to Strong Stability. *Quarterly Journal of the Royal Meteorological Society*, **96 (407)**, 67–90, doi:10.1002/qj.49709640708. → pages 6, 53

Whitaker, J. S., and A. F. Loughe, 1998: The Relationship between Ensemble Spread and Ensemble Mean Skill. *Monthly Weather Review*, **126 (12)**, 3292–3302, doi:10.1175/1520-0493(1998)126⟨3292:TRBESA⟩2.0.CO;2. → pages 43

Whiteman, C. D., and S. W. Hoch, 2014: Pseudovertical Temperature Profiles in a Broad Valley from Lines of Temperature Sensors on Sidewalls. *Journal of Applied Meteorology and Climatology*, **53 (11)**, 2430–2437, doi:10.1175/JAMC-D-14-0177.1. → pages 58

Whiteman, C. D., J. M. Hubbe, and W. J. Shaw, 2000: Evaluation of an Inexpensive Temperature Datalogger for Meteorological Applications. *Journal of Atmospheric and Oceanic Technology*, **17 (1)**, 77–81, doi:10.1175/1520-0426(2000)017⟨0077:EOAITD⟩2.0.CO;2. → pages 58

Wilczak, J., and Coauthors, 2015: The Wind Forecast Improvement Project (WFIP): A Public-Private Partnership Addressing Wind Energy Forecast Needs. *Bulletin of the American Meteorological Society*, **96 (10)**, 1699–1718, doi:10.1175/BAMS-D-14-00107.1. → pages 5, 29, 77

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 704 pp. → pages 16, 36, 38, 39

Yang, B., and Coauthors, 2016: Sensitivity of Turbine-Height Wind Speeds to Parameters in Planetary Boundary-Layer and Surface-Layer Schemes in the Weather Research and Forecasting Model. *Boundary-Layer Meteorology*, 1–26, doi:10.1007/s10546-016-0185-2. → pages 9

Yang, Q., L. K. Berg, M. Pekour, J. D. Fast, R. K. Newsom, M. Stoelinga, and C. Finley, 2013: 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 (8)**, 1753–1763, doi:10.1175/JAMC-D-12-0267.1. → pages 9, 12

Zhang, D., and R. A. Anthes, 1982: A High-Resolution Model of the Planetary Boundary Layer—Sensitivity Tests and Comparisons with SESAME-79 Data. *Journal of Applied Meteorology*, **21 (11)**, 1594–1609, doi:10.1175/1520-0450(1982)021⟨1594:AHRMOT⟩2.0.CO;2. → pages 6, 53, 76

Zhang, Y., J. Wang, and X. Wang, 2014: Review on Probabilistic Forecasting of Wind Power Generation. *Renewable and Sustainable Energy Reviews*, **32**, 255–270, doi:10.1016/j.rser.2014.01.033. → pages 5, 29, 31

Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Mylne, 2002: The Economic Value of Ensemble-Based Weather Forecasts. *Bulletin of the American Meteorological Society*, **83 (1)**, 73–83, doi:10.1175/1520-0477(2002)083⟨0073:TEVOEB⟩2.3.CO;2. → pages 28

Appendix A

Summary of Bias-Corrected Forecast Accuracy Scores

Tables A.1 and A.2 provide the full MAE dataset for wind-farm sites 1-4. MAE is provided for the annual and seasonal averages, for each PBL-grid combination, and for using the GFS (Table A.1) and NAM (Table A.2) as initial-condition sources for the bias-corrected wind forecasts.

Fig. A.1 shows the annual MAESS of the bias-corrected wind forecasts (skill score is relative to the corresponding raw forecast, with larger positive scores being better) at each station using the DMB post-processing method. For sites 1-3 improvement in MAE due to bias correction ranged from 0-12% over MAE of the raw forecasts. While this 12% may seem small, the resulting monetary value can be large (Marquis et al., 2011). Site 4 saw large improvements (up to 56% for the 4-km grid using the MRF PBL scheme) due to large systematic biases. While the bias correction scheme applied to some forecasts did make the MAE up to 3% worse at individual sites, the majority of configurations were improved. I attribute the cause of poorer verification after bias correction at individual locations to the training dataset and to the fact that some grids (mostly of 12-km grid length) have low bias prior to bias-correction (Siuta et al., 2017a). When averaged over all locations, bias correction always resulted in improved accuracy skill.

Fig. A.2 shows the variation in the annual MAESS at all four wind-farm sites averaged over all 48 bias-corrected forecasts. While I do not show training periods less than 5 days in length, I tested a training period of only 1 day and found MAESS to be significantly worse (-40% or worse), thus it is not included in Fig. A.2. I find that the benefits of a longer training period level out between 20 and 30 days.

Table A.1: Mean Absolute Error (MAE) for each bias-corrected forecast initialized off the GFS. Sites 1-4 are the anonymous wind farm locations. Statistics are divided into five blocks, Annual, Summer (June - August), Fall (September - November), Winter (December - February), and Spring (March - May). MAE is in m s^{-1}.

