The Pennsylvania State University

The Graduate School

College of Earth and Mineral Sciences

A SUB-KM-GRIDLENGTH ENSEMBLE FOR REPRESENTING MESOGAMMA HAZARD-PREDICTION UNCERTAINTY IN THE STABLE BOUNDARY LAYER OVER COMPLEX TERRAIN

A Thesis in

Meteorology

by

Eric Wendoloski

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2015

The thesis of Eric Wendoloski was reviewed and approved* by the following:

David R. Stauffer
Professor of Meteorology
Thesis Advisor

George Young
Professor of Meteorology and GeoEnvironmental Engineering

David J. Stensrud
Department Head and Professor of Meteorology

Johannes Verlinde
Professor of Meteorology
Associate Head, Graduate Program in Meteorology

*Signatures are on file in the Graduate School

ABSTRACT

Atmospheric transport and dispersion (AT&D) forecasting in the stable boundary layer over complex terrain is inherently uncertain. An ensemble prediction system with sub-km horizontal grid spacing and high vertical resolution is used with the goal of quantifying this uncertainty. Diversity in initial conditions and/or planetary boundary layer/surface layer physics within the Weather Research and Forecasting (WRF) model provides ensembles with up to 12 members. WRF explicit ensemble member data drive trajectory calculations and the Second-Order Closure Integrated Puff (SCIPUFF) model for hazard prediction. Explicit ensemble outcomes from SCIPUFF are compared against SCIPUFF forecasts driven by single members derived from the WRF ensemble and capable of leveraging ensemble wind-field uncertainty statistics. Ensemble performance is evaluated for both 1.3-km and 0.4-km horizontal-grid-spacing ensemble configurations for three case studies of differing flow regimes with respect to the Nittany Valley in central Pennsylvania. AT&D uncertainty is expected to be dependent on small-scale drainage flows and circulations related to trapped-lee wave activity over the valley.

Results demonstrate that a 12-member ensemble provides reasonable spread in AT&D outcomes. Additionally, single-member SCIPUFF forecasts reflect much of the spatial spread and impact probability given by explicit SCIPUFF ensemble forecasts but at a reduced computational cost. When evaluated against meteorological observations from the Rock Springs observation network within the valley, ensemble performance statistics demonstrate that low-level wind and temperature forecasts exhibit a statistically significant improvement in the 0.4-km configuration forecasts over the 1.3-km configuration forecasts. Analysis reveals that sub-km horizontal grid spacing better captures temperature and wind fluctuations related to drainage flows and trapped-lee wave activity that directly impact AT&D, and in general, the sub-km ensemble reliably samples the true state of these low-level variables. The best forecast skill is given by the full 12-

member ensemble configuration involving both initial-condition and physics diversity. In the trapped-lee wave cases, the use of localized data assimilation positively impacts overall probabilistic forecast skill while the drainage flow case appears more dependent on model physics than initialization strategy with the available observations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to first thank my advisor, Dr. David Stauffer, for his support, mentorship, and patience as I completed this thesis. His guidance and infectious love for science has allowed me to become a far more competent and capable scientist, and for that, I will be ever grateful. I would also like to thank Dr. George Young and Dr. David Stensrud for their insightful comments and suggestions during the thesis writing process. Thanks is also extended to Glenn Hunter, Dr. A.J. Deng, and Dr. Brian Gaudet for their useful discussions and insight particularly in the early stages of this research. Additionally, I would especially like to thank Astrid Suarez whose knowledge and intuition has been truly invaluable throughout this process.

This research was funded by the Defense Threat Reduction Agency under DTRA Grant No. HDTRA1-10-1-0033 under the supervision of Dr. John Hannan. Funding for the SODARs was provided by the US Army Research Office under DURIP Grant No. W911NF-10-1-0238.

CHAPTER 1

INTRODUCTION

The stable boundary layer (SBL), a strongly stratified layer that is often meters to tens of meters deep, generally develops within 30 minutes of sunset (Andre et al. 1978). Radiative cooling under clear sky conditions with weak synoptic forcing commonly promotes SBL generation. Because the SBL is characterized by the development of a surface temperature inversion, it decouples from the statically neutral residual layer remaining from the daytime. The residual layer is bounded by the SBL below and an inversion above (Stull 1988). Strong stratification in the SBL is often associated with quiescent surface conditions and intermittent turbulence (Stull 1988; Mahrt 1999). Mahrt and Vickers (2002) describe this boundary layer as "upside down" because turbulent bursts and subsequent mixing are often due to mechanical shear generation aloft as opposed to mixing related to surface forcing as within the daytime convective boundary layer. Intermittent turbulent bursting has been attributed to shear generating phenomena such as low-level jets and internal-gravity waves (e.g., Salmond and McHendry 2005). Sub-mesoscale (submeso) motions, non-turbulent motions generated by non-stationary shear events with time scales of minutes to tens of minutes and spatial scales from 0.02 km to 2 km, are also associated with shear generation in the SBL (e.g., Seaman et al. 2012, Hoover et al. 2015, Mahrt 2014).

Turbulent mixing can have a direct and complex impact on the atmospheric transport and dispersion (AT&D) of contaminants released within the SBL. Because of the stable stratification, contaminants and pollutants released within the SBL generally disperse horizontally as meandering plumes (Etling 1990; Mahrt 2007). During turbulent mixing events, dispersion may be enhanced in the vertical, and the SBL may even temporarily couple with the residual layer

aloft where neutral stratification allows pollutants to disperse equally in the horizontal and vertical (Salmond and McKendry 2005). Such AT&D processes can be further complicated by the influence of complex terrain (Nappo 1991, Hoover et al. 2015).

Common submeso motions within the SBL in complex terrain include wave-like motions, solitary waves, quasi-horizontal modes, microfronts, intermittent drainage flows, and more complex structures as surveyed in Mahrt (2014). Depending on wind speed and stability, mountain waves may be excited by flow over topography. Boundary layer separation processes related to mountain wave activity may lead to the development of rotors (e.g., Scorer 1949; Kuettner 1959; Doyle and Durran 2002). Hertenstein and Kuettner (2005) describe rotors as either Type 1 or Type 2. Type 1 rotors, characterized by moderate to severe turbulence, can develop when mountain waves become trapped close to the ground in a ducting region and induce pressure gradients that lead to circulations collocated beneath the crests of a lee wave (Doyle and Durran 2002). Type 2 rotors, characterized by severe to extreme turbulence and downslope windstorms, are associated with highly nonlinear processes such as hydraulic jumps downwind of a mountain crest (Chow et al. 2013). Regardless of the mechanism, rotors serve as an additional source of turbulent mixing and, in addition to creating hazardous conditions for aviation, play a complex role in AT&D of pollutants over regions of mountainous terrain (Durran 1990; Chow et al. 2013; Markowski and Richardson 2012).

Using numerical weather prediction (NWP) to forecast the outcome of a chemical or biological (chem-bio) release within the SBL is a task inherently fraught with uncertainty. Estimations of the initial state of the atmosphere, temporal and spatial model numerics, and physical parameterizations necessary for the prediction of subgrid-scale processes are all sources of uncertainty in an NWP model (e.g., Stauffer 2013). Whether a chem-bio release is from factory emissions, an industrial accident, or a terrorist attack, the implications for human life in the path

of a plume may be dire. Thus, it is essential to quantify the uncertainty surrounding an AT&D forecast in order to aid emergency managers in the decision-making process.

No single deterministic NWP forecast can perfectly cover the full range of this uncertainty (Murphy 1988; Leutbecher and Palmer 2007). One method that addresses this problem and estimates forecast uncertainty is the use of ensembles. Ensemble modeling involves multiple numerical simulations with the goal of spanning the range of plausible outcomes of the future atmospheric state. By perturbing various aspects of an NWP model, one can create a suitably large number of equally likely ensemble members (Leith 1974). Each individual, or explicit, ensemble member can be viewed as a separate deterministic forecast, and when aggregated, the ensemble of explicit members can be used to compute the probability of a particular event, and the ensemble skill can be validated by probabilistic skill scores (Marsigli et al. 2001). Single members derived from explicit ensemble members may be useful in operational forecasting as well as for highly time-sensitive forecasts depending on the application (e.g., AT&D forecasts following a chem-bio release). Although dynamic consistency is not assured, the ensemble mean, the average of explicit member outcomes, tends to provide a more skillful forecast than any one explicit member (e.g., Grimit and Mass 2002). To avoid dynamic inconsistency and over-smoothing issues of the ensemble mean, a "best member" may be defined. For this study, a best member refers to the deterministic member most similar to the ensemble mean for those meteorological parameters deemed vital to a particular forecast application.

There is no one universal method of creating ensemble members; however some methods include the use of breeding vectors (Toth and Kalnay 1997), singular vectors (Molteni et al. 1996), the ensemble Kalman filter (EnKF; Houtekamer and Mitchell 2005), the Ensemble Transform and Rescaling method (ETR, Ma et al. 2014), or using deterministic forecasts combined with past observations to create analog ensembles (e.g., Delle Monache et al. 2013). These methods often produce ensemble forecasts based on a variety of possible initial

atmospheric states with a focus usually on synoptic-scale features. However, mesoscale forecasts of shorter length (6-48 h) may not allow sufficient time for initial-condition perturbations to grow leading to a lack of spread among ensemble members (e.g., Stensrud et al. 2000). One approach to increase spread over shorter forecast lengths is the use of multi-model ensembles that combine output from more than one dynamical core or model physics scheme to achieve more spread among the forecasts (e.g., National Centers for Environmental Prediction Short Range Ensemble Forecast (NCEP-SREF); Du et al. 2014). Stensrud and Fritsch (1994) suggest that ensemble diversity in both initial conditions and model physics may be appropriate because of uncertainties in both of these aspects. Under weak synoptic forcing, Stensrud et al. (2000) demonstrate improved short-range mesoscale forecast skill when multi-physics diversity is included.

Ensembles for AT&D applications commonly include model, physics, or initial-condition diversity or combinations of the three (e.g., Warner et al. 2002; Delle Monache and Stull 2003; Jimenez-Guerrero et al. 2013). Lei et al. (2012) demonstrate improved AT&D forecasts based on Weather Research and Forecasting (WRF; Skamarock et al. 2008) ensembles including both initial-condition and physics diversity, and using continuous data assimilation methods produced reduced observation insertion noise levels compared to intermittent methods. Some AT&D ensembles produce spread by varying a release in space, time, or both (e.g., Draxler 2003; Bieringer et al. 2014).

Regardless of the method used to create a set of ensemble members, creating an ensemble that provides reliable forecasts of uncertainty may require at least several hundred members (e.g., Kolczynski et al. 2011). Here, the term reliable refers to an ensemble that produces an event probability equal to that of the true probability. The fewer number of members, a common limitation of the approach, contributes to the underdispersive nature of most ensembles (e.g., Kolczynski et al. 2011). Fortunately, previous studies have noted statistical improvements in forecast skill while employing ensembles of only 5-10 members (e.g., Leith 1974; Houtekamer

and Derome 1995; Du et al. 1997; Buizza and Palmer 1998; Grimit and Mass 2002). Because of the high computational cost of running an ensemble prediction system, the number of ensemble members in operational settings generally ranges anywhere from 5-50 members (e.g., Eckel et al. 2010; Clark et al. 2011; Jirak et al. 2012) while some EnKF-based ensembles in research settings employ as many as 60-80 members (e.g. Wang et al. 2013; Munsell and Zhang 2014).

To compensate for a limited number of ensemble members, calibration techniques are often used to improve forecasts and allow the ensemble to more accurately represent actual forecast uncertainty. Some examples of calibration techniques are Bayesian Model Averaging (Raftery et al. 2005), Ensemble Model Output Statistics (Gneiting et al. 2005), and Linear Variance Calibration (Kolczysnki et al. 2009). Calibration is difficult in a non-operational case study mode as a relatively large dataset of prior ensemble forecasts must be available to properly weight members and correct for bias (e.g., Mass et al. 2009). Thus, in order to properly assess ensemble performance in a case study mode, computation of ensemble bias and spread will be included.

Another strategy to glean the best forecast skill from a relatively small ensemble is to choose the members that best represent the true uncertainty space. Lee et al. (2012) present a down-selection technique that utilizes principal component analysis in order to select the ensemble members that best represent the atmospheric state for a specific model application, region, and time period, and Molteni et al. (2001) employ a statistical clustering technique for down selection with a similar goal in mind. Although the down-selection results may vary by season, domain, and resolution (e.g., Lee et al. 2012), these approaches subsequently allow for fewer members with minimal or no performance degradation at lower computational cost when compared to a larger ensemble. As there is no set standard for ensemble member selection, a common method, also used in this thesis, is to simply select members based on physical intuition regarding the atmospheric processes of greatest importance to the problem at hand.

In this thesis, an ensemble approach is employed to quantify model uncertainty within the SBL over the complex terrain of the Nittany Valley in central Pennsylvania. In order to more accurately resolve the SBL and some of the small scale phenomena (mesogamma and below), members require very fine horizontal and vertical resolution uncommon to most current ensemble configurations and even most deterministic predictions. For example, some of the highest horizontal resolution mesoscale ensembles employed by operational forecast centers include the 4 - 5.15-km Storm Prediction Center Storm-Scale Ensemble of Opportunity (SSEO; Jirak et al. 2012), the 4-km Center for Analysis and Prediction ensemble (CAPS) that has included a single member using 1-km horizontal grid spacing (Xue et al. 2013), the 3-km High Resolution Rapid Refresh Ensemble (HRRRE; Benjamin 2014) slated for operations in 2016, the 2.2-km United Kingdom Met Office Global and Regional Ensemble Prediction System (MOGREPS; Mylne 2013), and the 1-km NCEP Storm-Scale Ensemble (NSSE) that is still at least five years from implementation (Du et al. 2014). One of the highest horizontal resolution deterministic forecasts produced operationally is by the United Kingdom Variable (UKV) resolution version of the Unified Model (Mylne 2013) that employs 1.5-km horizontal grid spacing. This thesis explores the viability and performance characteristics of a 3-12 member ensemble with sub-km (0.444-km) horizontal grid spacing and high vertical resolution designed for use in the SBL over complex terrain. Ensemble output is then used to compute trajectories and drive AT&D forecasts in a way that reflects the model uncertainty surrounding the true atmospheric state.

The remainder of this thesis will proceed as follows: Chapter 2 presents descriptions of the WRF model, ensemble design, and the Second-order Closure Integrated Puff (SCIPUFF) model for hazard prediction (Sykes et al. 2006). Experimental design and evaluation methods are summarized in Chapter 3. Chapter 4 presents the Rock Springs, PA observation network and descriptions of the real data cases analyzed. Case study results including model trajectories,

SCIPUFF forecasts, and statistical evaluations of ensemble performance are discussed in Chapter 5. Chapter 6 provides a summary, conclusions, and recommendations for future work.

CHAPTER 2

MODEL DESCRIPTIONS AND ENSEMBLE DESIGN

2.1 WRF Model Description and Configuration

An ensemble is created in order to quantify forecast uncertainty within the SBL over complex terrain. Use of the ensemble is extended to AT&D forecast uncertainty quantification under the same atmospheric conditions. Ensemble members utilized in this study are based on initial-condition and physics diversity within a high-resolution configuration of version 3.6 of the Advanced Research WRF (Skamarock et al. 2008).

WRF is a three-dimensional, fully-compressible, nonhydrostatic primitive-equation model with terrain-following vertical coordinates and Arakawa-C horizontal grid-point staggering (Skamarock et al. 2008). The WRF configuration includes four one-way nested domains of 12-km, 4-km, 1.333-km (1.3-km), and 0.444-km (0.4-km) horizontal grid spacing with the 1.3-km and 0.4-km nests centered over central Pennsylvania and the Nittany Valley (Fig. 2-1). The staggered vertical grid features 44 full levels, with 2-m spacing in the lowest 10 m and 10 half model levels in the lowest 50 m with gradually increasing thicknesses up to the model top at 50 hPa for all domains. This fine vertical resolution is required to better resolve SBL evolution including stable stratification and cold pool development (Seaman et al. 2012). Initial and lateral boundary conditions are given by 6 h NCEP 0.5^{o}x0.5^{o} Global Forecast System (GFS) analyses and forecasts that are spatially and temporally interpolated to the WRF model grids.

WRF contains a large suite of physics options, and only those most relevant to SBL evolution are varied among ensemble members. Physics options held constant among members include Rapid Radiative Transfer Model (RRTM; Mlawer et al. 1997) longwave radiation,

Dudhia (Dudhia 1989) shortwave radiation, 3-class simple ice (Hong et al. 2004) microphysics, Kain-Fritsch (Kain 2004) cumulus parameterization (12-km domain only), Noah land surface model (LSM, Chen and Dudhia 2001), and 20-category Moderate Resolution Imaging Spectroradiometer (MODIS; Friedl et al. 2002) land use. Atmospheric radiation calculations are updated every five minutes for all domains.

2.2 Ensemble Design

Given the high computational cost of running mesoscale ensembles, members are carefully chosen to provide the most relevant spread using the fewest members. Because of uncertainties in initial conditions and model physics over mesoscale forecast lengths less than 48 h, both initialization and physics diversity are included in the ensemble (e.g., Stensrud et al. 2000). For this research, several nighttime SBL cases exhibiting differing flow and stability regimes are investigated over 12-h windows (0000 - 1200 UTC).

2.2.1 Initial-Condition Diversity

Initial-condition diversity is given by three WRF initialization strategies. The first strategy, Control (CTRL), is a 12-h free forecast from 0000 – 1200 UTC using GFS initial and lateral boundary conditions. The second strategy, Baseline (BSL), uses a 24-h free forecast initialized from GFS 12 h prior to the nighttime period of interest. The third strategy, Four Dimensional Data Assimilation (FDDA; Stauffer and Seaman 1994), refers to a 12-h nighttime forecast with a 12-h pre-forecast starting from GFS and employing analysis nudging over the 12-km domain and observation nudging over all domains until 0000 UTC when the free forecast begins. Each of these strategies offers different but plausible initial-conditions for each SBL case.

2.2.2 Physics Diversity

Subgrid-scale turbulent motions must be accounted for because turbulent mixing, or the lack of turbulent mixing, directly impacts SBL evolution. Therefore, the choice of subgrid-scale parameterization employed in planetary boundary layer (PBL) schemes within WRF is very important to account for this uncertainty and to create physics diversity among ensemble members. Four PBL physics schemes along with their respective coupled surface layer schemes are incorporated into the ensemble: the Mellor-Yamada-Janjic PBL scheme (MYJ; Janjic 1994) coupled with the Eta surface layer scheme (Janjic 1996), the Mellor-Yamada Nakanishi Niino PBL scheme (MYNN; Nakanishi and Niino 2004) coupled with the Eta surface layer scheme, the Quasi-Normal Scale Elimination PBL scheme (QNSE; Sukoriansky et al. 2005) coupled with the QNSE surface layer scheme (Sukoriansky 2008), and the Yonsei University PBL scheme (YSU; Hong et al. 2006; Hong 2010) coupled with the revised Pennsylvania State University – National Center for Atmospheric Research (MM5) surface layer scheme (Jimenez et al. 2012). When these one-dimensional (1D) PBL schemes are activated within WRF, the model vertical diffusion is accounted for by the PBL scheme. Coupled surface layer schemes provide exchange coefficients allowing for the calculation of heat, moisture, and momentum fluxes by the LSM. These fluxes provide the lower boundary condition for the PBL schemes (Skamarock et al. 2008).

Although the methods by which surface fluxes are calculated by the LSM may differ among parameterizations, LSM diversity is not included in the ensemble configuration. Sensitivity analyses of low-level wind and temperature errors over the Rock Springs, PA network (not shown) reveal little spread between the Noah LSM and thermal diffusion LSM (Dudhia 1996) when averaged over the nighttime period. Moreover, the ensemble down-selection process presented by Lee et al. (2012) reveals that the most representative ensemble members of the true

atmospheric state included the Noah LSM. Therefore, the Noah LSM is held constant for each ensemble member.

The PBL parameterizations offer diversity in turbulence closure methodology. The MYJ is a local 1.5-order closure scheme that includes a prognostic equation for Turbulent Kinetic Energy (TKE). Unknown turbulent quantities at a point are parameterized by known quantities or vertical gradients at that point within the 1D column for local closure schemes (Stull 1988). In the MYJ, PBL height is defined as the height at which TKE drops below a requisite value. In order to avoid excessive mixing within the SBL, background TKE is reduced from 0.1 to 0.01 $m^2 s^{-2}$ (Seaman et al. 2012). The MYJ is coupled with the Monin-Obukhov similarity-theory-based NCEP Eta surface layer scheme that uses the Beljaars correction to eliminate singularities when friction velocity goes to zero (Beljaars 1994).

The MYNN is configured with local 1.5-order closure within the ensemble although it can be configured to use 2nd-order closure within WRF (Nakanishi and Niino 2004). The 1.5-order configuration of the MYNN scheme includes the same prognostic TKE equation as the MYJ but employs a different method to calculate mixing lengths. Additionally, the MYNN accounts for horizontal and vertical TKE advection. In this study, the MYNN is also coupled with the NCEP Eta surface layer scheme.

The QNSE is a TKE-based local 1.5-order closure scheme that calculates eddy viscosities and diffusivities using a spectrally based theory that eliminates perturbation scales from the primitive equations (Sukoriansky et al. 2005). As described by Sun et al. (2015), this method considers spatially anisotropic flow and accounts for the effects of unresolved turbulence and waves simultaneously. As the QNSE scheme is designed to provide a better representation of turbulent physics within the SBL, these considerations become more important with increasing stratification. PBL height determination is similar to that of the MYJ, and the QNSE is coupled with the similarity-based QNSE surface layer scheme.

The YSU is a 1st-order, K-theory-based turbulence closure scheme. Unlike the MYJ, MYNN, and QNSE, the YSU does not include a prognostic equation for TKE. In the presence of a mixed layer, counter-gradient terms in the YSU allow up-gradient fluxes associated with eddies regardless of the background local gradients (Hong et al. 2006). The YSU is often considered a nonlocal scheme because of the inclusion of these counter-gradient terms, but the counter-gradient terms are dropped under stable conditions and the YSU reduces to a local K-theory approach (Hong 2010). PBL height in stable (unstable) conditions is defined as the height at which the bulk Richardson number exceeds 0.25 (0) (Hong 2010). Recent improvements have been made to the YSU including the reduction of excessive mixing allowing for more representative SBL forecasts (e.g., Hu et al. 2013). The YSU PBL scheme is coupled with the revised MM5 similarity-based surface layer scheme.

2.3 SCIPUFF Model Description

A total of 12 ensemble members, summarized in Table 2-1, are given by initial-condition and PBL and surface-layer physics diversity. Ensemble utility is extended to AT&D forecasting by using each member to drive trajectories and the SCIPUFF AT&D model in order to quantify forecast uncertainty for simulated chem-bio releases within the Nittany Valley.