Initial Condition	GFS																							
PBL Scheme	YSU			ACM2			MRF			MYJ			MYNN			QNSE			UW			GBM		
Grid Length (km)	36	12	4	36	12	4	36	12	4	36	12	4	36	12	4	36	12	4	36	12	4	36	12	4
Annual																								
Site 1	2.30	2.16	2.45	2.28	2.12	2.41	2.27	2.20	2.46	2.34	2.15	2.50	2.35	2.26	2.52	2.41	2.24	2.62	2.40	2.19	2.53	2.35	2.15	2.47
Site 2	1.83	1.82	1.79	1.76	1.77	1.68	1.81	1.86	1.82	1.83	1.85	1.77	1.92	1.92	1.86	1.91	1.93	1.82	1.85	1.88	1.85	1.78	1.80	1.79
Site 3	1.81	1.77	1.77	1.78	1.72	1.73	1.77	1.73	1.69	1.78	1.71	1.79	1.87	1.81	1.80	1.89	1.81	1.87	1.86	1.77	1.83	1.75	1.72	1.76
Site 4	1.87	1.61	1.62	1.85	1.61	1.60	1.86	1.66	1.66	1.82	1.60	1.62	1.83	1.65	1.64	1.88	1.62	1.63	1.85	1.64	1.67	1.80	1.61	1.61
Summer																								
Site 1	2.04	1.99	2.14	1.95	1.90	2.08	1.97	1.91	2.03	2.10	2.01	2.14	2.10	2.03	2.13	2.22	2.14	2.28	2.14	2.03	2.21	2.09	2.00	2.15
Site 2	1.70	1.70	1.64	1.58	1.59	1.53	1.69	1.65	1.58	1.72	1.72	1.62	1.80	1.78	1.68	1.80	1.81	1.71	1.75	1.77	1.69	1.72	1.71	1.64
Site 3	1.64	1.61	1.60	1.56	1.53	1.52	1.66	1.62	1.55	1.61	1.57	1.56	1.75	1.70	1.64	1.72	1.67	1.68	1.71	1.65	1.60	1.62	1.61	1.58
Site 4	1.58	1.21	1.32	1.57	1.21	1.28	1.64	1.25	1.34	1.64	1.21	1.37	1.64	1.26	1.36	1.62	1.22	1.36	1.59	1.23	1.38	1.57	1.17	1.29
Fall																								
Site 1	2.32	2.13	2.36	2.32	2.12	2.34	2.31	2.19	2.43	2.34	2.10	2.41	2.36	2.30	2.51	2.42	2.21	2.52	2.42	2.22	2.51	2.38	2.14	2.42
Site 2	1.79	1.83	1.82	1.72	1.77	1.71	1.76	1.92	1.87	1.77	1.87	1.78	1.93	2.00	1.94	1.86	1.93	1.81	1.82	1.92	1.86	1.76	1.84	1.81
Site 3	1.75	1.70	1.72	1.70	1.62	1.66	1.73	1.68	1.68	1.68	1.63	1.76	1.81	1.73	1.77	1.80	1.72	1.87	1.78	1.71	1.82	1.68	1.62	1.71
Site 4	1.96	1.84	1.74	1.97	1.86	1.73	1.95	1.90	1.79	1.86	1.82	1.72	1.89	1.88	1.74	1.97	1.88	1.76	1.99	1.91	1.81	1.85	1.82	1.73
Winter																								
Site 1	2.67	2.45	2.98	2.76	2.45	3.01	2.76	2.62	3.14	2.75	2.44	3.10	2.78	2.62	3.10	2.71	2.47	3.14	2.84	2.44	3.06	2.80	2.40	2.99
Site 2	2.03	1.90	1.92	2.03	1.93	1.80	2.05	2.05	2.05	2.01	1.97	1.87	2.11	2.01	1.94	2.04	1.99	1.87	2.01	1.98	1.95	1.90	1.87	1.88
Site 3	1.96	1.85	1.79	2.05	1.89	1.87	1.92	1.85	1.79	1.96	1.81	1.91	2.02	1.90	1.84	2.03	1.85	1.91	1.98	1.81	1.91	1.86	1.78	1.83
Site 4	2.16	1.97	1.98	2.10	1.93	1.94	2.06	1.99	2.02	2.08	1.98	1.97	2.07	1.98	1.98	2.15	1.97	1.95	2.10	1.95	1.99	2.03	1.96	1.96
Spring																								
Site 1	2.16	2.06	2.32	2.11	2.00	2.21	2.04	2.07	2.25	2.17	2.03	2.36	2.17	2.09	2.33	2.29	2.16	2.54	2.20	2.09	2.36	2.16	2.06	2.33
Site 2	1.82	1.84	1.78	1.74	1.77	1.69	1.76	1.81	1.80	1.80	1.85	1.82	1.85	1.88	1.86	1.94	1.99	1.89	1.82	1.87	1.92	1.76	1.79	1.82
Site 3	1.90	1.93	1.95	1.83	1.83	1.87	1.77	1.75	1.75	1.88	1.83	1.91	1.89	1.90	1.93	2.01	1.97	2.03	1.95	1.92	2.00	1.85	1.85	1.91
Site 4	1.68	1.33	1.35	1.67	1.35	1.34	1.72	1.41	1.39	1.64	1.29	1.35	1.68	1.38	1.38	1.71	1.31	1.37	1.67	1.35	1.42	1.67	1.35	1.38

Table A.2: Mean Absolute Error (MAE) for each bias-corrected forecast initialized off the NAM. Sites 1-4 are the anonymous wind farm locations. Statistics are divided into five blocks, Annual, Summer (June - August), Fall (September - November), Winter (December - February), and Spring (March - May). MAE is in m s^{-1}.

Initial Condition	NAM																							
PBL Scheme	YSU			ACM2			MRF			MYJ			MYNN			QNSE			UW			GBM		
Grid Length (km)	36	12	4	36	12	4	36	12	4	36	12	4	36	12	4	36	12	4	36	12	4	36	12	4
Annual																								
Site 1	2.39	2.17	2.48	2.37	2.14	2.47	2.32	2.19	2.49	2.40	2.15	2.53	2.39	2.23	2.52	2.47	2.27	2.65	2.49	2.23	2.56	2.42	2.16	2.47
Site 2	1.81	1.83	1.76	1.75	1.80	1.70	1.81	1.86	1.79	1.81	1.87	1.78	1.88	1.89	1.83	1.90	1.95	1.84	1.84	1.88	1.85	1.74	1.81	1.81
Site 3	1.87	1.81	1.80	1.83	1.75	1.76	1.81	1.77	1.72	1.85	1.76	1.84	1.90	1.84	1.82	1.92	1.82	1.89	1.88	1.80	1.86	1.82	1.77	1.79
Site 4	1.88	1.63	1.64	1.85	1.63	1.63	1.88	1.66	1.67	1.82	1.62	1.67	1.85	1.67	1.67	1.88	1.63	1.66	1.86	1.68	1.71	1.82	1.64	1.65
Summer																								
Site 1	2.07	1.95	2.13	1.99	1.86	2.08	2.01	1.90	2.09	2.09	1.94	2.16	2.12	1.99	2.17	2.18	2.06	2.29	2.16	2.02	2.20	2.10	1.95	2.11
Site 2	1.73	1.72	1.70	1.63	1.66	1.62	1.73	1.74	1.67	1.80	1.72	1.70	1.84	1.76	1.74	1.86	1.84	1.77	1.84	1.80	1.76	1.74	1.72	1.71
Site 3	1.80	1.80	1.71	1.70	1.69	1.63	1.75	1.75	1.66	1.77	1.74	1.69	1.88	1.84	1.76	1.81	1.76	1.75	1.84	1.78	1.69	1.78	1.75	1.68
Site 4	1.60	1.26	1.36	1.58	1.28	1.35	1.64	1.27	1.36	1.66	1.27	1.43	1.66	1.31	1.41	1.63	1.27	1.41	1.61	1.30	1.43	1.58	1.23	1.34
Fall																								
Site 1	2.44	2.19	2.51	2.45	2.15	2.48	2.37	2.19	2.52	2.46	2.16	2.52	2.42	2.29	2.59	2.52	2.28	2.66	2.53	2.27	2.62	2.50	2.20	2.52
Site 2	1.75	1.83	1.74	1.71	1.82	1.69	1.74	1.89	1.81	1.74	1.88	1.75	1.88	1.95	1.82	1.84	1.94	1.84	1.78	1.93	1.86	1.71	1.88	1.79
Site 3	1.80	1.73	1.77	1.75	1.67	1.70	1.80	1.75	1.75	1.74	1.64	1.77	1.88	1.82	1.85	1.85	1.73	1.87	1.84	1.74	1.87	1.74	1.68	1.77
Site 4	2.00	1.86	1.77	2.01	1.86	1.74	1.98	1.87	1.80	1.87	1.80	1.74	1.95	1.91	1.78	2.00	1.85	1.76	1.99	1.92	1.80	1.91	1.84	1.75
Winter																								
Site 1	2.73	2.39	2.90	2.81	2.45	3.01	2.74	2.56	3.06	2.77	2.42	3.06	2.80	2.52	2.99	2.75	2.48	3.12	2.92	2.45	2.98	2.81	2.39	2.94
Site 2	1.95	1.92	1.82	1.91	1.92	1.75	1.98	2.00	1.92	1.93	1.98	1.83	2.02	1.99	1.89	1.98	2.02	1.82	1.93	1.96	1.91	1.81	1.84	1.86
Site 3	1.98	1.82	1.76	2.05	1.84	1.84	1.94	1.83	1.74	1.99	1.81	1.92	2.03	1.89	1.80	2.04	1.84	1.91	1.96	1.80	1.91	1.91	1.76	1.77
Site 4	2.11	1.92	1.95	2.05	1.88	1.90	2.06	1.96	1.97	2.02	1.94	1.96	2.04	1.96	1.95	2.07	1.95	1.97	2.04	1.96	2.00	2.01	1.96	1.98
Spring																								
Site 1	2.32	2.17	2.40	2.25	2.09	2.32	2.16	2.11	2.27	2.26	2.1	2.37	2.22	2.14	2.33	2.41	2.27	2.55	2.34	2.18	2.44	2.25	2.11	2.33
Site 2	1.83	1.85	1.80	1.76	1.82	1.72	1.78	1.81	1.77	1.77	1.89	1.86	1.79	1.84	1.85	1.93	2.00	1.95	1.80	1.85	1.88	1.70	1.80	1.86
Site 3	1.89	1.90	1.95	1.80	1.81	1.88	1.75	1.73	1.74	1.88	1.86	1.98	1.83	1.82	1.88	1.99	1.94	2.03	1.88	1.89	1.96	1.86	1.87	1.95
Site 4	1.73	1.37	1.42	1.70	1.42	1.43	1.78	1.45	1.47	1.71	1.36	1.46	1.71	1.42	1.46	1.75	1.35	1.42	1.72	1.43	1.52	1.71	1.43	1.47