The Second-Order Closure Integrated Puff (SCIPUFF) model is a Lagrangian-puff model used to simulate AT&D (Sykes et al. 2006). The use of SCIPUFF offers several advantages over other AT&D simulation methods. The Lagrangian approach avoids artificial numerical diffusion present in Eulerian schemes, and SCIPUFF simulations involve the release of a collection of 3D Gaussian puffs that may split or merge as they grow in size increasing computational efficiency. In order to better resolve AT&D over a wide range of scales, SCIPUFF employs adaptive horizontal grids with higher resolution grids focused around the highest tracer concentrations. As

in Warner et al. (2002), Lee et al. (2009), and Kolczynski et al. (2009), SCIPUFF is driven by gridded meteorological data. In this case, gridded ensemble member data output by WRF are utilized. By treating the gridded wind field input as an average quantity, SCIPUFF models the statistics of unresolved turbulence via 2nd-order turbulence closure. By calculating the mean and variance of tracer concentration at a given point, SCIPUFF provides the prediction of concentration probability density functions in space and time (Sykes et al. 2006).

A large source of uncertainty in dispersion forecasting is the uncertainty in meteorological data driving AT&D models, particularly in the horizontal wind (Rao 2005). Transport over complex terrain and long distances adds additional dispersion uncertainty (e.g., Deng et al. 2004). In hazard mode, SCIPUFF can account for horizontal-wind uncertainty by augmenting and orienting a plume based on single-point wind variances derived from the WRF ensemble of gridded meteorological data. In hazard mode, larger values of u- and v-wind component ensemble variances, referred to as UUE and VVE, lead to larger puff areas and lower mean concentrations. The ensemble covariance of u- and v-wind components, UVE, alters plume orientation given the sign of the single-point values. For example, negative UVE values result in puffs stretched in a NW to SE direction (e.g., Kolczynski et al. 2009). Additionally, a decorrelation length scale estimate (SLE) for dispersion concentration variance must be provided in hazard mode. As described by Warner et al. (2002), larger values of SLE increase uncertainty in the concentration field because of decreased small-scale mixing and diffusion of the plume and increased variability in plume location.

Two types of SCIPUFF ensemble approaches are considered to quantify AT&D uncertainty as in Warner et al. (2002). The explicit ensemble approach involves aggregating AT&D simulations for each ensemble member and calculating the probability of meeting or exceeding a given concentration threshold at a point. Because running SCIPUFF for each ensemble member may be too costly given a time-sensitive situation, the second approach

involves the use of a single-member SCIPUFF simulation derived from the full ensemble. SCIPUFF hazard mode incorporating UUE, VVE, UVE, and SLE is employed in order to include ensemble spread information (meteorological uncertainty) in single-member AT&D forecasts.

A commonly used single member that can be derived from an ensemble is the ensemble mean. By averaging over each of the 12 WRF members, the ensemble mean is used to drive SCIPUFF transport while hazard mode accounts for the meteorological uncertainty around the mean conditions. Another single member, referred to as a "best member", is defined in order to retain dynamic consistency and reduce the smoothing effects due to averaging within the ensemble mean. The best member can be defined through user preference, and in this case, it is the member whose vector wind difference from the ensemble mean is smallest when averaged over a prescribed atmospheric depth defined below. Because SCIPUFF defines a concentration probability density function with concentration mean and variance at a given point, probabilities of meeting or exceeding a threshold concentration are derived using a clipped-normal distribution for single-member AT&D forecasts (Sykes et al. 2006).

Several considerations are given before running SCIPUFF for a case involving the SBL. Within SCIPUFF, Large Scale Variability (LSV) is typically used for mesoscale and synoptic flows with larger grid length scales (Sykes et al. 2006). LSV is used to represent dispersion over the subgrid model scales. SBL motions are often dominated by submeso flows and smaller mesogamma flows, and finer grid lengths reduce the need for LSV, especially for relatively short integrations. Furthermore, LSV ignores stratification and topographic implications. For these reasons, LSV is deactivated for all SCIPUFF runs. Additionally, the parameter uu_calm, which defines a minimum horizontal velocity fluctuation variance associated with horizontal meandering, is set to 0 m^2s^{-2}. In order to prevent excessive mixing within the SBL, wwtrop, which represents tropospheric vertical turbulent fluctuations, is reduced from 0.01 to 1E-4 m^2s^{-2}. Because of the shallow PBL heights associated with the SBL and the height of the terrain

surrounding the Nittany Valley, the depth over which vector wind differences are compared in order to select a best member is set at 300 m. The SLE, after testing values ranging from 2 km (smallest wavelength gravity waves) to 20 km (valley width), is set to 2 km in order to retain the full impacts of wave-turbulence interactions related to submeso motions and smaller mesogamma scale motions in the cases examined.

The performance of multiple ensemble configurations based on Table 2-1 is evaluated for several case studies. These configurations are detailed in Chapter 3 along with a summary of the statistical assessment metrics used to evaluate individual member and ensemble performance.

Table 2-1. Summary of the 12 WRF ensemble members given by initial conditions (IC) and planetary boundary layer (PBL) with coupled surface layer (SL) physics diversity.

Member	IC	PBL	SL
1	CTRL	MYJ	Eta
2	BSL	MYJ	Eta
3	FDDA	MYJ	Eta
4	CTRL	MYNN	Eta
5	BSL	MYNN	Eta
6	FDDA	MYNN	Eta
7	CTRL	QNSE	QNSE
8	BSL	QNSE	QNSE
9	FDDA	QNSE	QNSE
10	CTRL	YSU	Rev. MM5
11	BSL	YSU	Rev. MM5
12	FDDA	YSU	Rev. MM5

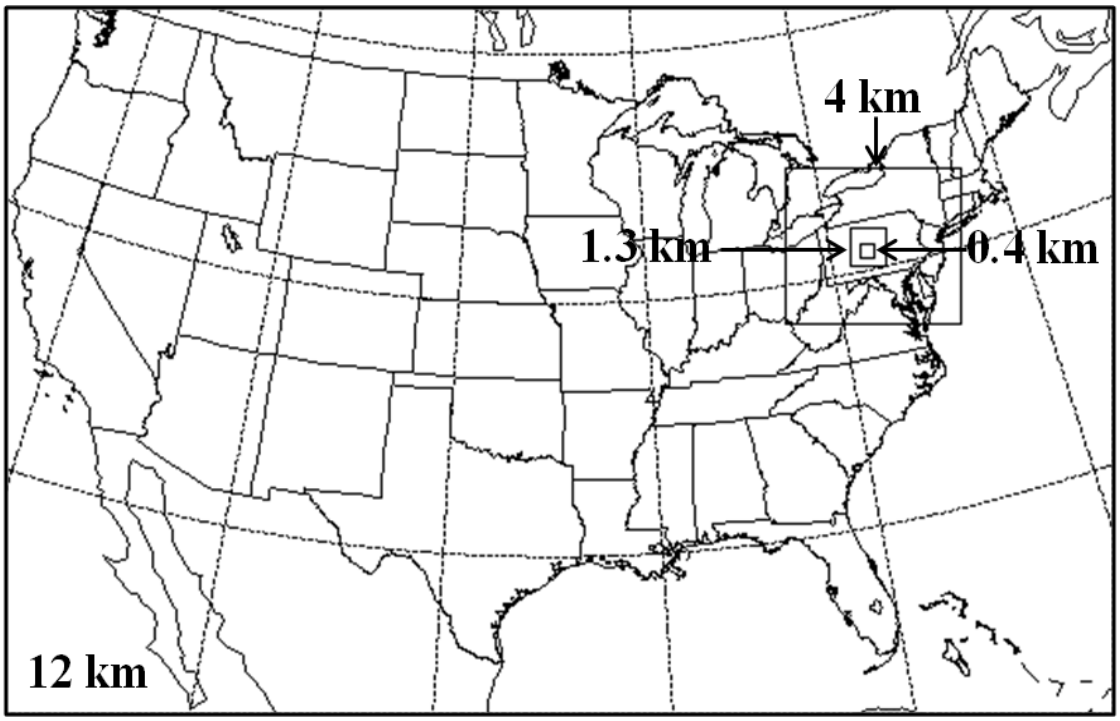

Figure 2-1. Locations of the 12-km, 4-km, 1.3-km, and 0.4-km one-way nested domains within the SBL WRF configuration.

CHAPTER 3

EXPERIMENTAL DESIGN AND EVALUATION METHODS

3.1 Experimental Design

Capturing the future state by generating a representative sample of possible future states of the atmosphere is the end goal of ensemble modeling. In order to evaluate full ensemble performance as well as the contributions of initialization and physics diversity, three ensemble configuration types are tested for several case studies. The first configuration type, the full ensemble, includes both initial-condition and physics diversity and is comprised of the 12 members summarized in Table 2-1. The second configuration type, referred to as multi-initialization, includes initial-condition diversity while holding each physics member constant. This configuration type allows testing of four three-member multi-initialization ensembles (e.g., members 1-3 in Table 2-1 form the three-member MYJ multi-initialization ensemble). The final configuration type, referred to as multi-physics, includes PBL/SL physics diversity while holding the initial-conditions constant. This permits testing of three four-member ensembles (e.g., members 3, 6, 9, and 12 in Table 2-1 form the four-member FDDA multi-physics ensemble). Examining the performance of each of these ensemble configurations allows a better understanding of the contributions of diversity within the full ensemble and provides insight into the degree to which the ensemble captures the future atmospheric state.

Because of submeso motions and small mesogamma phenomena that occur within the SBL, a 0.4-km horizontal grid spacing nest is included in the WRF model configuration (section 2.1). However, many operational centers consider less than about 4-km horizontal grid spacing to be relatively high resolution (see Chapter 1). Moreover, Grimit and Mass (2002) suggest

diminishing ensemble performance improvement for low-level wind direction forecasts as grid spacing is decreased from 12 km to 4 km for a study over the Pacific Northwest. Because the WRF configuration employed in this study includes 12-km, 4-km, 1.3-km, and 0.4-km grid spacing nests, it is important to investigate whether running a 12-member ensemble with 0.4-km grid spacing is worth the additional computational cost. In order to test the utility of enhanced horizontal resolution appropriate for modeling over the Nittany Valley and Rock Springs network (defined in section 4.1), both 0.4-km and 1.3-km horizontal grid spacing configurations of the full ensemble are evaluated and compared. These 12-member configurations are referred to as the 0.4-km ensemble and 1.3-km ensemble respectively.

3.2 Evaluation Methods

Individual member performance and ensemble performance are evaluated using deterministic and probabilistic verification techniques respectively. Ideally, the verified future atmospheric state should fall within the spread of the ensemble, and the amount of spread should be correlated to ensemble-mean forecast error. This spread-skill relationship yields information about the dispersive behavior of an ensemble that can be evaluated with several metrics. Here, ensemble performance for the various configuration types is evaluated using rank histograms (Hamill 2001, Wilks 2011a), the Continuous Ranked Probability Score (CRPS; Hersbach 2000; Wilks 2011b), and Reduced Centered Random Variable techniques (RCRV; e.g., Candille et al. 2010) while individual member performance is evaluated using Mean Error (ME) and Mean Absolute Error (MAE; Wilks 2011b). A bootstrap technique is employed to test for significant differences between the 0.4-km and 1.3-km ensemble configurations (Efron 1979; DiCiccio and Efron 1996; Candille et al. 2007).

3.2.1 Mean Error and Mean Absolute Error

Mean Error (ME) and Mean Absolute Error (MAE) measure deterministic forecast bias and accuracy respectively (Wilks 2011b). ME is the average difference between forecast and observed values and is given by

$$ME = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i) \tag{3.1}$$

where N is the number of forecasts with corresponding observations, *f* is the forecast value, and *o* is the corresponding observed value. A zero value is best, but because cancellation of error may occur due to averaging, ME must be interpreted carefully especially when used to evaluate variables that can take on both positive and negative values and when MAE is much greater than zero.

The MAE gives the average magnitude of deterministic forecast error and is given by

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|f_i - o_i| \tag{3.2}$$

with variables as in ME. The MAE is negatively oriented (i.e., a lower MAE indicates a better forecast). As opposed to other forecast verification metrics such as root mean square error (RMSE), MAE is advantageous as it is unambiguous and gives a clear representation of average or typical forecast error (Willmott and Matsuura 2005).

3.2.2 Rank Histograms

Rank histograms provide an assessment of the reliability and spread of an ensemble

prediction system. A reliable ensemble forecast probability is equal to the true probability of the event of interest. If an ensemble forecast is reliable, the observation is equally likely to fall into any position within the grouped and sorted ensemble and observation values where "rank" refers to the position of the observation. If reliable, a uniform rank histogram results when all ranks for forecasts of a variable over some spatial and/or temporal interval are combined (Hamill 2001). Deviations from rank uniformity can depict ensemble bias, underdispersive, or overdispersive behavior. Visually, an ensemble bias appears as a slope to one side of the rank histogram. Underdispersive behavior occurs when the observation is too often an outlier compared to the ensemble forecast and appears as a U-shaped rank histogram. An overdispersive ensemble depicts larger counts or frequencies toward the center of a rank histogram. This behavior is caused by verification values that are infrequently extreme in comparison to ensemble forecast values. Rank histograms are unable to quantify the ability of an ensemble to produce a specific forecast and should be used in conjunction with a technique that measures resolution, or sharpness, of a forecast (Hamill 2001).

3.2.3 Continuous Ranked Probability Score

The Continuous Ranked Probability Score (CRPS) is a strictly proper probabilistic score that quantifies the squared difference between the ensemble forecast cumulative distribution function (CDF) and the CDF of the corresponding observation (e.g., Gneiting and Raftery 2007). The CRPS is given by:

$$CRPS = \int_{-\infty}^{\infty} [P(x) - P_o(x)]^2 dx \tag{3.3}$$

where x is the variable of interest, P is the ensemble forecast CDF, and P_o is the observation CDF. Because it is negatively oriented, a perfect CRPS of zero is achieved when all ensemble forecast values are equal to the corresponding observed value. As a limited number of ensemble member forecasts exist in practice, ensemble CDFs are constructed as cumulative density functions often under the assumption that each ensemble member forecast is equally likely, and the CRPS is calculated over a discrete interval. The CRPS is then averaged over all forecast and observation pairs for a given spatial and/or temporal interval. Conveniently, the CRPS retains the units of the variable of interest and reduces to MAE for a single deterministic forecast. Therefore, the CRPS allows for the direct comparison of deterministic and probabilistic forecasts.

As demonstrated by Hersbach (2000), the CRPS can be decomposed into three components so that

$$CRPS = Reliability - Resolution + Uncertainty \tag{3.4}$$

Reliability refers to the position of the observed value within the ensemble of sorted forecast values; a perfect reliability value is zero with reliability decreasing with increasing value. The value of reliability is closely related to the shape of a corresponding rank histogram. Resolution, often referred to as the sharpness of a forecast, refers to the ability of the ensemble to produce a specific, unambiguous event forecast, and larger values are better. The value of uncertainty is proportional to the standard deviation of the observations. The negatively oriented potential CRPS given by

$$CRPS_{Pot} = Uncertainty - Resolution \tag{3.5}$$

combines the resolution and uncertainty terms into a single term that represents the best possible

ensemble performance given observational uncertainty (Hersbach 2000). Ideally, the ensemble forecast system strikes an optimal balance between reliability and resolution (e.g., an ensemble with a high resolution value and a poor (high) reliability value refers to a specific forecast for the wrong value). Because the CRPS provides an assessment of global ensemble skill and includes reliability, resolution, and uncertainty information, it is a highly desirable metric for ensemble evaluation.

3.2.4 Reduced Centered Random Variable Bias & Dispersion

The reliability property of an ensemble can be further explored using the first two moments, bias and dispersion, of the reduced centered random variable (RCRV; e.g., Candille et al. 2010). The RCRV is calculated for each ensemble forecast realization and is given by

$$RCRV = \frac{m - o}{\sqrt{\sigma_o^2 + \sigma^2}} \tag{3.6}$$

where m is the ensemble mean, o is the observed value, σ is the standard deviation of the corresponding ensemble prediction, and σ_o is the standard error of the observations over all realizations. The first moment of the RCRV, bias, is simply the mean of the RCRV values over a set of realizations and gives the forecast error of the ensemble mean normalized by ensemble spread and observational error. The numerator of the RCRV is reversed here from that of Candille et al. (2010) so that the interpretation of bias is more consistent with that of ME. Dispersion, the second moment of the RCRV, is given by the standard deviation of the RCRV values and measures the spread of the forecast error of the ensemble mean over a set of realizations. For a perfectly reliable ensemble forecast system, bias = 0 and dispersion = 1. If bias < 0, the ensemble forecast values are systematically less than observed values, and if bias > 0, the ensemble forecast

values are systematically greater than observed values. An overdispersive ensemble is indicated if dispersion < 1, and an underdispersive ensemble is indicated if dispersion > 1. Values of bias and dispersion correspond to the shape of rank histograms and can be used to diagnose the cause of reliability shortcomings indicated by the reliability component of the CRPS.

3.2.5 Bootstrapping and Confidence Intervals

Bootstrapping is a technique used to create approximations to the sampling distributions of sample statistics of interest (Wilks 2011b). Typically, bootstrapping is employed in order to ascertain statistical significance when a nonparametric test is not available. In bootstrapping, data are resampled with replacement, and the statistic of interest for the new sample is calculated. This process is generally repeated thousands of times to create a distribution of sample statistics known as a bootstrap distribution (Efron 1979). Under the assumption of data independence, approximate confidence intervals for the sample statistic can be set from percentiles of the bootstrap distribution. The unknown population statistic of interest exists within these confidence intervals, with a level of confidence defined by the percentiles. This is known as the percentile method. The bias corrected and accelerated (BCa) modification to the percentile method adjusts percentiles of interest based on the bias and skewness of the bootstrap distribution (DiCiccio and Efron 1996).

Here, the BCa technique is used to identify significant differences in the CRPS and its components between two ensemble configurations. The methodology behind this approach is based on Candille et al. (2007). Data are resampled 10,000 times to create 10,000 bootstrap samples. The CRPS is calculated for each of the 10,000 new samples for two separate ensembles. CRPS values from ensemble B are subtracted from those of ensemble A where a positive value indicates an improvement by ensemble B. A distribution of CRPS differences is created, and

confidence intervals are derived by calculating the 2.5th and 97.5th percentile to give a 95% confidence interval. If the CRPS difference and confidence intervals are both above zero, the result is interpreted as a statistically significant improvement because only improvement occurred within the confidence interval. If both confidence intervals fall below zero, the result is a statistically significant degradation from ensemble A to B. If the confidence intervals straddle zero, both improvement and degradation are noted within the confidence interval and there is no significant difference in ensemble performance. The method is repeated for both reliability and potential CRPS. Because a direct comparison of two ensemble configurations is based on the same observation data, uncertainty is equal in both ensembles and any improvement in potential CRPS is attributed to an improvement in resolution (e.g., Candille et al. 2007). This technique is used to quantify significant differences in performance between the 0.4-km and 1.3-km ensemble configurations described in section 3.1 and Table 2-1.

CHAPTER 4

ROCK SPRINGS OBSERVATION NETWORK AND CASE STUDY DESCRIPTIONS

4.1 The Nittany Valley and the Rock Springs Observation Network

Located in central PA, the Nittany Valley is approximately 20 km wide and is bordered by the Allegheny Mts. (~350 m valley to ridge-top height) to the northwest and Tussey Ridge (~300 m valley to ridge-top height) to the southeast. The long axis of the valley extends from roughly SW to NE (Fig. 4-1a). The surrounding ridgelines and slopes are forested while the valley itself is predominantly covered by croplands and grasslands. A special observing network located in Rock Springs, PA is located along the base of Tussey Ridge (Fig. 4-1b), and data gathered by the network are used for ensemble performance evaluation for several SBL case studies that involve flow interaction with complex terrain.

The Rock Springs network contains multiple instrumented towers with heights ranging from 2 - 50 m AGL, and sonic detection and ranging (SODAR) instrumentation at two locations. Rock Springs observation sites are distributed at various locations at different elevations along the slope of Tussey Ridge and within the valley in order to best sample mesogamma and submeso phenomena related to flow interactions with complex terrain in the SBL. Tracer measurements are unavailable within the Rock Springs network, and direct verification of AT&D ensemble forecasts is not possible with available data. Because low-level horizontal winds and PBL depth are two of the most important parameters impacting AT&D forecast uncertainty (Lewellen and Sykes 1989; Rao 2005), evaluation of ensemble low-level wind and temperature forecasts are used as proxy for direct AT&D forecast evaluation.

Low-level temperature and wind data are gathered from sites 3, 6, 7, 8, 9, and 12 within the Rock Springs network (Fig. 4-1b). Wind measurements are given by 2-dimensional (2D) and 3-dimensional (3D) sonic anemometers. Vaisala WS425 2D sonic anemometers provide 1-min-averaged horizontal wind speed and direction with accuracies of ± 0.1 ms^{-1} and $\pm 2^{o}$ respectively, and Campbell Scientific CSAT3 3D sonic anemometers provide 20-Hz *u*, *v*, and *w* wind component measurements with accuracies of ± 0.08 ms^{-1}, ± 0.08 ms^{-1}, and ± 0.04 ms^{-1} respectively. Campbell Scientific T107 thermistors provide 1-min temperature measurements with $\pm 0.4^{o}$C accuracy. Table 4-1 provides a summary of site locations and 2-m AGL Rock Springs instrumentation availability.

The Atmospheric Science Corporation WindExplorer 4500-Hz SODARs available within the Rock Springs network provide vertical profiles of 10-min-averaged *u*, *v*, and *w* (accuracy of ± 0.5 ms^{-1}) wind components and atmospheric stability information. The 3D wind observations range in height from 30 - 250 m AGL with 5-m vertical resolution. In this study, SODAR observations are used to identify the possible presence of wave motions and rotor circulations over the Rock Springs network. A SODAR is located on the lower slope of Tussey Ridge prior to 29 September 2011 and is located on the valley floor thereafter (Fig 4-1b). A more extensive discussion of Rock Springs observation network measurement capabilities can be found in Hoover et al. (2015).

For each of the case studies described in section 4.2, 2-m temperature and wind data gathered at the six Rock Springs sites are used for ensemble performance evaluation. The ensemble is evaluated over 12-h nighttime periods (0000 - 1200 UTC) during which an SBL and submeso and smaller mesogamma scale motions are present. Wind variables of interest include wind speed and wind components derived from 2D and 3D sonic anemometer data. In order to best describe the phenomena occurring over the Nittany Valley and Rock Springs network, wind components are rotated to an across-valley (perpendicular to Tussey Ridge) and along-valley

(parallel to Tussey Ridge) Cartesian coordinate system. Positive across-valley wind is directed from approximately NW to SE, and positive along-valley wind is directed from approximately SW to NE. The temperature and wind data are averaged using 12-min centered averaging giving a maximum of 360 nighttime observations per case for 2-m temperature, wind speed, across-valley, and along-valley wind over the six sites. These data are directly compared against 2-m temperature and wind data output every 12 min by the various configurations of the WRF ensemble using the metrics described in section 3.2.