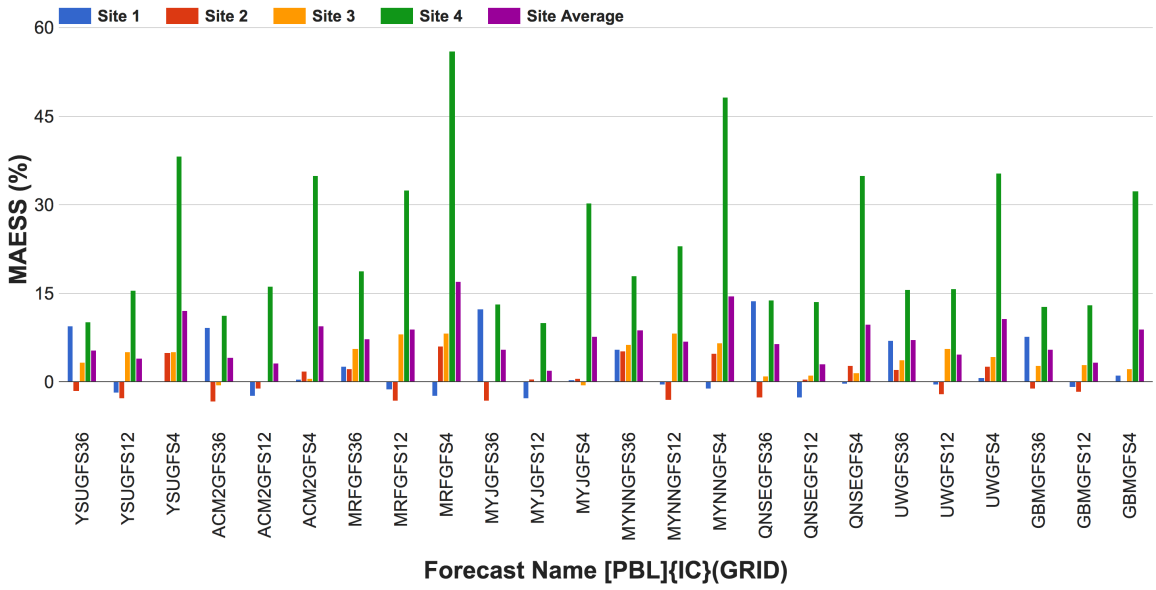

Figure A.1: Annual MAE skill score (MAESS), showing improvement in MAE resulting from bias correction. Skill is relative to the equivalent raw wind forecast. Colors represent individual wind farm sites, with purple bars indicating the site-averaged performance. Larger positive values are better.

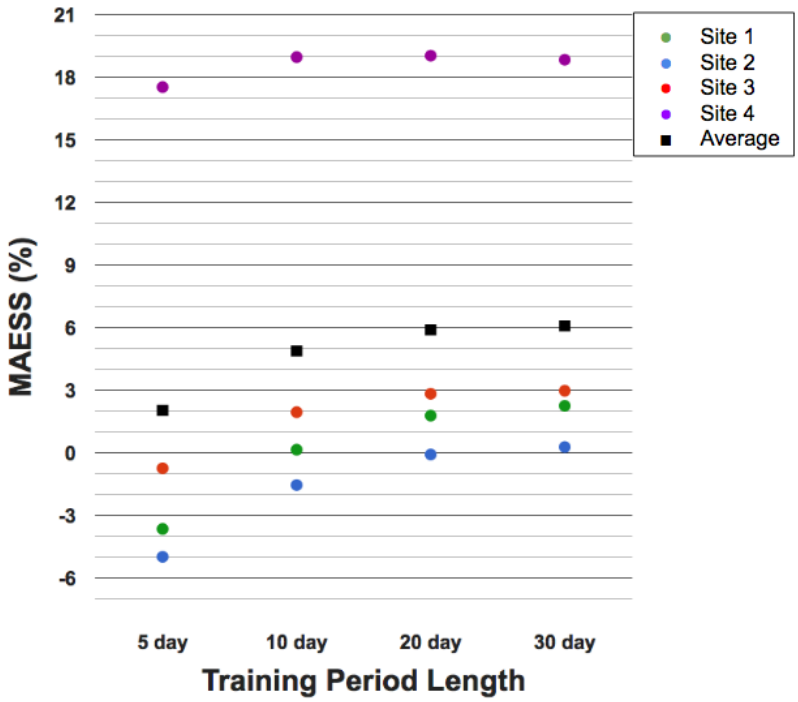

Figure A.2: Effect of the training period length on forecast accuracy (MAESS), averaged over all 48 bias-corrected deterministic wind forecasts at each wind-farm site.

Appendix B

The pq Probability Distribution

The meteorological community is introduced to a new frequency distribution that was devised for Chapter 3. The distribution has a simple un-normalized one-sided form:

$$f_s(x) = (1 - x^p)^q \tag{B1}$$

where p and q are the shape parameters and the distribution is defined between 0 and 1. It can be made into a normalized (unit area) symmetric probability density (f) with arbitrary scaling parameter S and center-location parameter x_o:

$$f(x) = \frac{1}{A}\left[1 - \left|\frac{x - x_o}{S}\right|^p\right]^q \quad \text{for} \quad (x_o - S) \leq x \leq (x_o + S) \tag{B2}$$

where the area A under the un-normalized symmetric curve is

$$A = \frac{2S\Gamma\left(q+1\right)\Gamma\left(1/p\right)}{p\Gamma\left[q+1+(1/p)\right]} \tag{B3}$$

and where Γ is the gamma function. The pq distribution is similar, but not identical, to the Kumaraswamy (1980) distribution and the beta distribution (Jambunathan, 1954).

The pq distribution has a theoretical mean absolute deviation of

$$MAD = S\frac{\Gamma\left[q+1+(1/p)\right]\Gamma\left(2/p\right)}{\Gamma\left[q+1+(2/p)\right]\Gamma\left(1/p\right)} \tag{B4}$$

and variance of

$$\sigma_x^2 = S^2\frac{\Gamma\left[q+1+(1/p)\right]\Gamma\left(3/p\right)}{\Gamma\left[q+1+(3/p)\right]\Gamma\left(1/p\right)}. \tag{B5}$$

The fourth statistical moment is

$$\mu_4 = S^4\frac{\Gamma\left[q+1+(1/p)\right]\Gamma\left(5/p\right)}{\Gamma\left[q+1+(5/p)\right]\Gamma\left(1/p\right)} \tag{B6}$$

from which the kurtosis can be found as μ_4/σ_x^4. Skewness is zero for this symmetric distribution. Sometimes a situation may exist where part of the distribution tail on one or both sides must be cut off. For this situation, numerically integrate to get the area under the curve and to get the higher moments, using evenly-spaced samples from (B2) instead of using the theoretical expressions (B3) - (B6).

If you are given an observed distribution with scatter, then you can find the best-fit parameters (p, q, S, x_o) for the pq distribution as follows. If you know apriori (or want to fix) one or more of the parameters (typically S and/or x_o), then do so and find the remaining parameters by one of two methods. You can use the method of moments, where you first calculate the statistical moments of the observation data, and then use the the Newton-Raphson or other root-finding algorithm to find the pq-distribution parameters that give the same statistical moments. You will need as many statistical moments as unknown parameters.

A second approach is brute force, which is used here to help learn more about the sensitivity of the distribution to the choice of parameters. Namely, I loop through all reasonable values of p and q, and for each (p, q) combination to find the mean absolute error between the pq distribution and the observed frequency distribution (after normalizing the observations to have unit area under its histogram). The best (p, q) combination is the one with minimum error. Fig. B.1 shows an example for one of the wind-farm forecast-error distributions.

While stepping through each (p, q) combination, a plot of the pq-distribution MAD, variance, and kurtosis is produced (Fig. B.2). You can use this as a poor-mans method of moments. Namely, if you know the variance and kurtosis (or the MAD and kurtosis) from your observation data, then you can use those values in Fig. B.2 to find the approximate p and q values.

Although the pq distribution cannot fit skewed distributions, its versatility for symmetric distributions is remarkable. Fig. B.3 shows extremes that approximate a delta function, a linear ramp shape, and a near-uniform distribution. Fig. B.4 shows that q controls the slope of the tails, while p controls the shape of center of the distribution. The last frame in Fig. B.4 shows how the pq distribution can approximate Gaussian distributions, particularly for values of $p \approx 1.85$.