4.2 Case Study Descriptions

SBL cases of interest are chosen based on three distinct flow regimes in relation to the topography of the Nittany Valley: flow parallel to the valley, flow across the Allegheny Mts., and flow across Tussey Ridge. These regimes influence the presence and amplitude of submeso and mesogamma wave motions that lead to AT&D forecast uncertainty within the SBL (Hoover et al. 2015). The cases include 13 September 2011 (SEP13) characterized by valley-parallel flow, 16 September 2011 (SEP16) characterized by NW flow from the Allegheny Mts., and 6 November 2011 (NOV06) characterized by S-SE flow across Tussey Ridge. Figure 4-2 depicts mean sea level pressure, surface temperature, surface wind barbs, and total cloud fraction at 0000 UTC and 1200 UTC as indicated by NCEP North American Regional Reanalysis (NARR; Mesinger et al. 2004) data for each of the three cases examined.

4.2.1 Valley-Parallel Flow: SEP13

SEP13 features weak SW surface flow roughly parallel to Tussey Ridge from 0000 - 1200 UTC (Fig. 4-2a, b). Westerly winds at 850 hPa (not shown) persist throughout the night as a

high-pressure ridge extends toward PA from the south and a low-pressure system crosses eastern Canada. Rock Springs observations reveal 2-m wind speeds less than 0.5 ms^{-1} throughout most of the nighttime period (not shown). Although NARR total cloud fraction indicates enhanced cloud cover extending over central PA at 0000 UTC (Fig. 4-2a), it is inconsistent with surface observations at state-wide WMO sites that reveal mostly clear skies across the majority of PA including the Nittany Valley (not shown). Light winds in conjunction with clear sky conditions lead to strong thermal stratification near the surface and stable conditions. Because SEP13 lacks strong terrain-crossing synoptic flow over the Nittany Valley, it represents an SBL case with AT&D dominated by submeso motions such as drainage flows.

4.2.2 Cross-Mountain NW Flow: SEP16

SEP16 exhibits NW flow from the Allegheny Mts. over central PA and the Nittany Valley. Pressure gradients over central PA weaken overnight as an area of high pressure approaches from the west (Fig. 4-2c, d) while 10-15 ms^{-1} NW winds remain present at 850 hPa (not shown). Figures 4-2c and 4-2d reveal clear skies throughout the night with surface wind speeds over the Rock Springs network generally less than 1.5 ms^{-1} (not shown). These conditions allow the development of the SBL although thermal stratification is weaker than that of SEP13. As evidenced by rotor circulations observed by SODAR observations over the Rock Springs network (Fig. 4-3), atmospheric conditions support the excitation of trapped-lee waves by flow over the Allegheny Mts. The impacts of these circulations on AT&D are discussed in section 5.2.

4.2.3 Cross-Mountain S-SE Flow: NOV06

NOV06 demonstrates S-SE flow from Tussey Ridge over central PA and the Nittany Valley. Southeasterly flow is related to a zone of high pressure centered east of Rock Springs at 0000 UTC (Fig. 4-2e). By 1200 UTC, the high-pressure center moves over the Atlantic Ocean and surface winds become more southerly (Fig. 4-2f). At 850 hPa, 5-10 ms^{-1} winds shift from southerly to southwesterly overnight (not shown). Total cloud fraction indicates clear skies overnight and weak (~1ms^{-1}), omnidirectional surface flow is present over the Rock Springs network leading to strong thermal stratification at the surface. As evidenced by shallow rotor circulations observed by the SODAR, atmospheric conditions support the excitation of trapped-lee waves by flow interaction with Tussey Ridge (Fig. 4-4). Because wind data are unavailable at Site 8 for NOV06, there are 300 wind observations over the other five sites to be used for the ensemble performance evaluation.

Table 4-1. Rock Springs observation sites 2-m instrumentation summary including location, elevation, instrument availability (denoted by "X"), and instrument temporal resolution.

Station	**Lat. /Long. (deg)**	**Elevation (m above sea level)**	**2D Sonic** (1-min)	**3D Sonic** (20-Hz)	**Thermistor** (1-min)
Site 3	40.71035N/ 77.95762W	368	X	_	X
Site 6	40.70042N/ 77.95728W	416	X	_	X
Site 7	40.70296N/ 77.95925W	385	_	X	X
Site 8	40.71816N/ 77.94582W	364	_	X	X
Site 9	40.70848N/ 77.95892W	368	_	X	X
Site 12	40.70371N/ 77.96684W	362	X	_	X

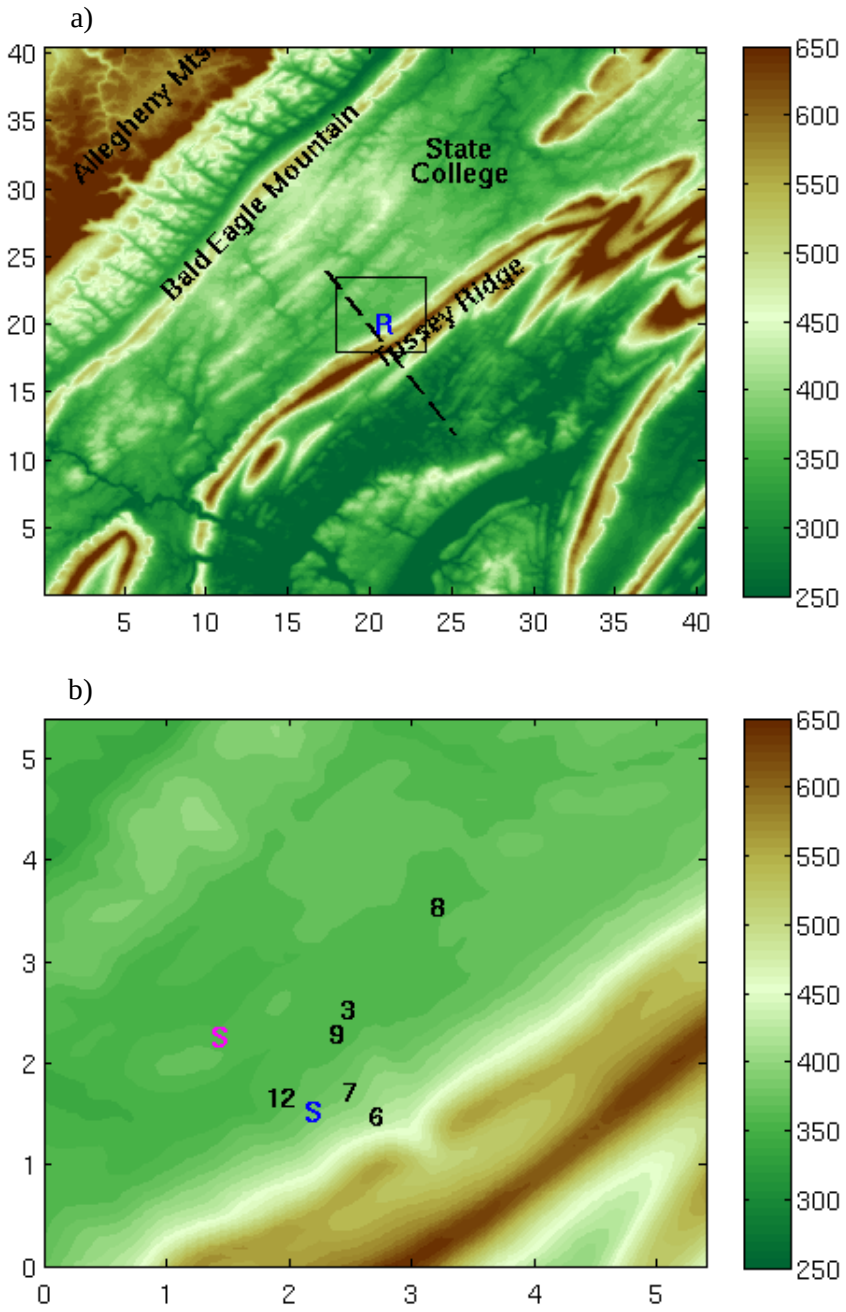

Figure 4-1. 90-m resolution terrain (MSL) for a) 40 km by 40 km region surrounding the Nittany Valley including major topographical features and b) a 5 km by 5 km region denoted by the black square in a) showing the distribution of instrumented sites used for this study within the Rock Spring observation network. The blue "S" denotes the location of the SODAR prior to 29 September 2011 and the magenta "S" denotes the location of the SODAR thereafter. The dashed line in a) corresponds to the location of trajectory cross sections discussed in Chapter 5.

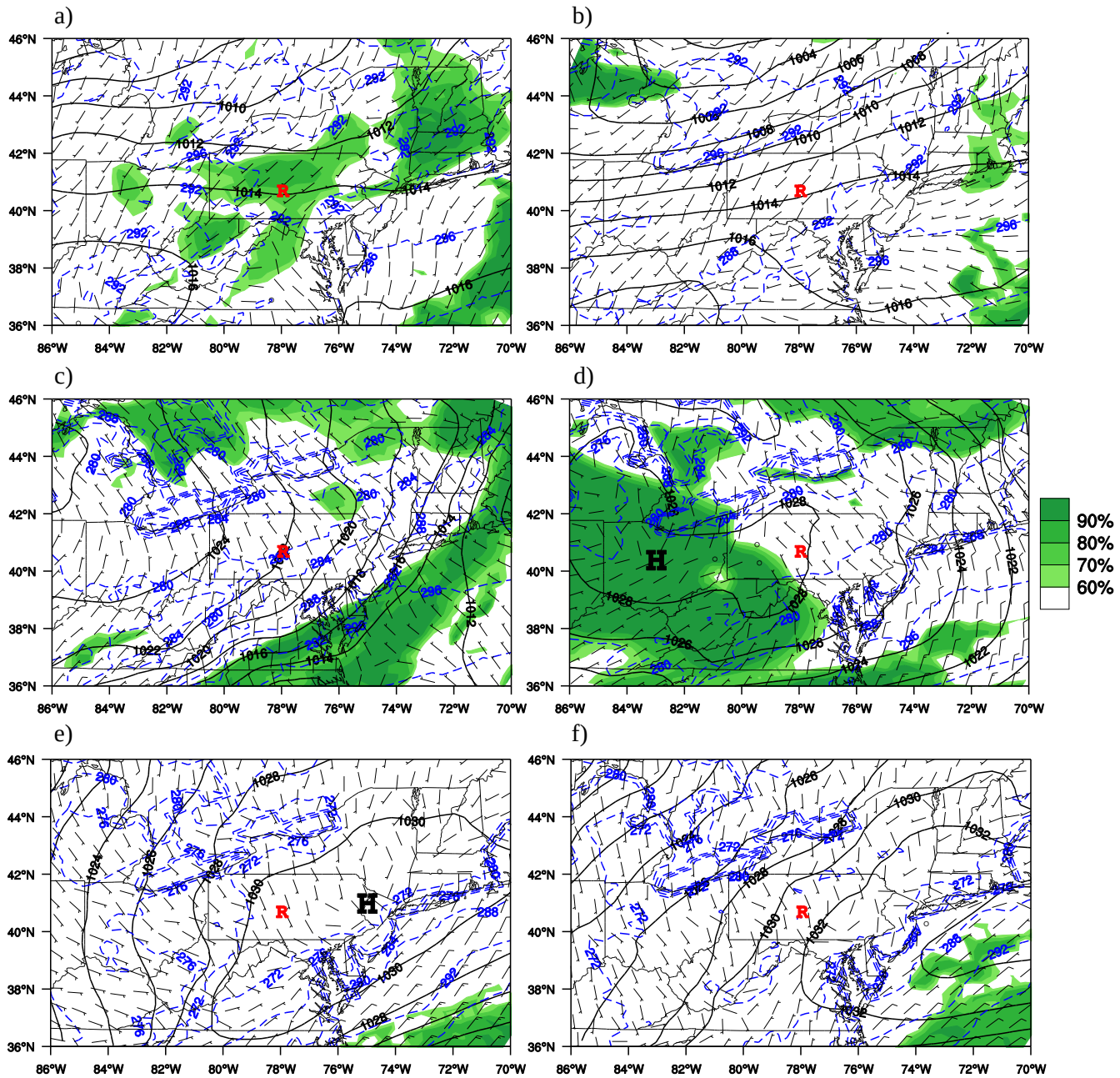

Figure 4-2.Surface wind barbs (ms^{-1}), surface temperature (K), MSL pressure (hPa), and total cloud fraction given by North American Regional Reanalysis (NARR) data at 0000 UTC (left column) and 1200 UTC (right column) for SEP13 (a, b), SEP16 (c, d), and NOV06 (e, f).

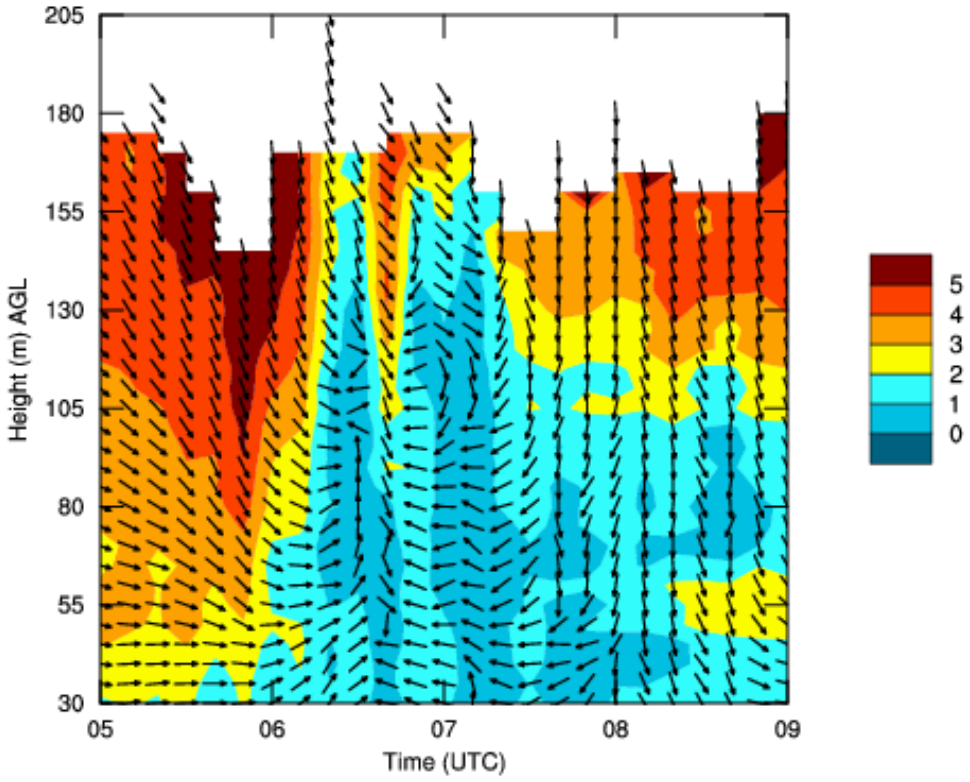

Figure 4-3. SEP16 SODAR observed wind speed (ms^{-1}, color-filled contours) and horizontal wind direction (plan-view wind vectors) with height (AGL) from 0500 – 0900 UTC.

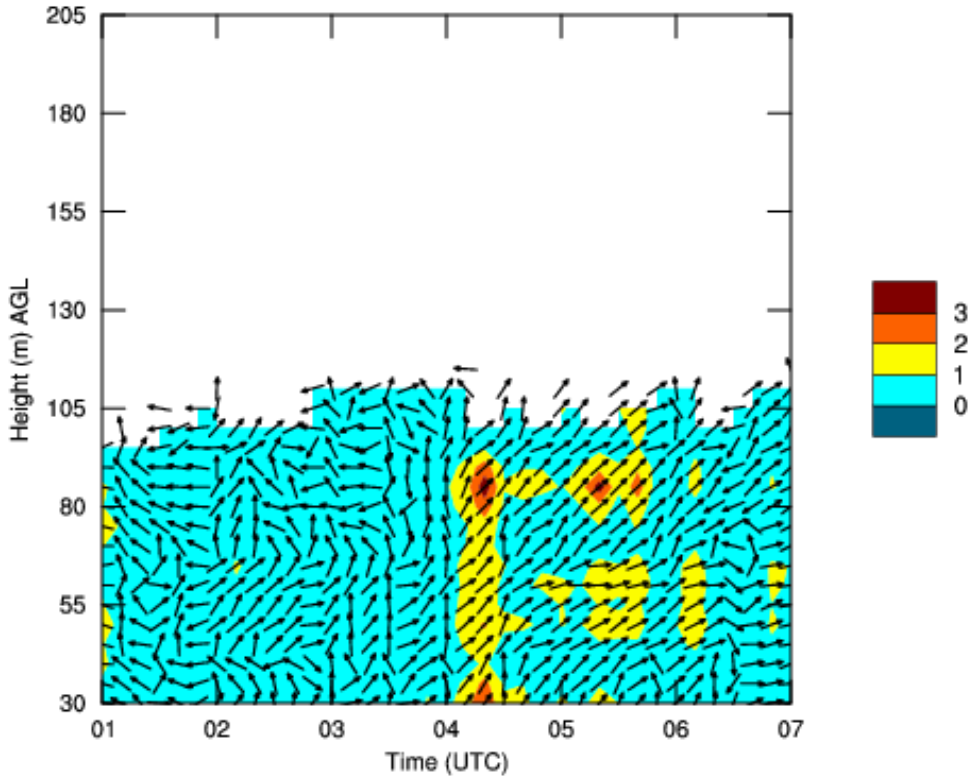

Figure 4-4. NOV06 SODAR observed wind speed (ms^{-1}, color-filled contours) and horizontal wind direction (plan-view wind vectors) with height (AGL) from 0100 – 0700 UTC.

CHAPTER 5

RESULTS: ENSEMBLE EVALUATION AGAINST CASE STUDY OBSERVATIONS

The three case studies are performed to assess ensemble performance at two grid resolutions under differing flow regimes impacting the SBL over the Nittany Valley. Each case study focuses on a simulated near-surface release from Site 9 in the Rock Springs observation network (section 4.1). For each member of the 0.4-km ensemble, particle trajectories that simulate the movement of air parcels are initialized from 3 m AGL at nine grid cells surrounding Site 9. Then, corresponding SCIPUFF surface dosage plumes are analyzed following 3-m, 12-min continuous releases of the passive tracer C_7F_{14}. Single-member surface dosage threshold probabilities are then compared against threshold probabilities derived from the explicit ensemble.

Because tracer data are unavailable over the Rock Springs network, the surface variable ensemble forecasts are evaluated with WRF ensemble output for various ensemble configurations using the metrics described in section 3.2. Observational error, which includes both instrumental and representativeness error, may impact ensemble evaluation results (e.g., Hamill 2001). Instrumental error can be accounted for by adding random noise to ensemble forecast values given the accuracies specified by measurement instrumentation. Representativeness error can be accounted for by essentially determining the error associated with interpolating model forecast values to specific points and adding this error randomly to ensemble forecast values as with instrumental error. For this study, instrumental error (see accuracies in section 4.1) is similar in magnitude to representativeness error for both 1.3-km and 0.4-km ensemble configurations, and ensemble evaluation results demonstrate very weak dependence on the inclusion of observational

error. Therefore, all ensemble evaluation results are presented without the inclusion of observational error.

Bootstrap resampling (section 3.2.5) is used to ascertain whether statistically significant differences exist between the forecast performance of the 1.3-km and 0.4-km ensemble configurations. Under the assumption of data independence, most performance differences between the two ensemble configurations are shown to be significant. However, meteorological data are not necessarily independent due to spatial and temporal correlations, although these correlations may be minimized for wind variables in cases of weak, omnidirectional flow. One way to alleviate the data independence problem is to employ block bootstrapping where geographic or temporal blocks of data are resampled instead of individual data vectors (Wilks 2011b). However, Hamill (1999) demonstrates that significant spatial correlations still exist between blocks for precipitation forecasts even when a relatively large domain is divided into coarse partitions. Moreover, Romine et al. (2014) note nearly unchanged results when employing block bootstrapping as compared to assuming data independence when verifying precipitation and radiosonde wind and temperature forecasts. Because observational data are gathered here for single 12-h forecast periods over a small geographic region, block bootstrapping is not expected to yield meaningful results. Therefore, as in Romine et al. (2014) and Lee et al. (2012), all statistical significance results are presented under the assumption of data independence.

5.1 Valley-Parallel Flow: SEP13

5.1.1 SEP13 AT&D Results

Because of weak winds parallel to Tussey Ridge and clear skies over central PA leading to radiational cooling and strong surface thermal stratification, AT&D outcomes for SEP13 are

dominated by drainage flows. Figure 5-1 depicts horizontal trajectories after a 2-h integration (0600 – 0800 UTC) for each of the 0.4-km ensemble members along with 3-m horizontal wind vectors at 0800 UTC. In general, the parcel trajectories converge at the base of Tussey Ridge before moving NE within the valley. Horizontal wind vectors suggest drainage-flow convergence as the cause of this trajectory behavior as flow from the slopes of Tussey Ridge meets weak flow of the opposite direction draining from a small region of enhanced topography within the valley. Horizontal trajectories are projected onto a vertical cross section oriented perpendicular to Tussey Ridge (depicted by the dashed line in Fig. 4-1). The trajectory vertical cross sections are plotted in Fig. 5-2 along with potential temperature contours at 0800 UTC. Trajectory vertical cross sections reveal little or no vertical extent for the trajectories as they are caught in a strong surface cold pool within the valley as indicated by the tightly stacked potential temperature contours. A slight enhancement in trajectory height can be noted in the YSU-BSL and YSU-FDDA members. For these members, parcels are lofted just above the region of strongest thermal stratification allowing them to be advected further down the valley so that trajectories extend more NE than those of the other members. Horizontal wind vector convergence along with enhanced upward vertical motion suggest that this lofting is related to stronger drainage flow convergence at low levels as compared to other ensemble members (Fig. 5-1).

Over the same time window, SCIPUFF surface dosage plumes are shown in Fig. 5-3 corresponding to the particle trajectories in Fig. 5-1. The highest surface dosages exist along Tussey Ridge in the region of drainage flow convergence. Because of the simulation of diffusion in SCIPUFF, the spatial extent of the dosage plumes is enhanced when compared to the extent of horizontal trajectory transport. Strong agreement among ensemble members causes high explicit ensemble surface dosage probabilities of meeting or exceeding a concentration threshold value of 10^{-9} m^3-s m^{-3} along Tussey Ridge and downwind (Fig. 5-4). This threshold value is arbitrary and

would change based on exposure risk tolerances for a specific substance in the event of a chem-bio release.