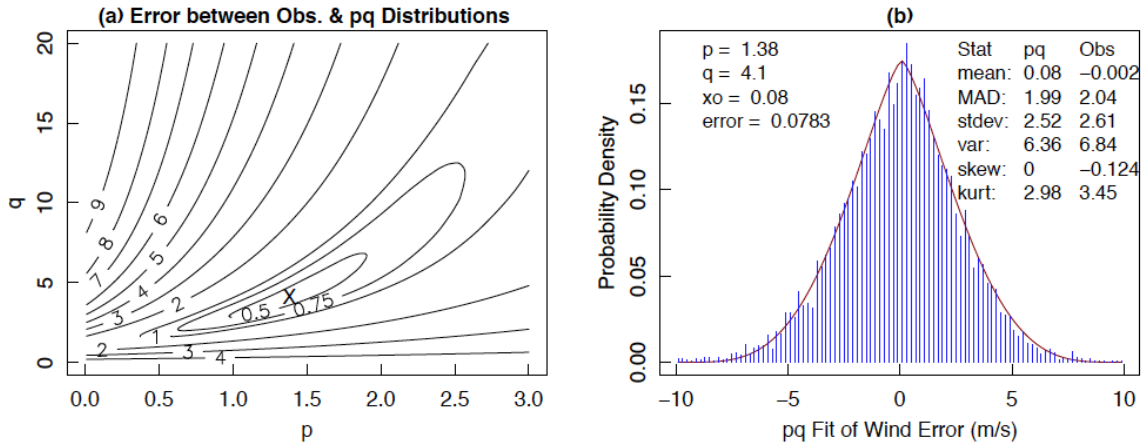

Figure B.1: (a) Relative error (contoured, arbitrary units) between the observed and pq distributions as a function of parameters p and q. X marks the minimum error; namely, the best p and q values. S was fixed apriori at 10 m s^{-1} wind-forecast error, and p, q, and x_o were varied to find the distribution errors. (b) Resulting best-fit pq distribution (curve) to the observed histogram of wind-speed forecast errors (vertical bars). Note that this fitting method of minimizing the mean absolute error does not overly weight the tails of the distribution.

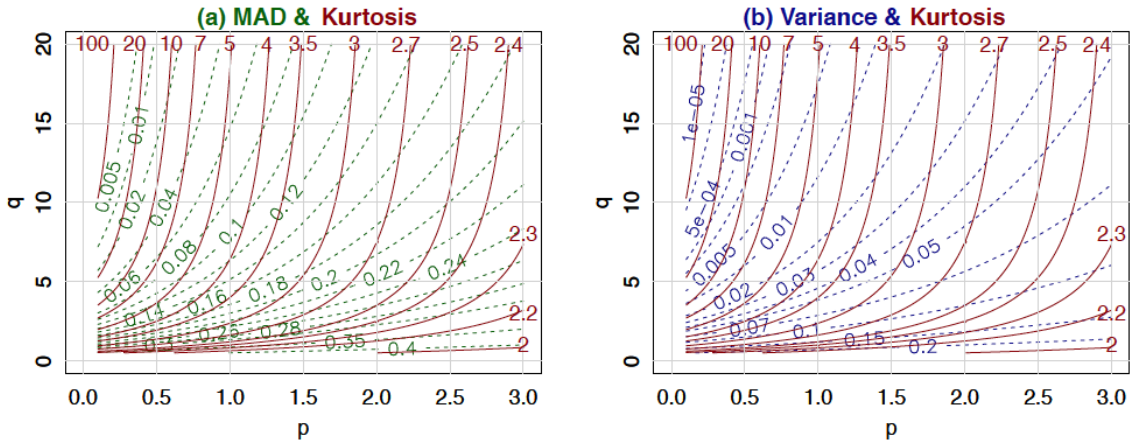

Figure B.2: Contour plots of (a) MAD (dashed line) and kurtosis (solid line), and (b) variance (dashed line) and kurtosis (solid line) for the pq distribution as a function of the p and q shape parameters.

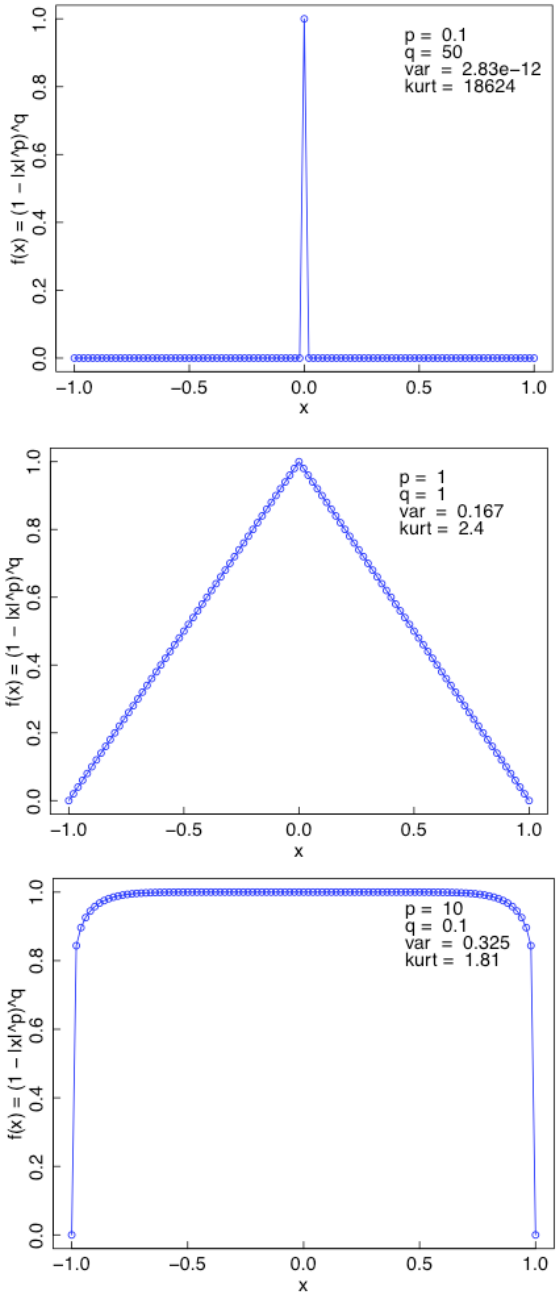

Figure B.3: Some extreme examples of the pq distribution.