Reasonable agreement exists between explicit ensemble probabilities and those derived from a SCIPUFF hazard mode simulation using a single-member approach. Figure 5-5 presents surface dosage probabilities for the ensemble mean and ensemble best member (MYJ-CTRL in this case) approaches. Recall that the best member is defined by the member with low-level winds closest to that of the ensemble mean. The probability swaths given by both the ensemble mean and best member approaches are quite similar. Because of plume augmentation and orientation impacts of the wind field uncertainty statistics employed in SCIPUFF hazard mode (section 2.3), the single-member probabilities do not match the fine detail of the explicit ensemble and demonstrate a somewhat spatially larger and probabilistically more diffuse solution. Moreover, the enhanced probabilities to the south of Tussey Ridge in both single-member approaches are likely due to somewhat artificial plume dispersion results linked to the use of SCIPUFF hazard mode. Fine details aside, the single members provide a clear indication of the overall range of potential AT&D outcomes from the explicit ensemble and the spatial area most likely to be impacted in the event of a chem-bio release and at a much reduced computational cost.

5.1.2 SEP13 Ensemble Performance Evaluation

Strong agreement among trajectories and high probability surface dosage probabilities may seem to indicate a high-confidence AT&D forecast; however, forecast outcomes must be compared against observed values to gain insight into the true ensemble forecast accuracy. Lacking tracer observations, we examine Fig. 5-6 presenting nighttime CRPS values for 2-m temperature, wind speed, across-valley wind, and along-valley wind. The CRPS calculation is aggregated over all six Rock Springs sites for the 1.3-km and 0.4-km ensembles. For each

variable, CRPS values are smaller for the 0.4-km ensemble compared to the 1.3-km ensemble indicating that higher grid spacing resolution produces more accurate forecasts. Table 5-1 contains CRPS, reliability, and potential CRPS for each variable for SEP13 along with percentage improvement and statistical significance derived using the bootstrap methodology described in section 3.2.5. Results show a statistically significant improvement in CRPS and reliability for each variable. These improvements are all mainly due to large bias corrections noted in reduced RCRV (section 3.2.4) bias values (Table 5-2). A statistically significant degradation in potential CRPS is present for each wind variable indicating a reduction in resolution or sharpness from the 1.3-km to the 0.4-km ensemble; however, as evidenced by reliability values (Table 5-1), the 1.3-km ensemble produces a higher-resolution or "sharp" forecast for a more incorrect value.

Wind components demonstrate the most notable statistical improvements from the 1.3-km to 0.4-km ensemble. Along-valley wind, the most relevant variable for downwind AT&D in this case, demonstrates a large improvement in RCRV bias and dispersion (see section 3.2.4), and the across-valley wind forecast undergoes a similar improvement. Moreover, large MAE reductions in along-valley wind exist for each explicit ensemble member (Fig. 5-7). The bias reduction for along-valley wind is indicated in the corresponding rank histograms in Fig. 5-8. The large frequency value in rank “1” indicates that for over 90% of the ensemble forecasts throughout the night, every 1.3-km ensemble forecast value is greater than observed. Although still present, this bias is reduced by over 40% in the 0.4-km ensemble forecast. It is this bias reduction that allows the CRPS reliability component to approach zero and the value of RCRV bias to tend closer to zero. Additionally, RCRV dispersion shows improvement in underdispersive behavior as values tend closer to unity (Table 5-2).

In order to obtain an overall sense of initialization- and physics-diversity contributions to the full ensemble, CRPS results for the multi-initialization and multi-physics ensembles are presented in Figs. 5-9 and 5-10 respectively. The QNSE multi-initialization ensemble appears to

produce the best temperature forecast by a relatively large margin while it performs similarly to the MYNN and MYJ configurations for wind speed variables. The YSU multi-initialization ensemble produces the worst forecast for each variable. The multi-physics ensemble CRPS values (Fig. 5-10) show that no one configuration consistently produces the best wind variable forecasts while the CTRL multi-physics configuration produces a superior temperature forecast. This may indicate that low-level ensemble wind forecasts are more dependent on PBL/SL physics than initialization strategy with available observations for this case.

Results of a surface-layer scheme sensitivity analysis (not shown) reveal that poor YSU multi-initialization ensemble performance does not appear to be related to the coupling of the YSU with the revised MM5 surface layer scheme (Jimenez et al. 2012). Because the YSU can only be coupled with the MM5 or revised MM5 surface layer schemes, this finding is deduced by comparing multi-initialization CRPS values for the MYNN coupled with the Eta surface layer scheme and the MYNN coupled with the revised MM5 surface layer scheme. Additional slight degradations in probabilistic skill demonstrated by coupling the MYNN with the revised MM5 surface layer scheme show that deficiencies introduced by the use of this scheme cannot account for the comparatively poor probabilistic skill of the YSU multi-initialization ensemble. Therefore, although recent revisions to the YSU scheme have been shown to improve nighttime boundary layer forecasts (Hu et al. 2013), for SEP13 it seems that TKE-based local turbulence closure schemes (MYJ, MYNN, QNSE) still have some advantage over the non-TKE YSU scheme. This finding is consistent with previous studies that have shown, in general, that local TKE-based PBL schemes better represent the nighttime boundary layer than non-TKE schemes (e.g., Shin and Hong 2011).

Additionally, it is worth expounding upon the finding that the QNSE multi-initialization ensemble performed best for temperature for SEP13. Ensemble results as a whole reveal a strong warm forecast bias in this case (see RCRV bias in Table 5-2) although the QNSE members

produce a reduced warm bias compared to the other members as evidenced by ME (not shown). In both the SEP16 and NOV06 cases discussed in sections 5.2 and 5.3, the full ensemble produces a cold temperature bias compared to observations and the QNSE members produce a stronger cold bias than the other members, consistent with QNSE performance findings from previous studies (e.g., Cintineo et al. 2014). It seems that for SEP13, the cold bias typically produced by the QNSE works to balance a warm temperature bias possibly related to systematic initialization errors. Therefore, the superior temperature forecasts produced by QNSE members for SEP13 may be more lucky than skillful.

5.2 Cross-Mountain NW Flow: SEP16

5.2.1 SEP16 AT&D Results

Trapped-lee waves excited by NW flow from the Allegheny Mts. and perpendicular to the Nittany Valley are the main source of AT&D uncertainty for SEP16. Figure 4-3 presents SODAR observations of horizontal wind from 0500 - 0900 UTC near the base of Tussey Ridge (see Fig. 4-1 for location). After 0600 UTC, weakening winds at low-levels and spreading upwards accompany the onset of flow reversal with height. By 0700 UTC, low-level winds become southeasterly while winds aloft remain northwesterly. Persisting until about 0800 UTC, this flow reversal pattern is indicative of a rotor circulation over the Rock Springs network with the reversal flow below about ~100 m and perpendicular to Tussey Ridge into the Nittany Valley.

Figure 5-11 presents horizontal trajectories from 0600 - 0800 UTC as in Fig. 5-1 but for SEP16. In most members, the trajectories move southeast across Tussey Ridge and into the adjacent valley; however, trajectories in several members (e.g., QNSE-CTRL) deviate from the mean flow from the NW and track along Tussey Ridge and into the Nittany Valley.

Corresponding trajectory vertical cross sections reveal the influence of a circulation on trajectories in several members during this particular release (e.g., MYJ-BSL, MYNN-BSL, QNSE-CTRL, QNSE-BSL; Fig. 5-12). Trajectory vertical cross sections are also presented for the 1.3-km ensemble configuration for comparison to the 0.4-km ensemble results (Fig. 5-13). For the 1.3-km ensemble, none of the members demonstrate the presence of a rotor circulation for this release. Of all of the members, only a single trajectory for the QNSE-CTRL even demonstrates notable vertical extent. This discrepancy between the two ensemble configuration outcomes may stem from differences in the terrain resolved by each configuration (Fig. 5-14). Although the 1.3-km nested grid (see Fig. 2-1) provides relatively high horizontal resolution, it is less capable of resolving the fine structure of the terrain and broadens and lowers the ridges surrounding the Nittany Valley. Therefore, it is less capable of resolving the presence and structure of small mesogamma scale circulations related to flow interaction with this terrain. As indicated by potential temperature contours in Fig. 5-13, the 1.3-km ensemble members appear to resolve trapped-lee waves over the Nittany Valley but are unable to produce rotor circulations for this case.

SCIPUFF surface dosage plumes for a 0600 - 0800 UTC release demonstrate a broad range of AT&D outcomes exhibited by the ensemble members (Fig. 5-15). The surface dosage plumes that include diffusion effects show strong agreement with their corresponding horizontal transport trajectories. These surface dosage values and the spatial extent of the surface dosage plumes are closely related to the variations in transport associated with the rotor circulations resolved by the model. For example, the MYNN-CTRL surface dosage plume crosses Tussey Ridge and moves SE in accordance with synoptic scale NW flow. Conversely, the dosage plume for the QNSE-CTRL member that depicted a rotor circulation in its trajectories actually moves back into the Nittany Valley before drifting SE over Tussey Ridge.

Figure 5-16 presents explicit ensemble surface dosage probabilities and those probabilities derived from a single-member (ensemble mean) approach. The single-member approach misses the western extent of the 5% contour in Fig. 5-16a. Only one member, the YSU-FDDA, demonstrates spatial plume extent into the western half of the domain. Although the single-member ensemble-mean dosage probabilities do not exactly match those of the explicit ensemble, the approach again gives a clear indication of the overall range of potential outcomes from the explicit ensemble and the spatial area most likely to be impacted in a chem-bio release and at a much reduced cost. Use of the best member (MYJ-CTRL) yields a very similar result (not shown). From a hazard prediction standpoint, wind reversal regions and along-valley flow associated with rotor circulations are both very important to surface dosage prediction. Predicting the correct location and timing of these motions is especially difficult, but a 12-member ensemble appears capable of providing reasonable spread among AT&D outcomes important for quantifying forecast uncertainty for a chem-bio release.

5.2.2 SEP16 Ensemble Performance Evaluation

CRPS values for temperature, wind speed, across-valley, and along-valley wind are presented in Fig. 5-17. Temperature, wind speed, and across-valley wind forecasts again improve from the 1.3-km ensemble to the 0.4-km ensemble. Significance testing reveals that improvements in CRPS, reliability, and resolution (potential CRPS) are statistically significant for wind speed and across-valley wind while improvements in CRPS and reliability for temperature are also statistically significant (Table 5-3). As opposed to the warm bias reduction demonstrated in SEP13, temperature improvements in this case are mainly due to a cold forecast bias reduction from the 1.3-km to 0.4-km ensemble (see RCRV bias values in Table 5-4).

The largest CRPS improvement is for the across-valley wind forecast, the variable most directly impacted by rotor reversal flow over the Rock Springs network. This ~47% improvement in CRPS is largely due to a reduction in forecast bias and to a lesser extent is due to a resolution improvement. The bias reduction is evidenced by the across-valley wind rank histogram for the 0.4-km ensemble that tends much closer to rank uniformity when compared to that of the 1.3-km ensemble (Fig. 5-18). For the 0.4-km ensemble winds, RCRV dispersion values close to unity (Table 5-4) and CRPS reliability values of near zero (Table 5-3) indicate almost perfectly reliable ensemble wind forecasts where each member is nearly equally likely to verify.

In this case, the degradation in along-valley wind can be attributed to forecast amplitude errors. Even though the 0.4-km ensemble members resolve more fine physical structure in the along-valley wind time series, numerics associated with the coarser 1.3-km ensemble lead to smoother and less risky forecasts. Because of susceptibility to phase and amplitude errors, a shortcoming of most traditional verification approaches when evaluating fine-scale mesoscale model performance (Mass et al. 2002), the finer 0.4-km ensemble is penalized more harshly by the CRPS. More representative evaluation metrics must be used or developed to better understand model performance when large phase and amplitude errors are likely (Casati et al. 2008).

CRPS results for the multi-initialization ensemble reveal mixed results among configurations for each variable (Fig. 5-19). The QNSE based configuration performs markedly worse for temperature but performs comparably to or better than the MYJ and MYNN multi-initialization configurations for wind variables. As in SEP13, the YSU multi-initialization ensemble performs poorly for each wind speed variable when compared to the other three configurations. Multi-physics ensemble CRPS results show that the FDDA multi-physics ensemble configuration produces superior forecasts for temperature, wind speed, and along-valley wind while performing comparably to the CTRL multi-physics configuration for across-valley wind (Fig. 5-20). This result highlights the importance of data assimilation in complex

forecasting situations involving small mesogamma scale features. Ensemble skill generally appears to depend more on initialization strategy than in the previous SEP13 case. This may not be a surprising result as SEP16 is more synoptically forced, and the presence of rotor circulations within the model is dependent on the correct representation of upstream conditions.

5.3 Cross-Mountain S-SE Flow: NOV06

5.3.1 NOV06 AT&D Results

AT&D forecast uncertainty for NOV06 stems from the presence of rotor circulations over the Rock Springs network due to trapped-lee waves excited by S-SE flow over Tussey Ridge. Figure 4-4 presents SODAR observations of horizontal wind from 0100 - 0700 UTC from the valley location (see Fig. 4-1). The SODAR observations show that wind direction is highly variable throughout most of the night. From about 0200 – 0400 UTC, flow is predominantly W-SW up to ~65 m AGL while flow is predominantly E above ~80 m AGL. This wind reversal with height suggests the presence of low-amplitude rotor circulations over the Rock Springs network. After about 0400 UTC until 0540 UTC, low-level wind direction generally steadies and remains from the SW. Increased wind speed values during this time may be related to enhanced TKE and mixing as a rotor circulation moves closer to the valley SODAR location. Periods of disturbed horizontal flow with height indicative of the presence of weak rotor circulations persist through 0700 UTC.

Figure 5-21 presents horizontal trajectories from 0500 - 0700 UTC. In some members, the trajectories remain relatively close to the Site 9 release point and move along Tussey Ridge toward to the NE; however, trajectories in about half of the members move further into the valley over the 2-h integration. For the latter members, corresponding trajectory vertical cross sections

reveal the complex influence of a rotor circulation over the Rock Springs network (Fig. 5-22). The circulations cause parcels to be lofted above strong thermal stratification at the surface (as indicated by the potential temperature contours) and to be subsequently advected toward the NE. These members exhibit meandering trajectories and some even follow a helical path (e.g., YSU-BSL) which further exemplifies AT&D forecast complications. These motions are consistent with meandering motions described by Seaman et al. (2012).

SCIPUFF surface dosage plumes for a 0500 - 0700 UTC release again demonstrate a broad range of AT&D outcomes exhibited by ensemble members (Fig. 5-23). The surface dosage plumes show strong agreement with corresponding transport trajectories. High surface dosages occur where the trajectories remain within the valley close to the point of release while those members that demonstrate trajectory lofting correspond to more diffuse surface dosages with plumes that extend down valley. For the MYJ members, a combination of weak wind speeds and the lack of a vertical lofting mechanism causes very high surface-dosages near the point of release and plumes that disperse very little over the 2-h integration.

Explicit ensemble dosage probabilities show enhanced values along Tussey Ridge and they fan outward to lower probabilities downwind (Fig. 5-24a) while the ensemble mean single-member approach shows a very broad field of enhanced surface dosage probability downwind indicating the uncertainty in plume dispersion outcomes and in the low-level wind field (Fig. 5-24b). Again, plume augmentation due to the use of wind-field uncertainty statistics appears to somewhat artificially enhance surface dosages (and thus probabilities) south of Tussey Ridge and along the ridge line. Use of the best member, the MYNN-CTRL, again produces a very similar outcome (not shown).

5.3.2 NOV06 Ensemble Performance Evaluation

CRPS results for each variable of interest for the 1.3-km and 0.4-km ensembles are presented in Fig. 5-25. Temperature, wind speed, across-valley wind, and along-valley wind forecasts again improve from the 1.3-km ensemble to the 0.4-km ensemble. Improvements in CRPS and reliability are statistically significant for temperature and improvements in CRPS, reliability, and potential CRPS are all statistically significant for wind speed, across-, and along-valley wind (Table 5-5). As in SEP16, temperature improvements are due to a cold forecast bias reduction from the 1.3-km to 0.4-km ensemble (see RCRV values in Table 5-6).

As in SEP16, a large CRPS improvement is shown for the across-valley wind forecast, although the along-valley wind forecast in this case experienced a similar improvement. Overall, across-valley wind CRPS values are reduced from about 0.40 ms^{-1} to 0.28 ms^{-1}. This is a considerable improvement given the weak, omnidirectional nature of the winds in this case. RCRV bias and dispersion values (Table 5-6) tend closer to 0 and 1 respectively. The across-valley wind rank histogram for the 0.4-km ensemble, although still slightly underdispersive, tends much closer to rank uniformity when compared to that of the 1.3-km ensemble (Fig. 5-26). Low overall CRPS as well as reliability values of near zero (Table 5-5), RCRV bias values of near zero, and dispersion values close to unity (Table 5-6) again indicate accurate and reliable ensemble wind variable forecasts where each ensemble member is nearly equally likely to verify.

As in SEP16, the other perpendicular flow case, CRPS results for the multi-initialization ensembles are mixed among configurations for each variable as no one configuration demonstrates a clear advantage (Fig. 5-27). However, it is noted that the YSU multi-initialization performs markedly worse than the other configurations for the across-valley wind component, and this may also be related to non-TKE turbulence closure as in SEP13. The QNSE multi-initialization ensemble performs markedly worse for temperature due to a strong cold bias among

members. Multi-physics ensemble CRPS values show that the FDDA multi-physics ensemble configuration generally produces the best ensemble forecasts for each of the four variables again highlighting the positive impact of data assimilation when modeling small scale, complex circulations induced by synoptic scale forcing (Fig. 5-28).

Table 5-1. SEP13 summary of CRPS, Reliability, and Potential CRPS for 2-m temperature, wind speed, across-valley wind, and along-valley wind for the 1.3-km ensemble (1.3-km) and the 0.4-km ensemble (0.4-km) along with percent improvement (+)/degradation(-) (%) of the 0.4-km ensemble over the 1.3-km ensemble and significance (Sig.) of the improvement/degradation at the 95% confidence level.

	1.3-km	**0.4-km**	**%**	**Sig.**
CRPS				
Temperature	1.93	1.41	26.9%	Y
Wind Speed	0.94	0.31	67.0%	Y
Across-Valley	0.62	0.27	56.5%	Y
Along-Valley	0.99	0.44	55.6%	Y
Reliability				
Temperature	1.31	0.79	39.7%	Y
Wind Speed	0.89	0.10	88.8%	Y
Across-Valley	0.47	0.02	95.7%	Y
Along-Valley	0.97	0.23	76.3%	Y
Potential CRPS				
Temperature	0.62	0.62	0%	N
Wind Speed	0.05	0.21	-320%	Y
Across Valley	0.15	0.25	-66.7%	Y
Along Valley	0.02	0.20	-900%	Y

Table 5-2. SEP13 summary of RCRV bias and dispersion for 2-m temperature, wind speed, across-valley wind, and along-valley wind for the 1.3-km ensemble (1.3-km) and the 0.4-km ensemble (0.4-km).

	1.3-km Bias	0.4-km Bias	1.3-km Dispersion	0.4-km Dispersion
Temperature	4.09	2.29	3.63	1.94
Wind Speed	5.99	1.20	3.78	2.15
Across-Valley	-2.94	0.27	1.98	1.65
Along-Valley	4.50	1.80	2.09	1.74

Table 5-3. SEP16 summary of CRPS, Reliability, and Potential CRPS for 2-m temperature, wind speed, across-valley wind, and along-valley wind for the 1.3-km ensemble (1.3-km) and the 0.4-km ensemble (0.4-km) along with percent improvement (+)/degradation(-) (%) of the 0.4-km ensemble over the 1.3-km ensemble and significance (Sig.) of the improvement/degradation at the 95% confidence level.

	1.3-km	**0.4-km**	**%**	**Sig.**
CRPS				
Temperature	0.81	0.71	12.3%	Y
Wind Speed	0.47	0.32	31.9%	Y
Across-Valley	0.75	0.40	46.7%	Y
Along-Valley	0.23	0.26	-13.0%	Y
Reliability				
Temperature	0.17	0.08	52.9%	Y
Wind Speed	0.18	0.08	55.6%	Y
Across-Valley	0.34	0.04	88.2%	Y
Along-Valley	0.01	0.01	0%	N
Potential CRPS				
Temperature	0.64	0.63	1.56%	N
Wind Speed	0.29	0.24	17.2%	Y
Across Valley	0.41	0.36	12.2%	Y
Along Valley	0.23	0.25	-8.70%	Y

Table 5-4. SEP16 summary of RCRV bias and dispersion for 2-m temperature, wind speed, across-valley wind, and along-valley wind for the 1.3-km ensemble (1.3-km) and the 0.4-km ensemble (0.4-km).

	1.3-km Bias	0.4-km Bias	1.3-km Dispersion	0.4-km Dispersion
Temperature	-0.73	-0.55	2.05	1.47
Wind Speed	1.25	0.68	1.64	1.21
Across-Valley	1.77	0.45	1.88	1.11
Along-Valley	-0.19	0.23	1.24	1.09

Table 5-5. NOV06 summary of CRPS, Reliability, and Potential CRPS for 2-m temperature, wind speed, across-valley wind, and along-valley wind for the 1.3-km ensemble (1.3-km) and the 0.4-km ensemble (0.4-km) along with percent improvement (+)/degradation(-) (%) of the 0.4-km ensemble over the 1.3-km ensemble and significance (Sig.) of the improvement/degradation at the 95% confidence level.

	1.3-km	**0.4-km**	**%**	**Sig.**
CRPS				
Temperature	1.39	1.30	6.47%	Y
Wind Speed	0.29	0.19	34.5%	Y
Across-Valley	0.40	0.28	30.0%	Y
Along-Valley	0.31	0.24	22.6%	Y
Reliability				
Temperature	0.81	0.63	22.2%	Y
Wind Speed	0.06	0.01	83.3%	Y
Across-Valley	0.09	0.03	66.7%	Y
Along-Valley	0.08	0.02	75.0%	Y
Potential CRPS				
Temperature	0.58	0.67	-15.5%	Y
Wind Speed	0.23	0.18	21.7%	N
Across Valley	0.31	0.25	19.4%	Y
Along Valley	0.23	0.21	8.70%	Y

Table 5-6. NOV06 summary of RCRV bias and dispersion for 2-m temperature, wind speed, across-valley wind, and along-valley wind for the 1.3-km ensemble (1.3-km) and the 0.4-km ensemble (0.4-km).