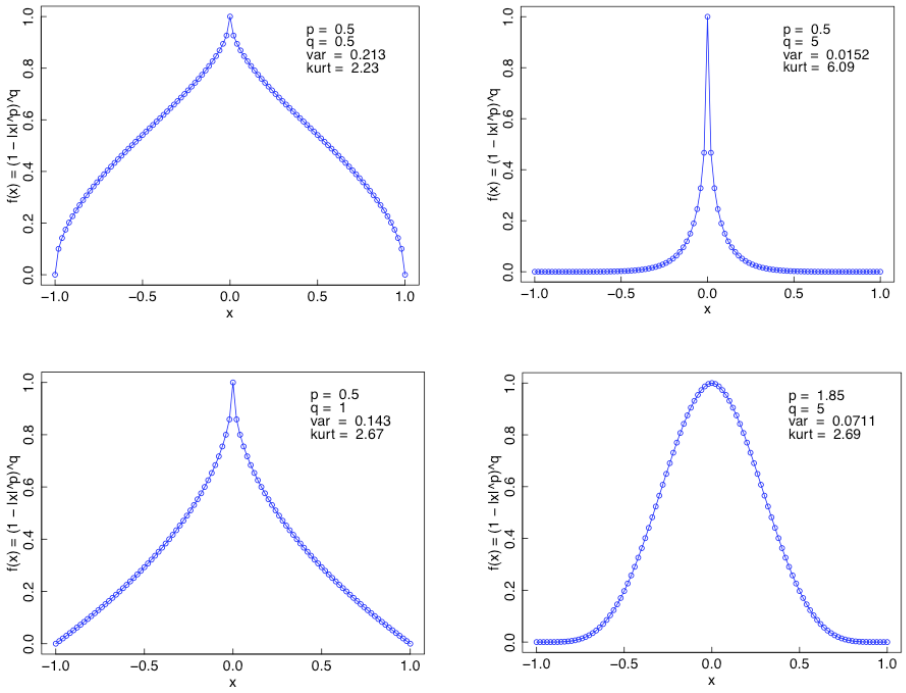

Figure B.4: Sample of some of the shapes produced by the pq distribution. The last shape is approximately Gaussian.

Appendix C

Terrain cross sections

To provide more insight into the terrain characteristics of the anonymous wind-farm locations, I have provided terrain cross sections for all four locations (Figs. C1-C4) in this appendix. Each cross section is approximately 80 to 90 km in the horizontal, taken out of the predominant wind direction. Four plots are provided in each of Figs. C1-C4. The panels are cross sections taken from a digital elevation model (DEM) at 30s resolution, the WRF-model terrain from the 4-km grid, the WRF-model terrain from the 12-km grid, and the WRF-model terrain from the 36-km grid. To maintain confidentiality of the locations, actual terrain elevations are not shown. Instead, the elevation is plotted with respect to the lowest elevation available in the cross section area. I also do not link the images to any of the specific site numbers used in Chapter 2-4.

Figs. C1-C4 show that the wind-farms are located along mountain ridges in the DEM. In addition, all of the wind farms have higher terrain in their immediate vicinity (e.g., within 40-50 km in the horizontal). Likewise, all locations show that the 4-km grids do represent the general shape of the terrain, but do not capture many of the small valleys and highest peaks. Nor do the 4-km WRF grids match the correct elevation. For the 12- and 36-km grids, the model terrain often takes the shape of an elongated slope, not putting the wind-farms on any type of peak.

In Chapter 4, I find non-dimensional wind shear and temperature gradient values are smaller than accepted theory for statically-stable conditions. Because of the complexity of the terrain shown in Figs. C1-C4, it is possible that internal boundary layers generated by roughness features upstream could be one of the causes on enhanced turbulence during statically-stable conditions. Dörenkämper et al. (2015) found similar internal boundary layers could be advected several 10s of kilometers over water surfaces. Also, under statically-stable conditions, mountain wave activity generated off of any of the mountain tops shown in the cross sections may also result in enhanced turbulence. Finally, slope flows are a common feature in complex terrain. Upslope and downslope flows will result in enhanced vertical mixing near mountain tops due to flow convergence or divergence, depending on the time of day. Future work should be focused on determining the effects of these features on surface-layer similarity theories.

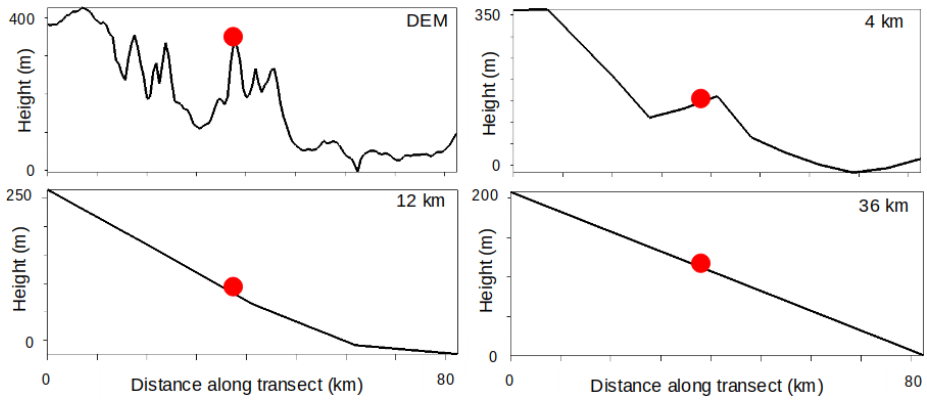

Figure C.1: Representation of terrain from a digital elevation model (DEM), the 4-km WRF grid, the 12-km WRF grid, and the 36-km WRF grid at anonymous site 1 (site label does not match other chapters). The red dot represents the wind farm location.

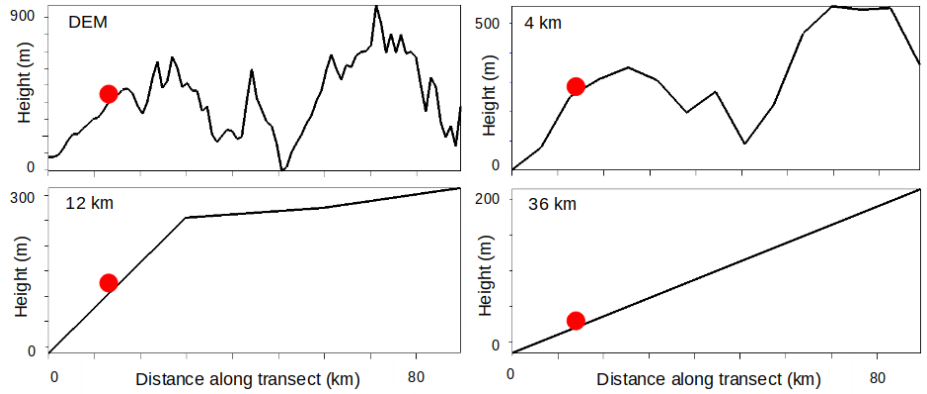

Figure C.2: Representation of terrain from a digital elevation model (DEM), the 4-km WRF grid, the 12-km WRF grid, and the 36-km WRF grid at anonymous site 2 (site label does not match other chapters). The red dot represents the wind farm location.

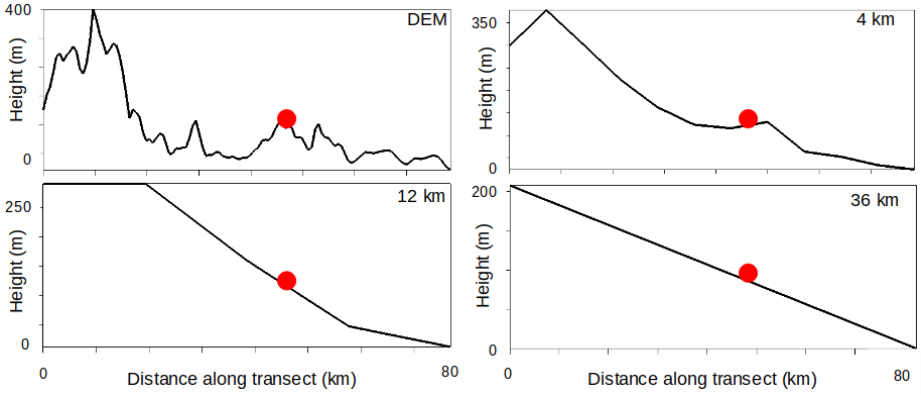

Figure C.3: Representation of terrain from a digital elevation model (DEM), the 4-km WRF grid, the 12-km WRF grid, and the 36-km WRF grid at anonymous site 3 (site label does not match other chapters). The red dot represents the wind farm location.

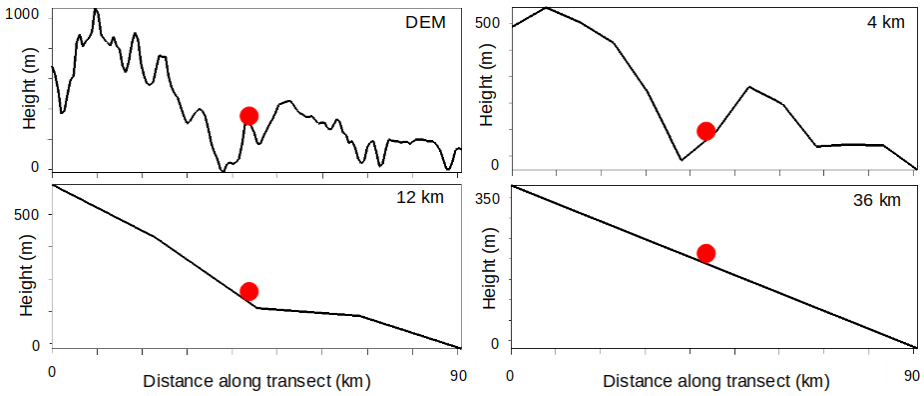

Figure C.4: Representation of terrain from a digital elevation model (DEM), the 4-km WRF grid, the 12-km WRF grid, and the 36-km WRF grid at anonymous site 4 (site label does not match other chapters). The red dot represents the wind farm location.