	1.3-km Bias	0.4-km Bias	1.3-km Dispersion	0.4-km Dispersion
Temperature	-1.97	-1.50	1.55	1.40
Wind Speed	0.86	-0.08	1.81	1.23
Across-Valley	-1.44	0.08	2.30	1.31
Along-Valley	1.06	0.33	2.55	1.62

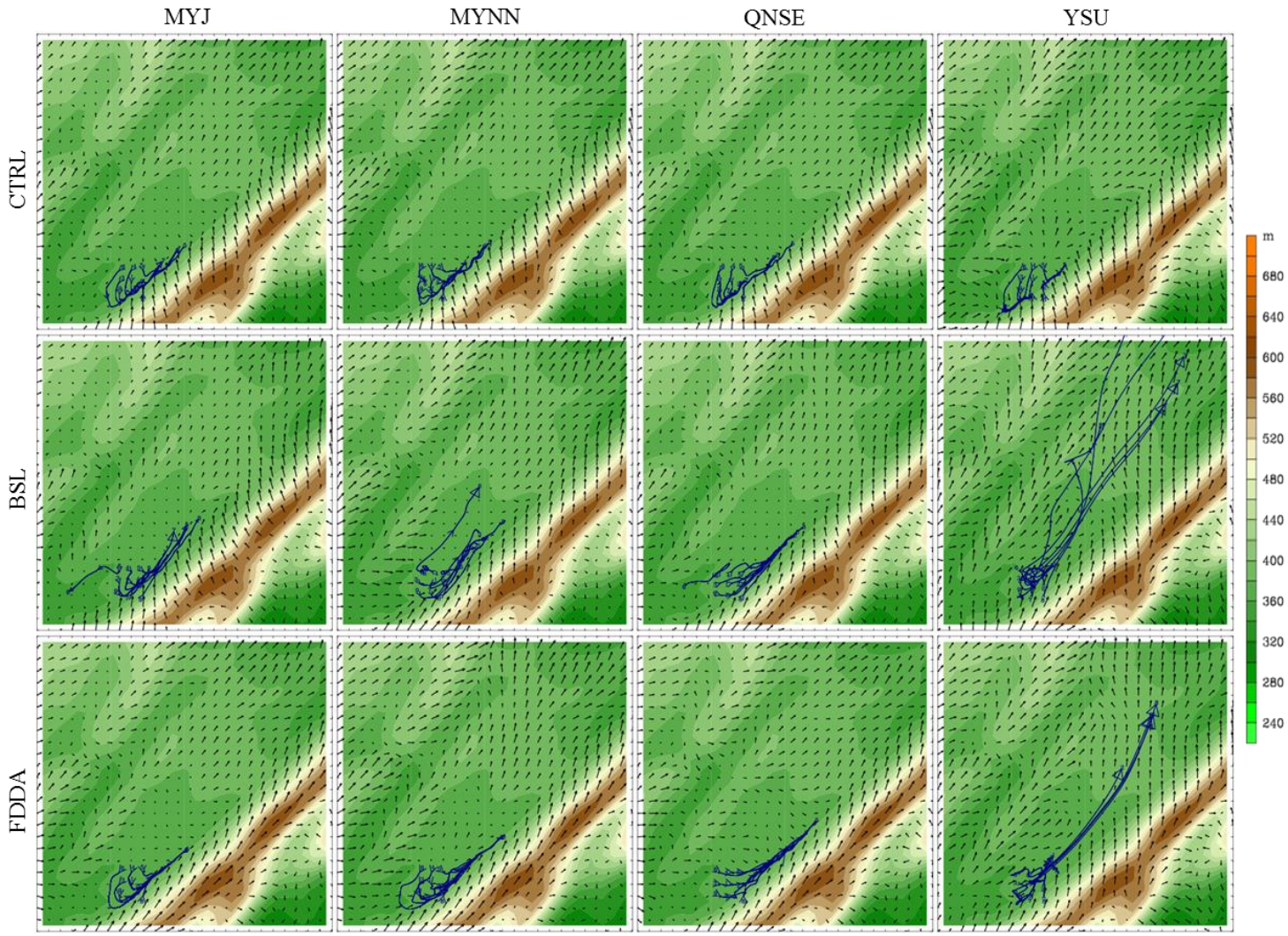

Figure 5-1. SEP13 horizontal trajectories (blue) for a 3-m AGL release from nine grid cells surrounding Site 9 from 0600 – 0800 UTC for each of the 0.4-km ensemble members summarized in Table 2-1. Wind vectors (black) at 3 m AGL at 0800 UTC with identical scaling are included. The plot area is approximately 13 km x 13 km, and the region of enhanced topography is Tussey Ridge.

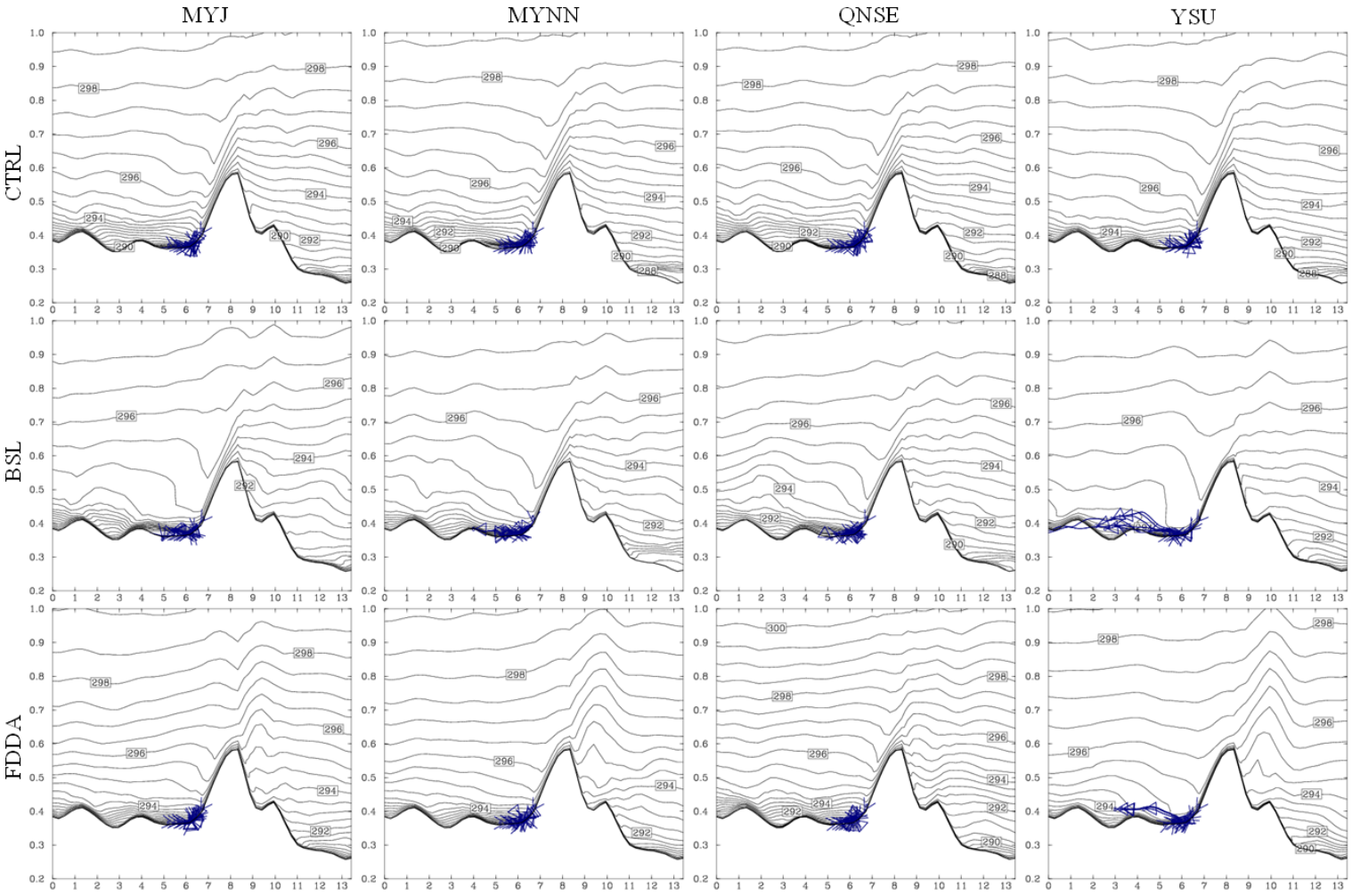

Figure 5-2. SEP13 trajectory (blue) vertical cross sections for a 3-m AGL release from nine grid cells surrounding Site 9 from 0600 – 0800 UTC for each of the 0.4-km ensemble members summarized in Table 2-1. Potential temperature (K) contours (black; 0.5 K intervals) are plotted at 0800 UTC for reference. The NE to SW cross section corresponds to the location of the dashed line in Fig. 4-1, and the region of enhanced topography is Tussey Ridge. The y-axis label indicates height (km; MSL) and the x-axis label indicates distance (km) along the cross-section.

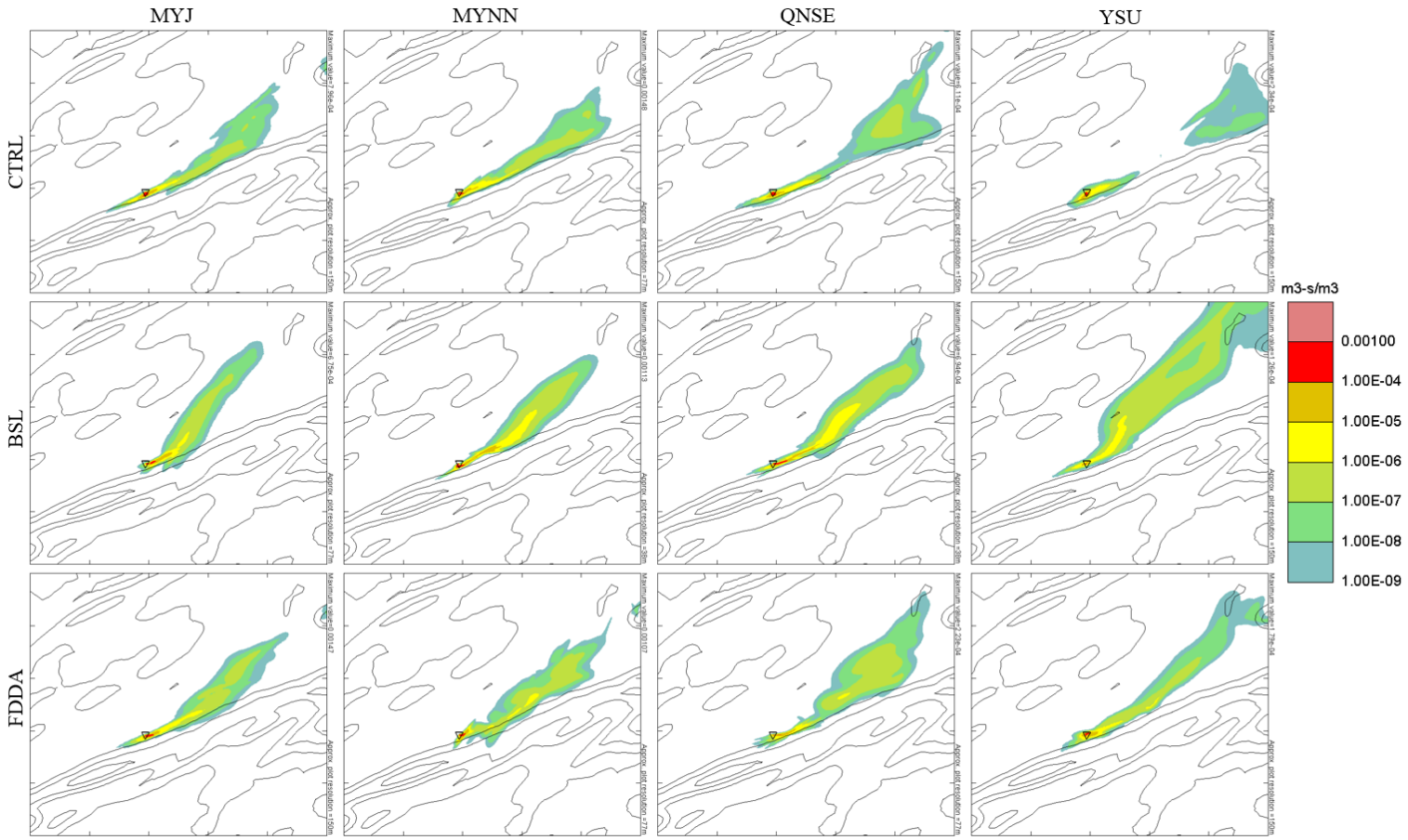

Figure 5-3. SEP13 0.4-km ensemble derived SCIPUFF surface dosage concentration plumes following a 3-m AGL Site 9 release of the passive tracer C_7F_{14} from 0600 – 0800 UTC for each of the ensemble members summarized in Table 2-1. The plot area is approximately 25 km x 25 km.

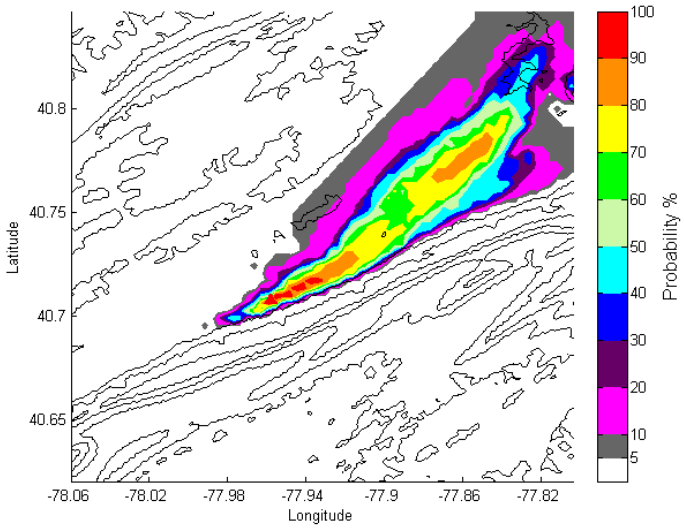

Figure 5-4. SEP13 probabilities of meeting or exceeding a surface dosage threshold concentration of 10^{-9} m^3-s m^{-3} for a 3-m AGL Site 9 release of the passive tracer C_7F_{14} for the explicit ensemble. Terrain contours are from 200 – 600 m MSL by 100 m using 90-m resolution data.

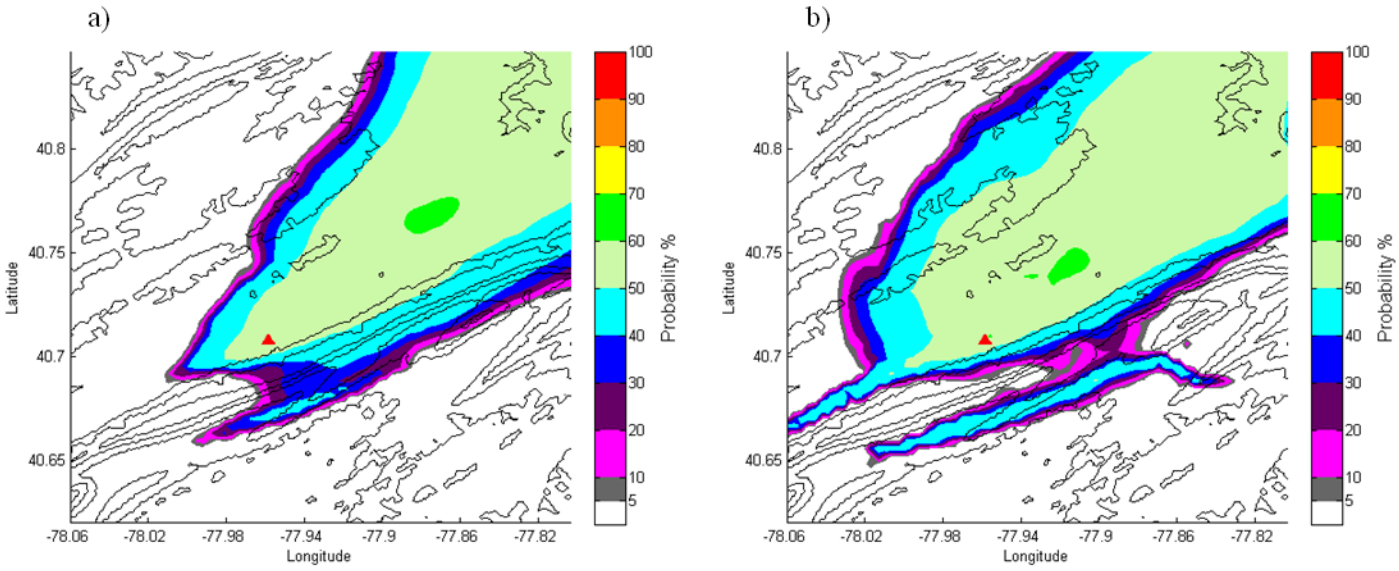

Figure 5-5. SEP13 probabilities of meeting or exceeding a surface dosage threshold concentration of 10^{-9} m^3-s m^{-3} for a 3-m AGL Site 9 release of the passive tracer C_7F_{14} for the a) ensemble mean and the b) ensemble best member (MYJ-CTRL) single-member SCIPUFF hazard mode approaches. Terrain contours are from 200 – 600 m MSL by 100 m using 90-m resolution data.

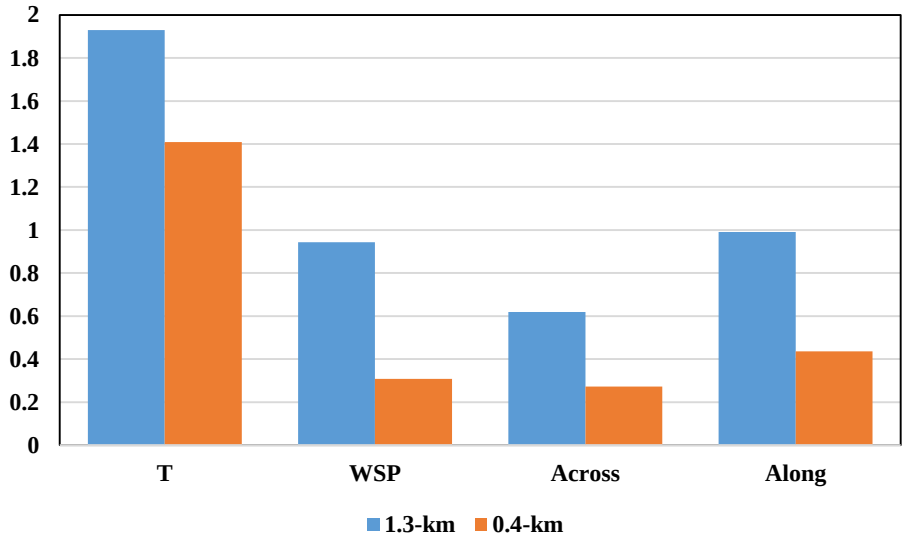

Figure 5-6. SEP13 nighttime CRPS values for 2-m temperature (T), wind speed (WSP), across-valley wind (Across), and along-valley wind (Along) for both the 1.3-km full ensemble configuration and the 0.4-km full ensemble configuration.

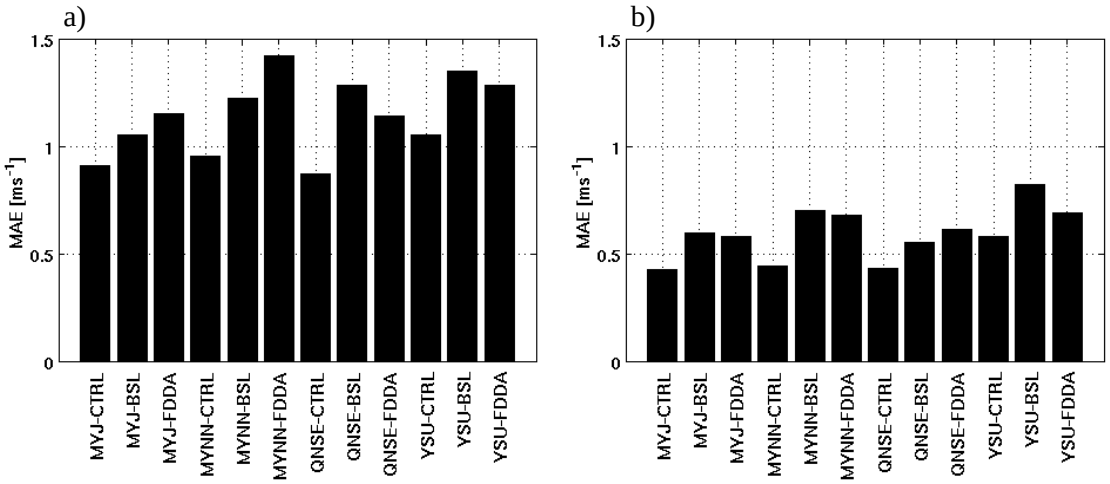

Figure 5-7. SEP13 along-valley wind MAE for each member of the a) 1.3-km ensemble and the b) 0.4-km ensemble.

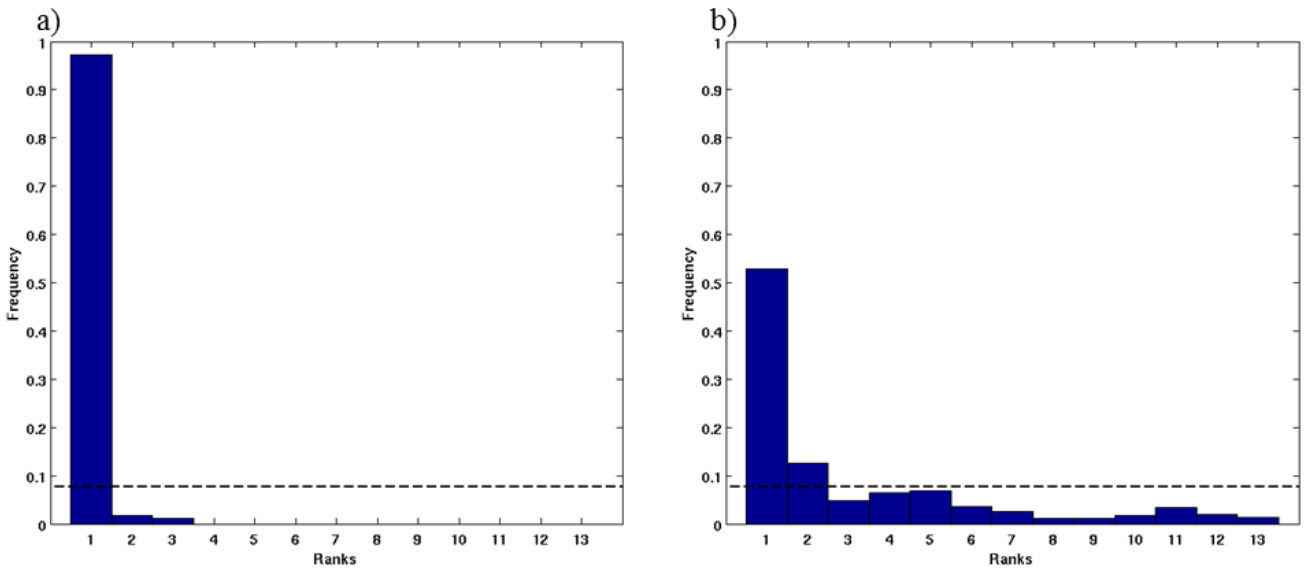

Figure 5-8. SEP13 along-valley wind rank histograms for the a) 1.3-km ensemble and the b) 0.4-km ensemble.

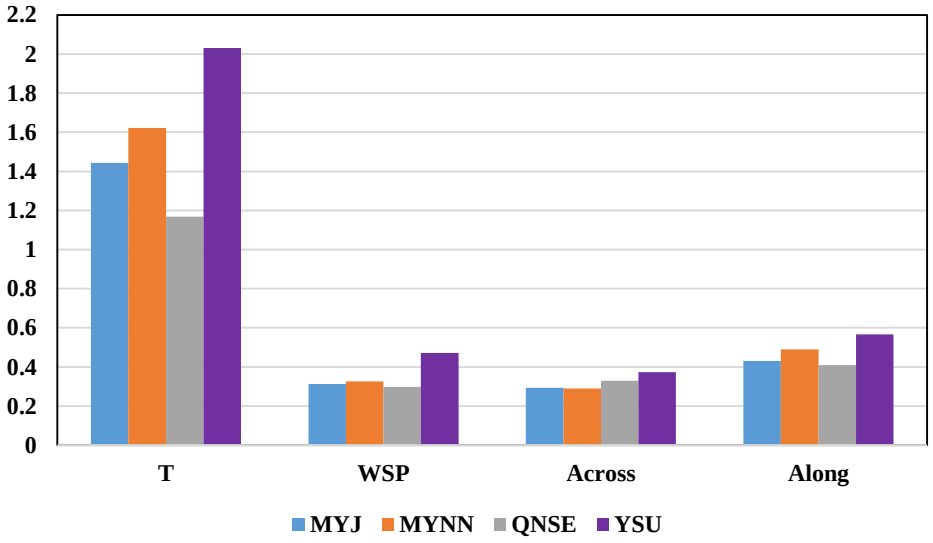

Figure 5-9. SEP13 nighttime CRPS values for 2-m temperature (T), wind speed (WSP), across-valley wind (Across), and along-valley wind (Along) for four 0.4-km multi-initialization ensemble configurations.

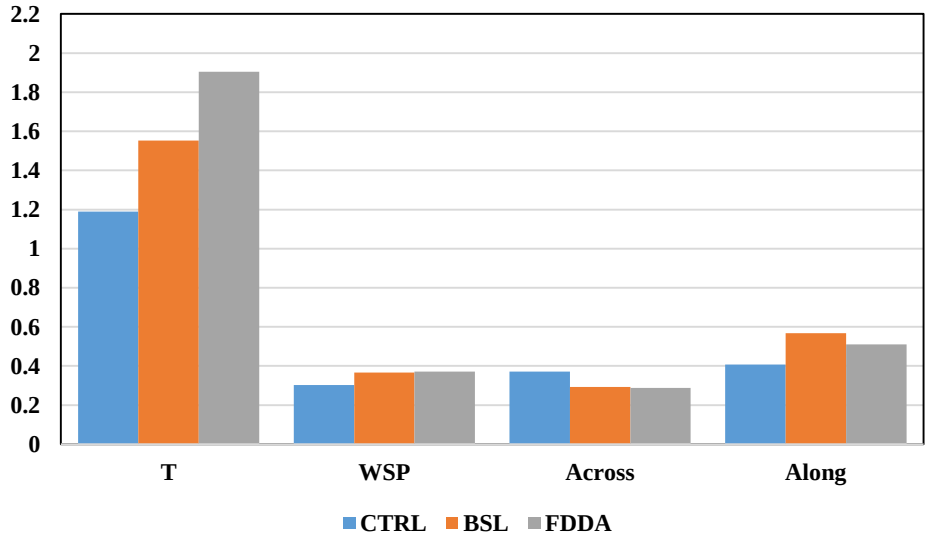

Figure 5-10. SEP13 nighttime CRPS values for 2-m temperature (T), wind speed (WSP), across-valley wind (Across), and along-valley wind (Along) for three 0.4-km multi-physics ensemble configurations.

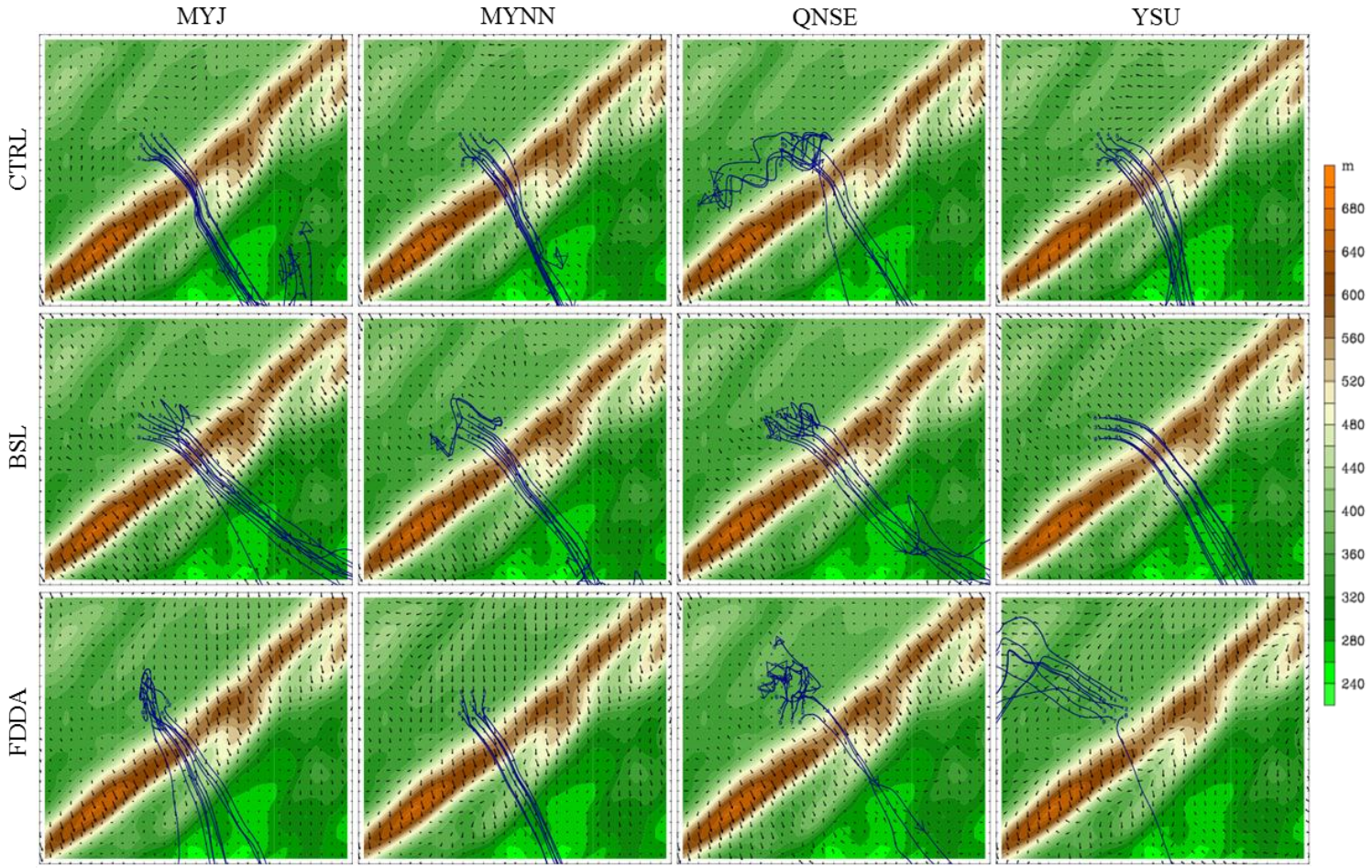

Figure 5-11. SEP16 horizontal trajectories (blue) for a 3-m AGL release from nine grid cells surrounding Site 9 from 0600 – 0800 UTC for each of the 0.4-km ensemble members summarized in Table 2-1. Wind vectors (black) at 3 m AGL at 0800 UTC with identical scaling are included. The plot area is approximately 15 km x 15 km, and the region of enhanced topography is Tussey Ridge.

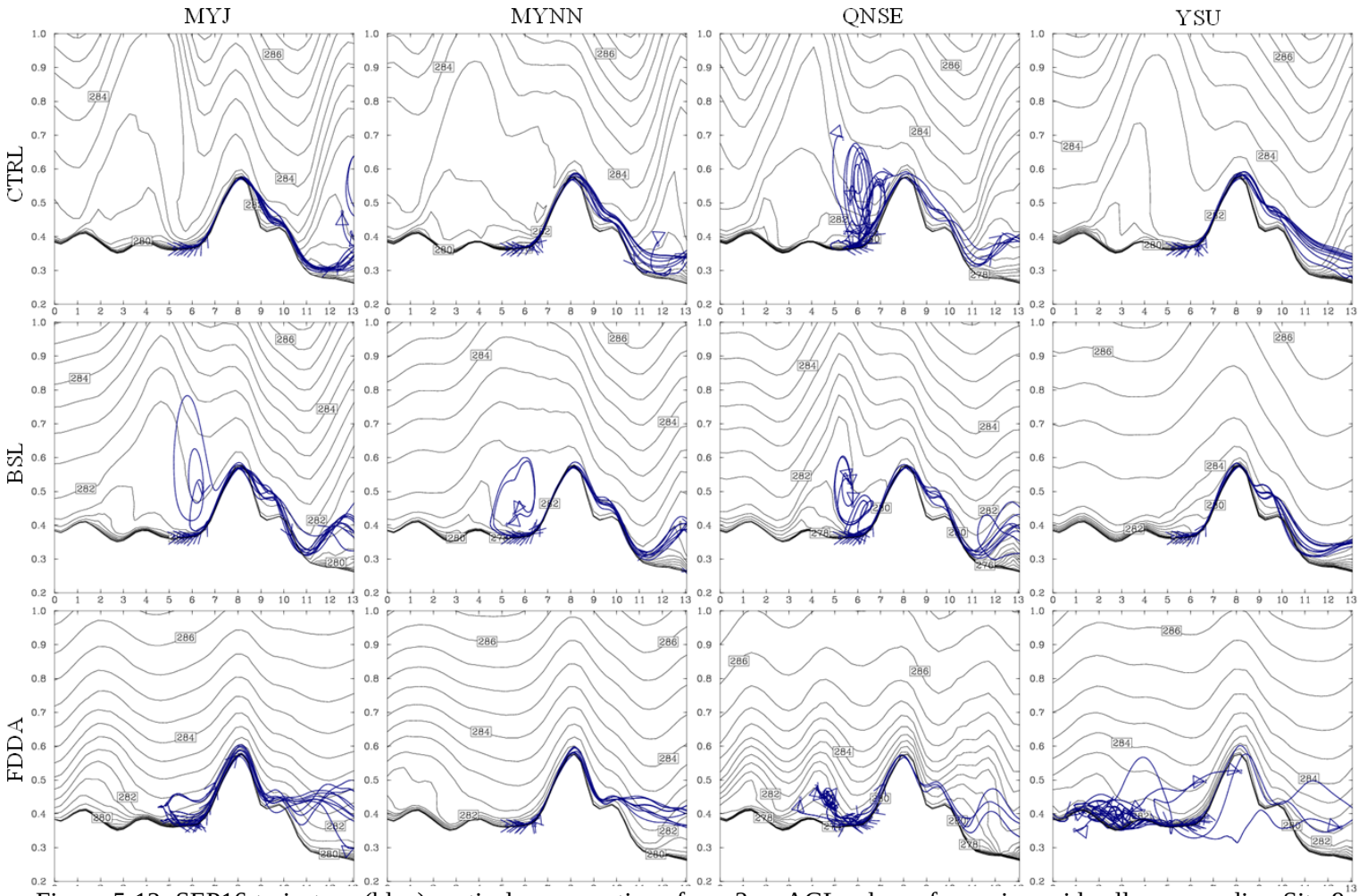

Figure 5-12. SEP16 trajectory (blue) vertical cross sections for a 3-m AGL release from nine grid cells surrounding Site 9 from 0600 – 0800 UTC for each of the 0.4-km ensemble members summarized in Table 2-1. Potential temperature (K) contours (black; 0.5 K intervals) are plotted at 0800 UTC for reference. The NE to SW cross section corresponds to the location of the dashed line in Fig. 4-1, and the region of enhanced topography is Tussey Ridge. The y-axis label indicates height (km; MSL) and the x-axis label indicates distance (km) along the cross-section.

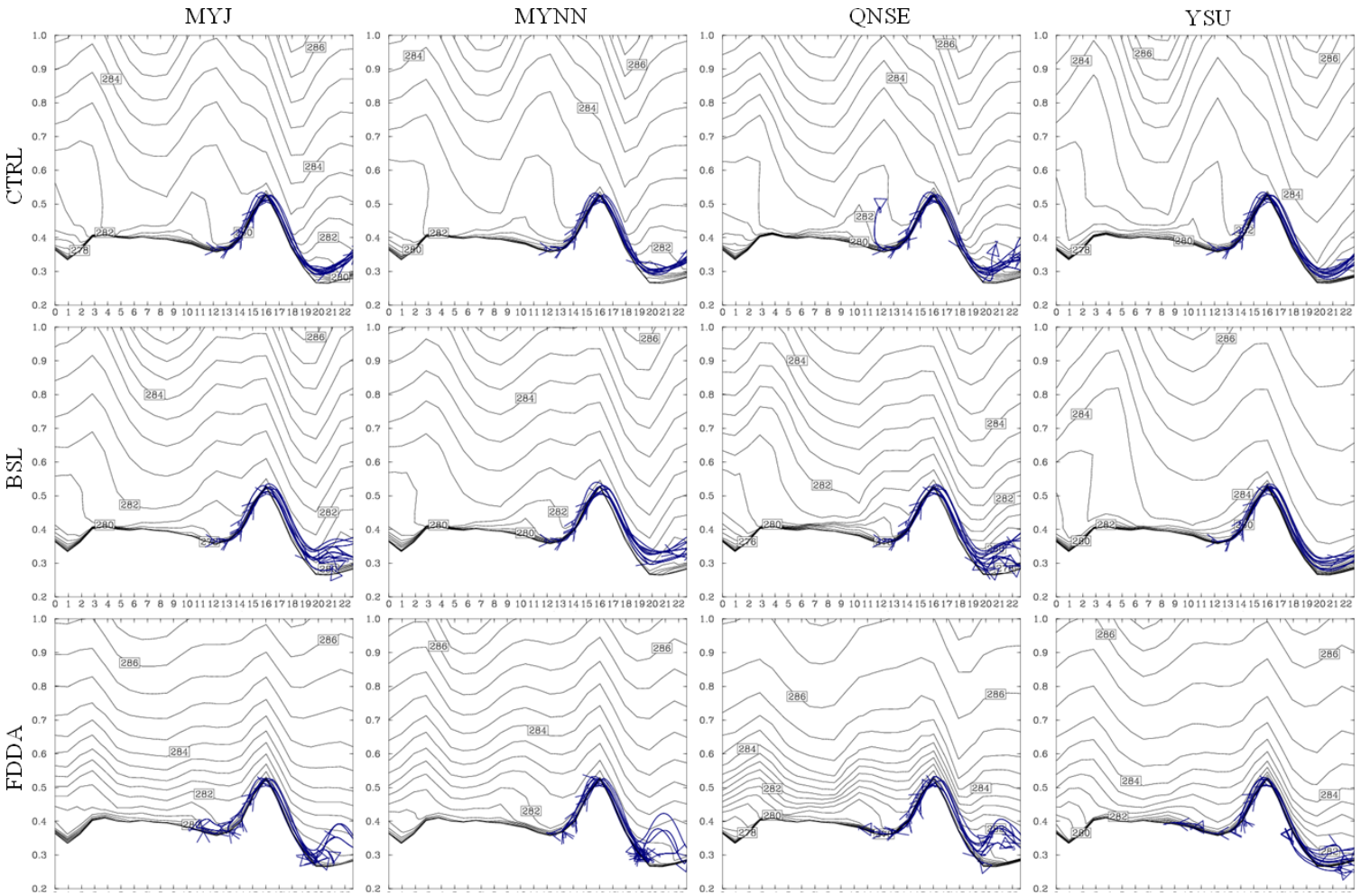

Figure 5-13. SEP16 trajectory (blue) vertical cross sections for a 3-m AGL release from nine grid cells surrounding Site 9 from 0600 – 0800 UTC for each of the 1.3-km ensemble members summarized in Table 2-1. Potential temperature (K) contours (black; 0.5 K intervals) are plotted at 0800 UTC for reference. The NE to SW cross section corresponds closely to the location of the dashed line in Fig. 4-1, and the region of enhanced topography is Tussey Ridge. The y-axis label indicates height (km; MSL) and the x-axis label indicates distance (km) along the cross-section.

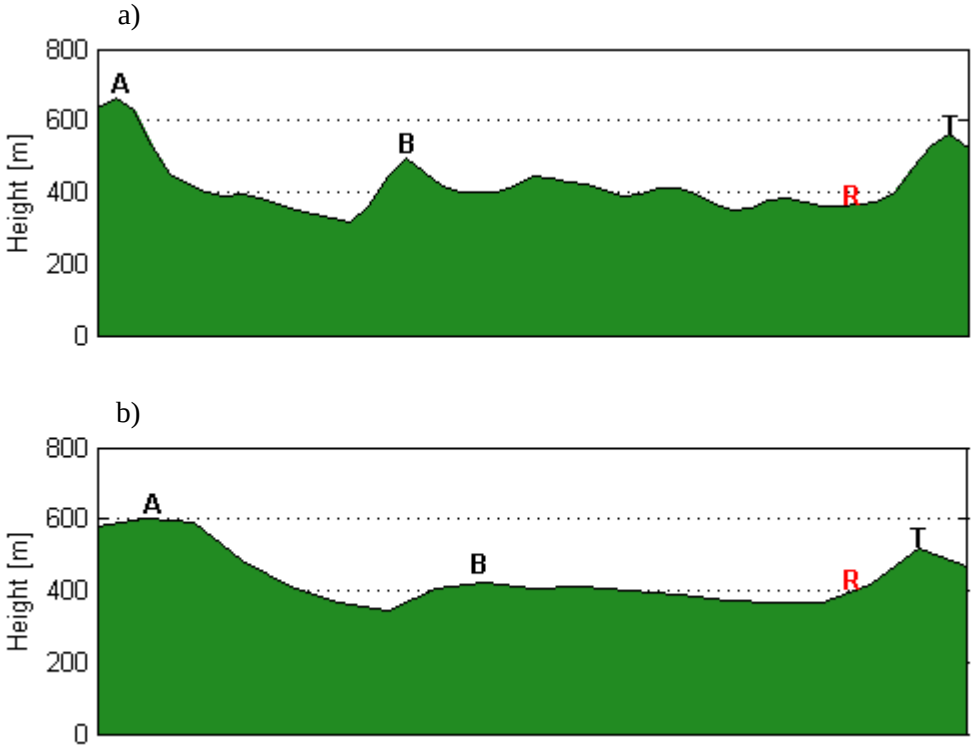

Figure 5-14. Nittany Valley NW to SE terrain cross section as resolved by the a) 0.4-km nested grid and the b) 1.3-km nested grid. "A" represents the Allegheny Mts., "B" represents Bald Eagle Mtn., "T" represents Tussey Ridge, and the red "R" denotes the location of the Rock Springs observation network. These features can be referenced in Fig. 4-1.

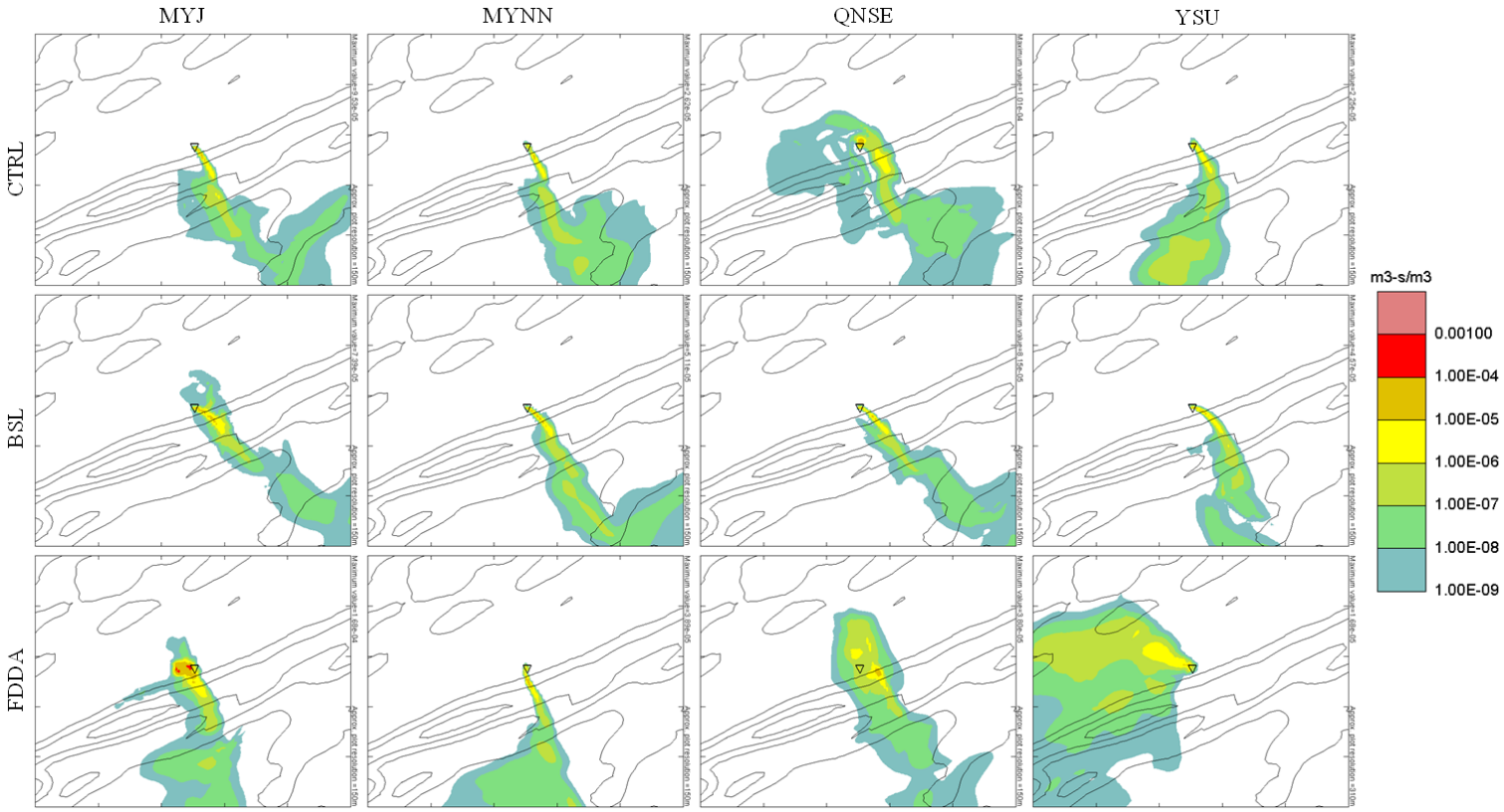

Figure 5-15. SEP16 0.4-km ensemble derived SCIPUFF surface dosage concentration plumes following a 3-m AGL Site 9 release of the passive tracer C_7F_{14} from 0600 – 0800 UTC for each of the ensemble members summarized in Table 2-1. The plot area is approximately 17 km x 17 km.

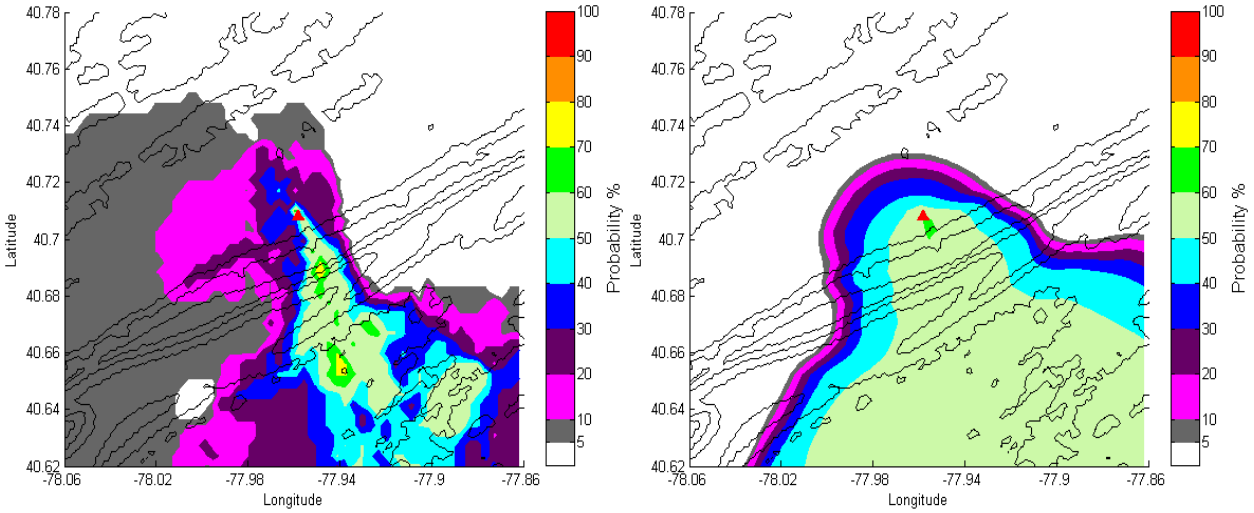

Figure 5-16. SEP16 probabilities of meeting or exceeding a surface dosage threshold concentration of 10^{-9} m^3-s m^{-3} for a 3-m AGL Site 9 release of the passive tracer C_7F_{14} for the a) explicit ensemble and the b) ensemble mean single-member SCIPUFF hazard mode approach. Terrain contours are from 200 – 600 m MSL by 100 m using 90-m resolution data.

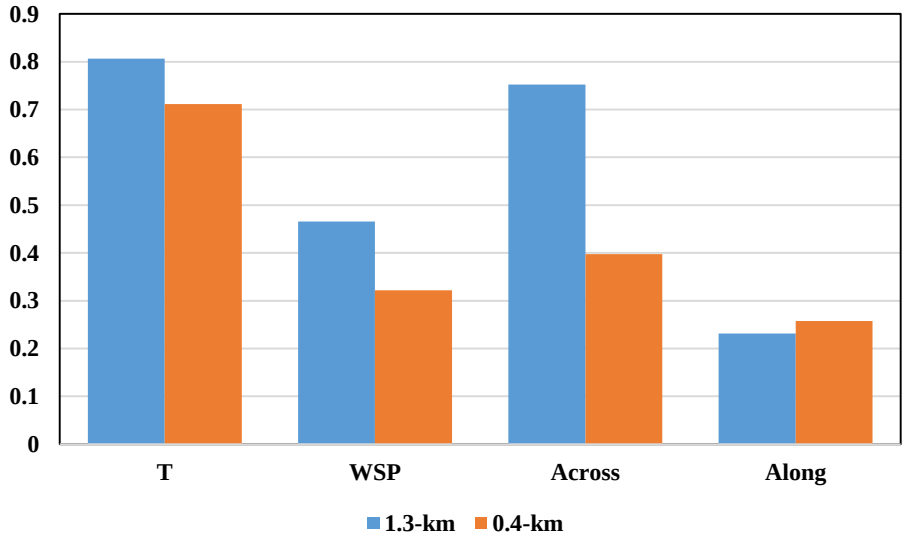

Figure 5-17. SEP16 nighttime CRPS values for 2-m temperature (T), wind speed (WSP), across-valley wind (Across), and along-valley wind (Along) for both the 1.3-km ensemble and the 0.4-km ensemble.

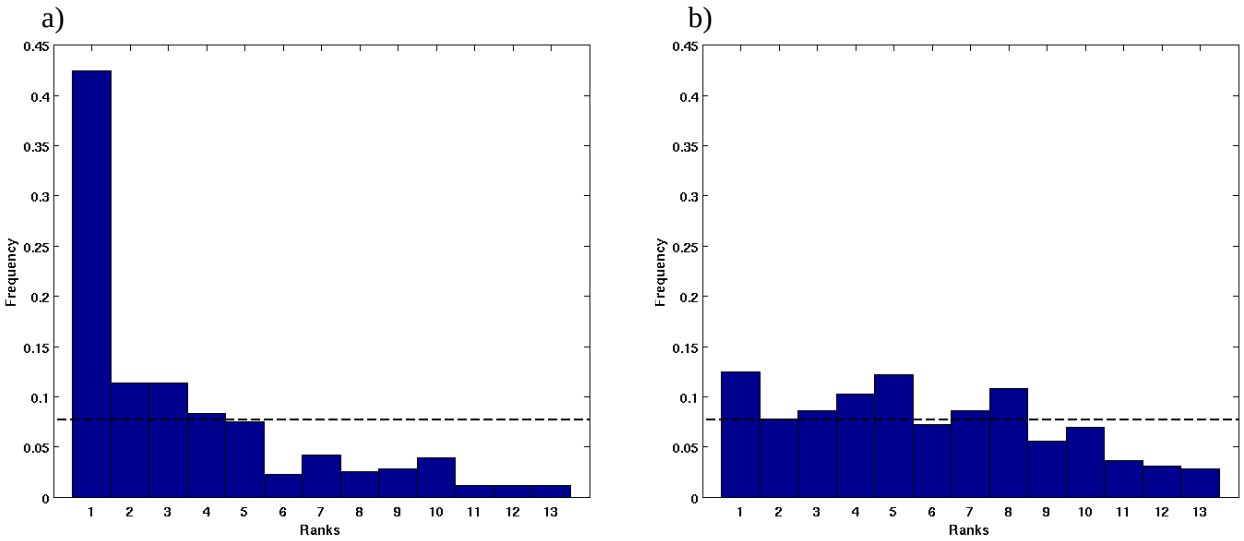

Figure 5-18. SEP16 across-valley wind rank histograms for the a) 1.3-km ensemble and the b) 0.4-km ensemble.

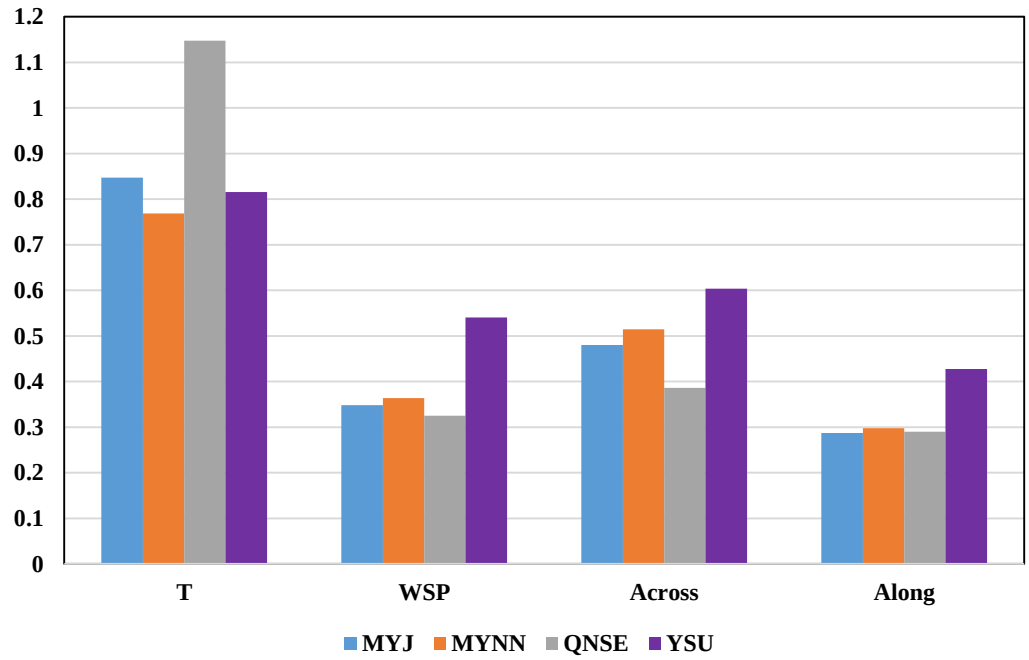

Figure 5-19. SEP16 nighttime CRPS values for 2-m temperature (T), wind speed (WSP), across-valley wind (Across), and along-valley wind (Along) for four 0.4-km multi-initialization ensemble configurations.

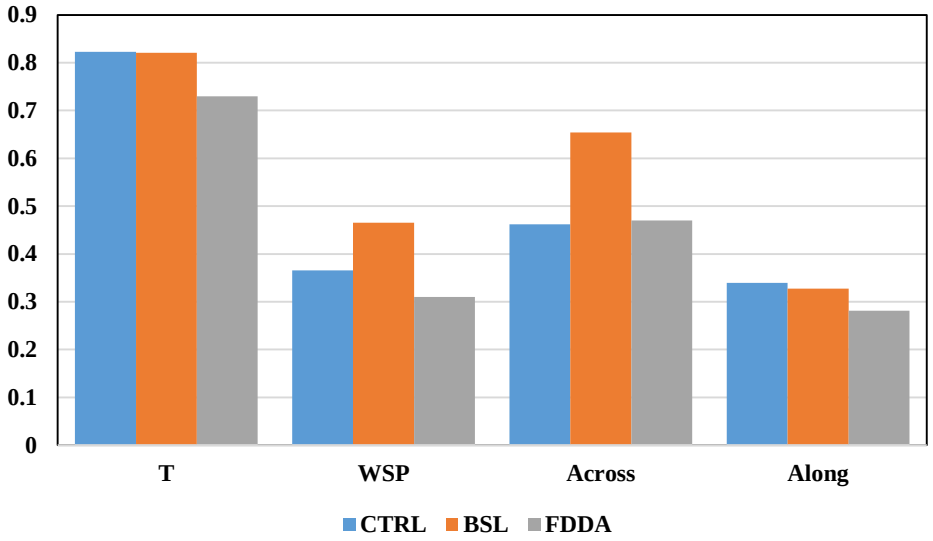

Figure 5-20. SEP16 nighttime CRPS values for 2-m temperature (T), wind speed (WSP), across-valley wind (Across), and along-valley wind (Along) for three 0.4-km multi-physics ensemble configurations.

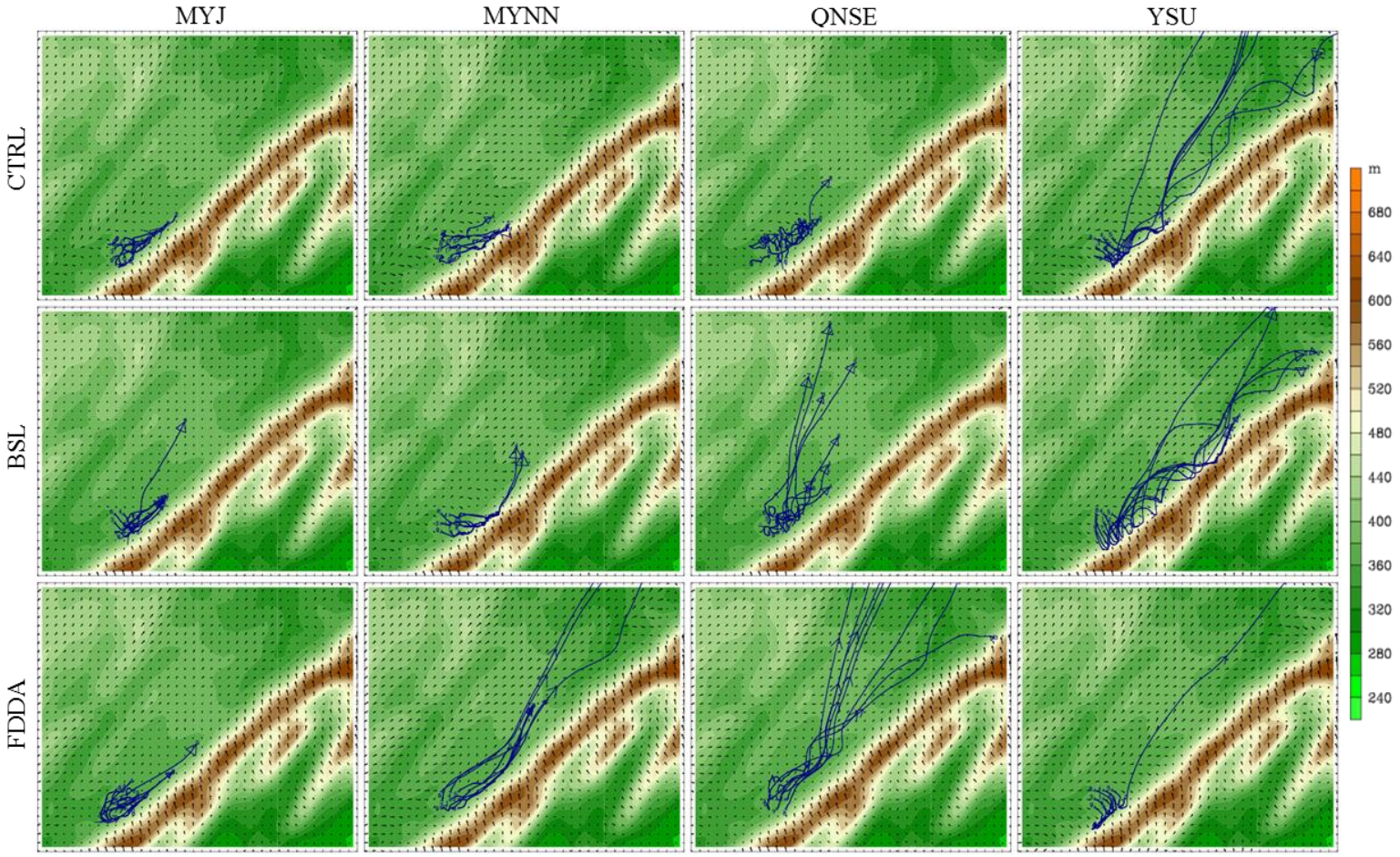

Figure 5-21. NOV06 horizontal trajectories (blue) for a 3-m AGL release from nine grid cells surrounding Site 9 from 0500 – 0700 UTC for each of the 0.4-km ensemble members summarized in Table 2-1. Wind vectors (black) at 3 m AGL at 0700 UTC with identical scaling are included. The plot area is approximately 15 km x 15 km, and the region of enhanced topography is Tussey Ridge.

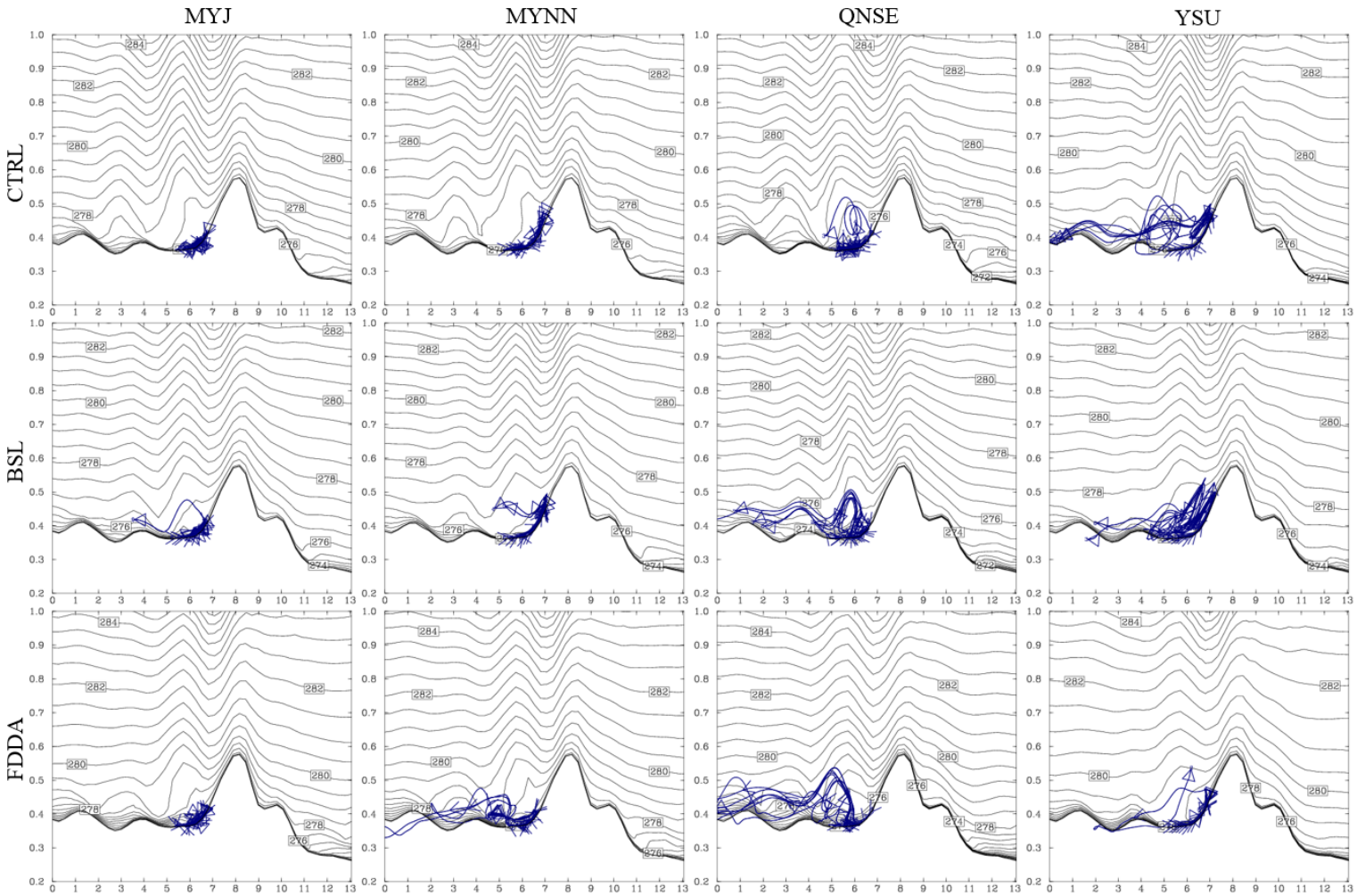

Figure 5-22. NOV06 trajectory (blue) vertical cross sections for a 3-m AGL release from nine grid cells surrounding Site 9 from 0500 – 0700 UTC for each of the 0.4-km ensemble members summarized in Table 2-1. Potential temperature (K) contours (black; 0.5 K intervals) are plotted at 0700 UTC for reference. The NE to SW cross section corresponds to the location of the dashed line in Fig. 4-1, and the region of enhanced topography is Tussey Ridge. The y-axis label indicates height (km; MSL) and the x-axis label indicates distance (km) along the cross-section.

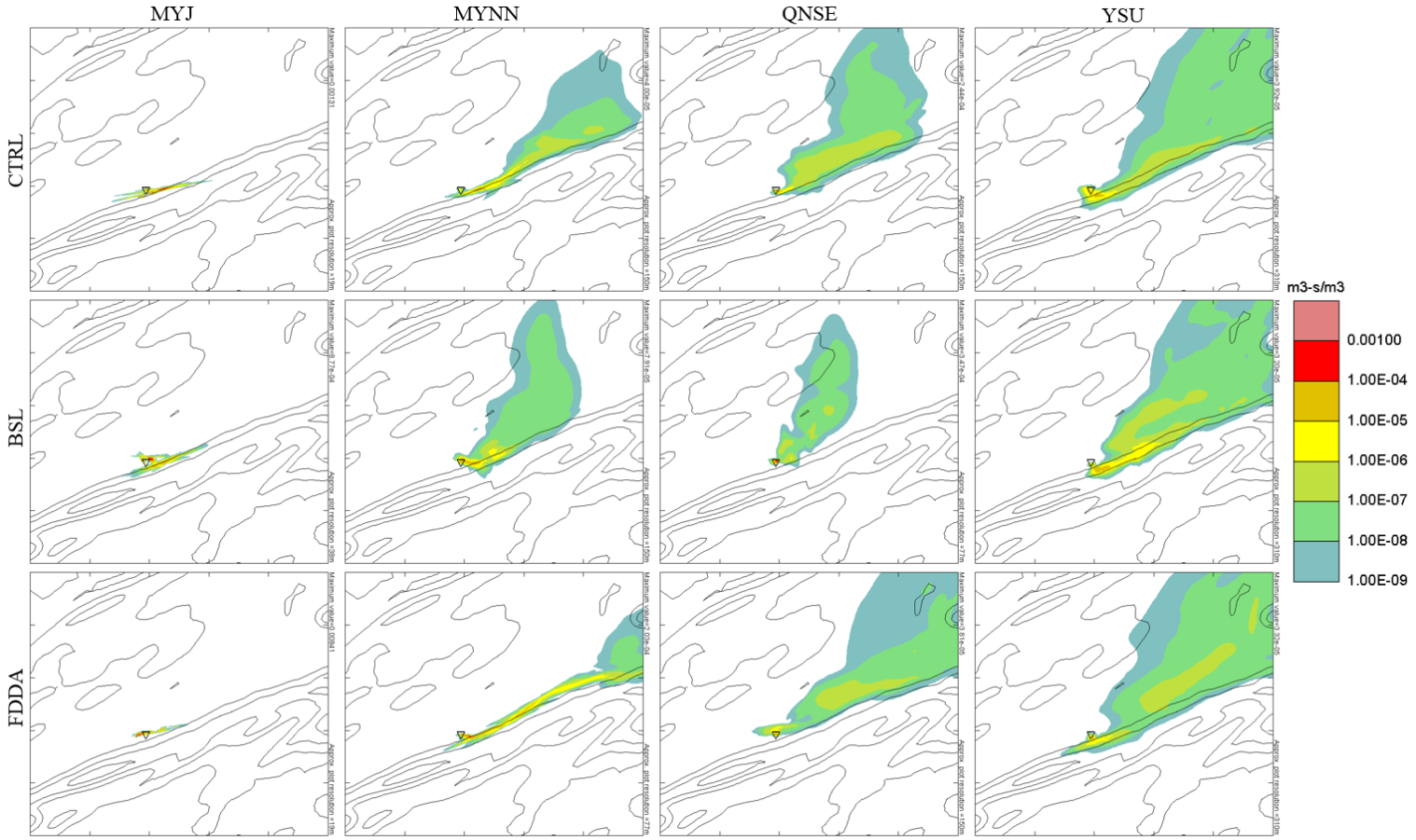

Figure 5-23. NOV06 0.4-km ensemble derived SCIPUFF surface dosage concentration plumes following a 3-m AGL Site 9 release of the passive tracer C_7F_{14} from 0500 – 0700 UTC for each of the ensemble members summarized in Table 2-1. The plot area is approximately 25 km x 25 km.

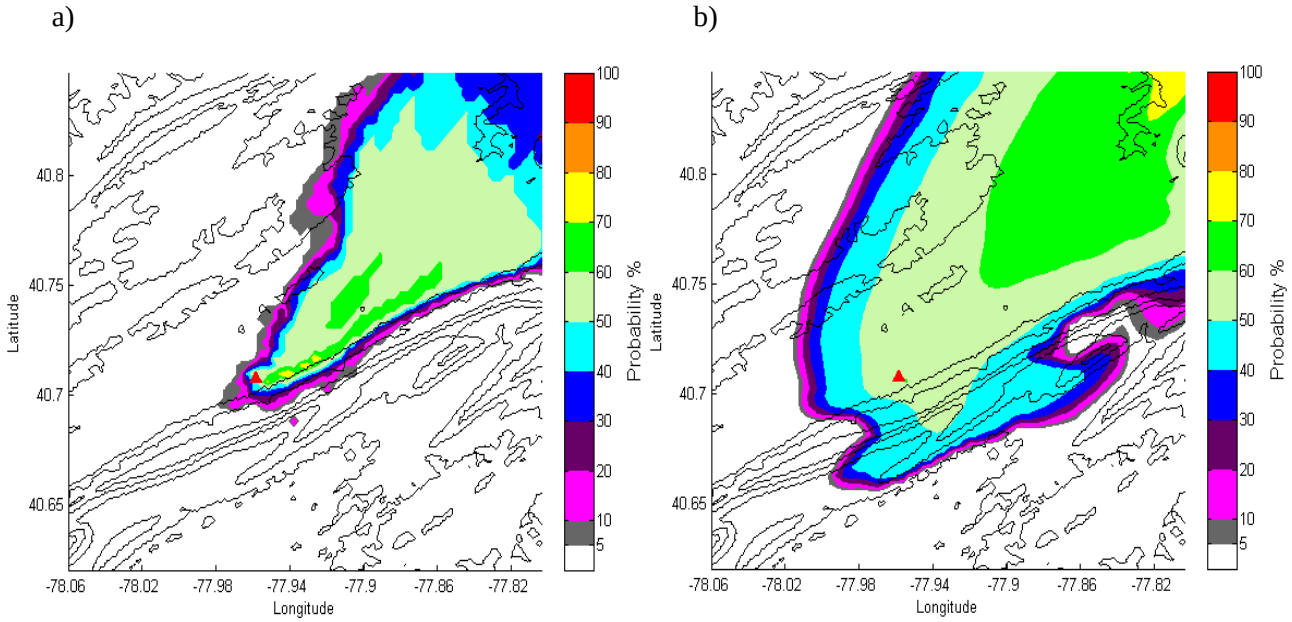

Figure 5-24. NOV06 probabilities of meeting or exceeding a surface dosage threshold concentration of 10^{-9} m^3-s m^{-3} for a 3-m AGL Site 9 release of the passive tracer C_7F_{14} for the a) explicit ensemble and the b) ensemble mean single-member SCIPUFF hazard mode approach. Terrain contours are from 200 – 600 m MSL by 100 m using 90-m resolution data.

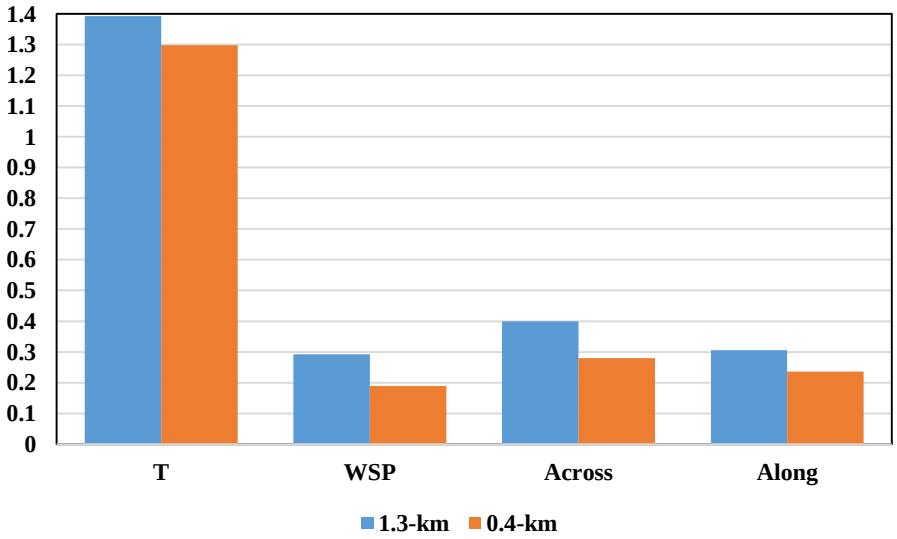

Figure 5-25. NOV06 nighttime CRPS values for 2-m temperature (T), wind speed (WSP), across-valley wind (Across), and along-valley wind (Along) for both the 1.3-km ensemble and the 0.4-km ensemble.

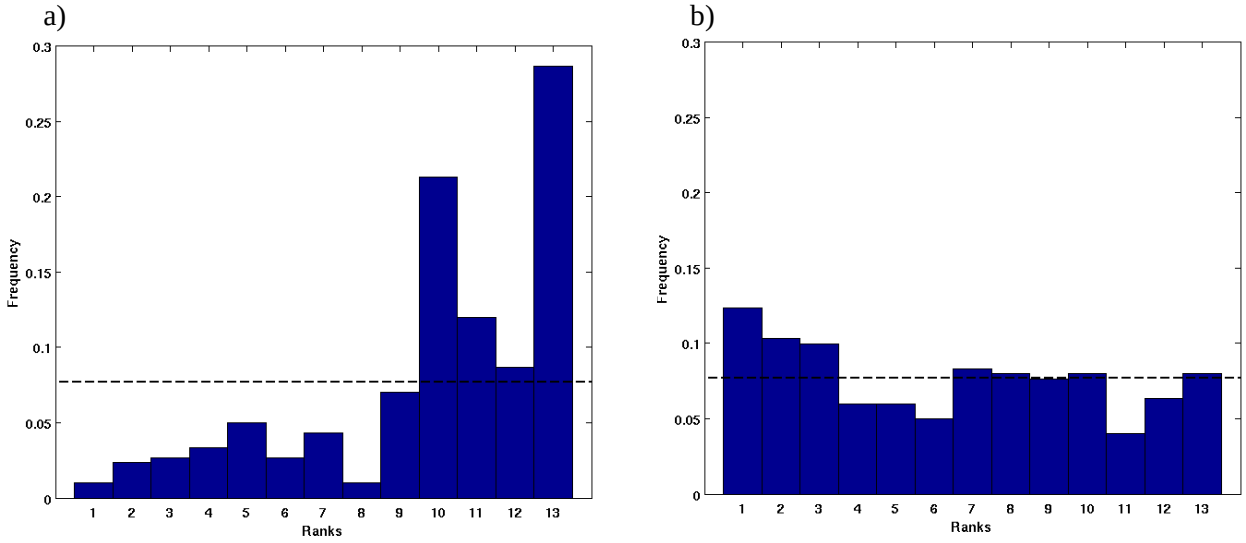

Figure 5-26. NOV06 across-valley wind rank histograms for the a) 1.3-km ensemble and the b) 0.4-km ensemble.

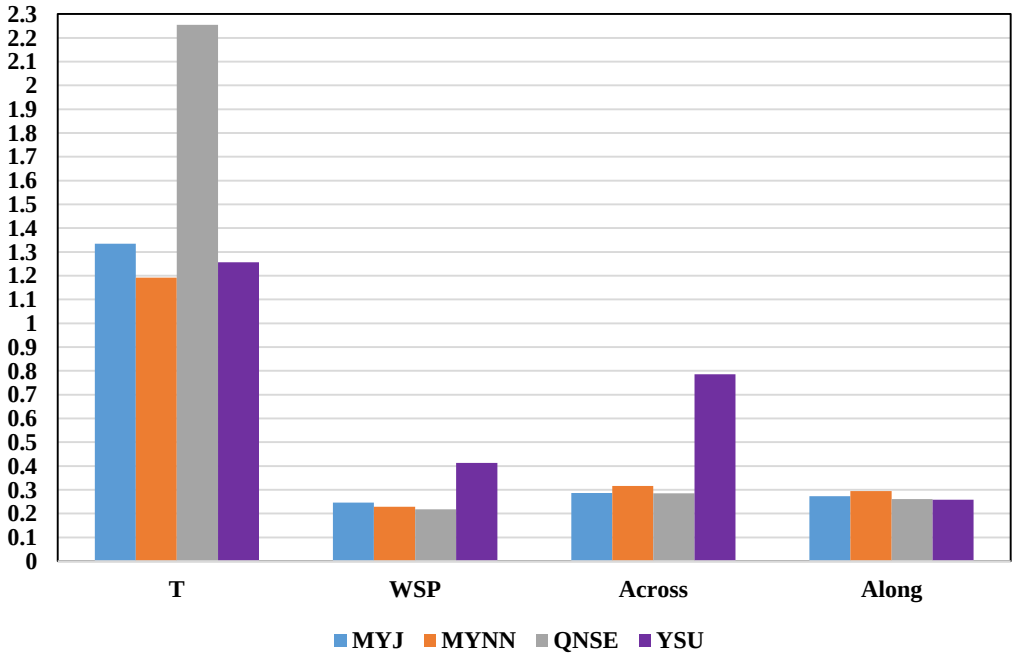

Figure 5-27. NOV06 nighttime CRPS values for 2-m temperature (T), wind speed (WSP), across-valley wind (Across), and along-valley wind (Along) for four 0.4-km multi-initialization ensemble configurations.

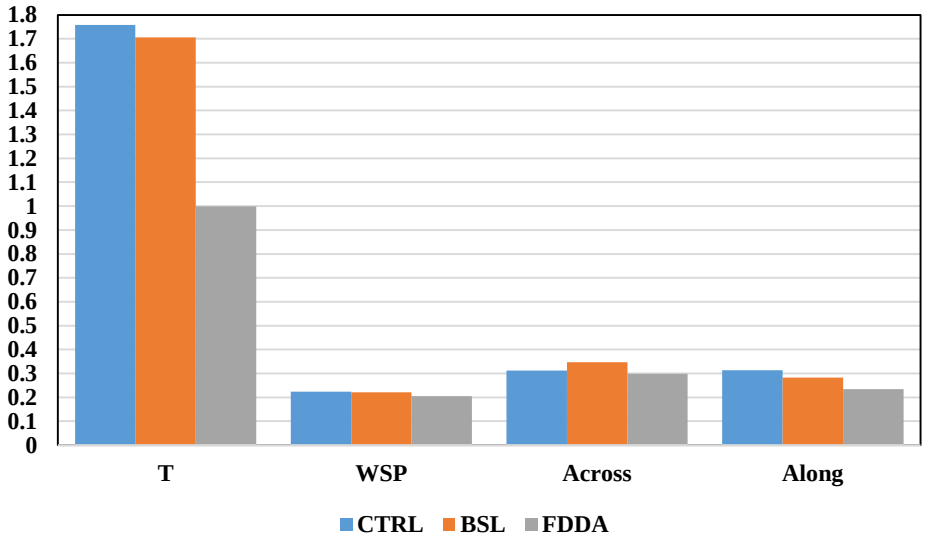

Figure 5-28. NOV06 nighttime CRPS values for 2-m temperature (T), wind speed (WSP), across-valley wind (Across), and along-valley wind (Along) for three 0.4-km multi-physics ensemble configurations.

CHAPTER 6

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS FOR FUTURE WORK

6.1 Summary and Conclusions

A high horizontal and vertical resolution multi-initialization, multi-physics ensemble approach for representing hazard-prediction uncertainty within the SBL over complex terrain is presented. The innermost meteorological model nest has a sub-km horizontal grid length (0.4 km), and all domains feature 2-m vertical grid spacing through the lowest 10 m and 10 vertical levels in the lowest 50 m. The ensemble is designed to sample the range of potential AT&D outcomes related to small-scale drainage flows and mesogamma-scale rotor circulations linked to trapped-lee waves that occur within and above the SBL over complex terrain. Ensemble configurations using both 0.4-km and 1.3-km horizontal grid spacing are tested and compared. The ensembles are created using WRF version 3.6, and ensemble meteorological data are used to compute trajectories and to drive the SCIPUFF AT&D prediction model for three case studies involving distinct flow regimes in reference to the Nittany Valley in central PA. These flow regimes include flow parallel to the Nittany Valley (SEP13), NW flow perpendicular to the Allegheny Mts. (SEP16), and S-SE flow perpendicular to Tussey Ridge (NOV06). AT&D uncertainty is expected in these cases due to drainage flows, trapped-lee waves excited by the Allegheny Mts., and trapped-lee waves excited by Tussey Ridge respectively. Tracer observations are unavailable, and thus, ensemble output for low-level temperature and wind variables is evaluated against low-level temperature and wind data gathered at the Rock Springs, PA observation network using both deterministic and probabilistic verification techniques. It is

assumed that improved low-level wind and temperature forecasts should produce improved SCIPUFF AT&D forecasts for the SBL.

Results show that a 12-member ensemble with sub-km horizontal grid spacing and diversity in initial conditions and planetary boundary layer/surface layer physics is able to produce reasonable spread in AT&D forecasts for the three cases. Sub-km horizontal grid spacing is necessary for resolving the flows most relevant to AT&D uncertainty in each case study. In the drainage flow case, the 0.4-km ensemble AT&D forecasts include the effects of drainage flow convergence, and in both trapped-lee wave cases, 0.4-km ensemble member predictions include the impacts of reversal flows associated with mesogamma-scale wave-induced rotor-circulations over the Nittany Valley and Rock Springs network. In both trapped-lee wave cases, the 1.3-km ensemble members are unable to produce rotor circulations over the Rock Springs network at the assumed release time and location.

For each case study, explicit-ensemble surface dosage probabilities are derived from an aggregation of the 12-member SCIPUFF AT&D ensemble outcomes. Single-member ensemble mean and best-member approaches using meteorological ensemble-derived wind-field uncertainty statistics in SCIPUFF hazard mode are capable of representing the regions of highest explicit-member dosage probabilities with less fine-scale detail but at a much reduced computational cost. The single-member SCIPUFF approach is a valuable tool for a short time-fuse emergency response in the event of a chemical or biological release. The additional detail given by the explicit-ensemble AT&D approach may be more useful for retrospective studies following an incident.

Continuous Ranked Probability Score (CRPS), rank histograms, and Reduced Centered Random Variable (RCRV) results show that sub-km horizontal grid spacing is necessary for providing adequate ensemble spread and for capturing temperature and wind fluctuations directly related to drainage flows and rotor-circulation flows over the Rock Springs network. For each

case study, CRPS reliability values of near zero for low-level variables using the sub-km ensemble configuration generally indicate that forecasted ensemble probabilities are close to the true probabilities of the observed values. In most instances, improvements in CRPS and its components are statistically significant from the 1.3-km ensemble to the 0.4-km ensemble demonstrating added value with increasing horizontal resolution and justifying the increased computational resources necessary to run a sub-km horizontal grid spacing nest in the WRF meteorological model.

Overall, the 12-member WRF ensemble produced the best forecast results. The 12-member is broken into three four-member multi-initialization ensembles and four three-member multi-physics ensembles to get a better sense of the ensemble performance dependence on diversity strategy. In the weakly forced parallel-flow case (SEP13), CRPS results indicate that improved low-level ensemble wind forecasts are generally more dependent on the TKE-based planetary boundary layer/surface layer physics than initialization strategy. In the two synoptically forced cases involving trapped-lee waves and corresponding rotor circulations over the Rock Springs network (SEP16 & NOV06), a dependence on ensemble member initialization strategies is noted, and the use of localized Four Dimensional Data Assimilation (FDDA) generally leads to the best CRPS results and thus superior forecast skill.

6.2 Recommendations for Future Work

Promising case study results illustrate the utility of a sub-km ensemble to quantify AT&D uncertainty in the SBL over complex terrain. However, future work should include tracer data and the creation of a large database of ensemble forecasts over different locations and seasons with verification at and above the surface. This type of dataset would allow for additional tuning of ensemble member diversity strategy and also calibration of ensemble output.

Alternative data assimilation strategies including variational, ensemble Kalman filer, and hybrid methods should be explored on these very fine scales for the SBL. Additional upwind observations including profiles of wind and stratification should be included in future work in order to better assess the utility of data assimilation and to enhance prediction of different wave and rotor types as well as their effects on the SBL and associated hazard predictions. Also, the performance of model configurations including increased horizontal and vertical resolution (i.e., less than 0.4-km horizontal grid spacing and fine vertical spacing up to ridge-top height) should be examined to allow for further evaluation against current model configurations.

REFERENCES

Andre, J.C., G. De Moor, P. Lacarrere, G. Therry, and R. Du Vachat, 1978: Modeling the 24-hour evolution of the mean and turbulent structures of the planetary boundary layer. *J. Atmos. Sci.*, **35**, 1861-1883.

Beljaars, A.C.M., 1994: The parameterization of surface fluxes in large-scale models under free convection, *Quart. J. Roy. Meteor. Soc.*, **121**, 255–270.

Benjamin, S., 2014: From RAPv3/HRRR-2014 to the NARRE/HRRRE eta. Proceedings, *Warn-on-Forecast High Impact Weather Workshop*, Norman, OK, National Weather Center, [Available online at http://ruc.noaa.gov/pdf/Benj_DTC_RAP-NARRE-Apr14.pdf]

Bieringer, P.E., A.J. Annunzio, N. Platt, G. Bieberbach, and J. Hannan, 2014: Contrasting the use of single-realization versus ensemble-average atmospheric dispersion solutions for chemical and biological defense analyses. *J. Appl. Meteor. Climatol.*, **53**, 1399–1415.

Buizza, R., and T.N. Palmer, 1998: Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, **126**, 2503-2518.

Candille, G., C. Cote, P.L. Houtekamer, and G. Pellerin, 2007: Verification of an ensemble prediction system against observations. *Mon. Wea. Rev.*, **135**, 2688-2699.

___, S. Beauregard, and N. Gagnon, 2010: Bias correction and multiensemble in the NAEFS context or how to get a "free calibration" through a multiensemble approach. *Mon. Wea. Rev.*, **138**, 4268-4281.

Casati, B., and Coauthors, 2008: Forecast verification: current status and future directions. *Meteorological Applications*, **15**, 3-18.

Chen, F. and J. Dudhia, 2001: Coupling of an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569-585.

Chow, F.K., S.F.J. DeWekker, and B.J. Snyder, 2013: *Mountain weather research and forecasting: Recent progress and current challenges*. Springer Atmospheric Sciences, 121-205.

Cintineo, R., J.A. Otkin, M. Xue, and F. Kong, 2014: Evaluating the Performance of Planetary Boundary Layer and Cloud Microphysical Parameterization Schemes in Convection-Permitting Ensemble Forecasts Using Synthetic GOES-13 Satellite Observations. *Mon. Wea. Rev.*, **142**, 163–182.

Clark, A.J., J.S. Kain, D.J. Stensrud, M.X., F. Kong, M.C. Coniglio, K.W. Thomas, Y. Wang, K. Brewster, J. Gao, X. Wang, S.J. Weiss, and J. Du, 2011: Probabilistic Precipitation Forecast Skill as a Function of Ensemble Size and Spatial Scale in a Convection-Allowing Ensemble. *Mon. Wea. Rev.*, **139**, 1410–1418.

Delle Monache, L., F.A. Eckel, and D.L. Rife, 2013: Probabilistic weather prediction with an analog ensemble. *Mon. Wea. Rev*., **141**, 3498-3516.

___, L., and R.B. Stull, 2003: An ensemble air-quality forecast over western Europe during an ozone episode. *Atmospheric Environment*, **37**, 3469-3474.

Deng, A., N. L. Seaman, G. K. Hunter, and D. R. Stauffer, 2004: Evaluation of interregional transport using the MM5-SCIPUFFsystem. *J. Appl. Meteor*., **43**, 1864–1886.

Doyle, J.D., and D.R. Durran, 2002: The dynamics of mountain-wave-induced rotors. *Journal of the Atmospheric Sciences*, **59**, 186-201.

Draxler, R.R., 2003: Evaluation of an ensemble dispersion calculation. *J. of Appl. Meteor.*, **42**, 308-317.

Du, J., G. DiMego, B. Zhou, D. Jovic, B. Ferrier, B. Yang, and S. Benjamin, 2014: NCEP Regional Ensembles: Evolving toward hourly-updated convection-allowing scale and storm-scale predictions within a unified regional modeling system. 22nd Conf. on Numerical Weather Prediction and 26th Conf. on Weather Analysis and Forecasting, Atlanta, GA, Amer. Meteor. Soc., J1.4. [Available online at https://ams.confex.com/ams/94Annual/webprogram/Manuscript/Paper239030/NCEP%20Regional%20Ensembles.pdf]

___, J., S.L. Mullen, and S. Frederick, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev*. **125**, 2427-2459.

Dudhia, J, 1996: A multi-layer soil temperature model for MM5. *Sixth Annual PSU/NCAR Mesoscale Model Users' Workshop*, Boulder CO, July 1996, 49-50.

___, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077-3107.

Durran, D.R., 1990: Mountain waves and downslope winds. *Atmospheric Processes Over Complex Terrain*, **23**, 59-83.

Eckel, F. A., H. R. Glahn, T. M. Hamill, S. Joslyn, W. M. Lapenta, and C. F. Mass, 2010: National mesoscale probabilistic prediction: Status and the way forward. *National Workshop on Mesoscale Probabilistic Prediction*, Boulder, CO, National Center for Atmospheric Research. [Available online at www.weather.gov/ost/NMPP_white_paper_28May10_with%20sig%20page.pdf.]

Efron, B., 1979: Bootstrap method: another look at the jackknife. *The Annals of Statistics*, **7**, 1-26.

Etling, D., 1990: On plume meandering under stable stratification. *Atmospheric Environment*, **24A**, 1979-1985.

Friedl, M.A., D.K. McIver, J.C. Hodges, X.Y. Zhang, D. Muchoney, A.H. Strahler, et al., 2002:

Global land cover mapping from MODIS: Algorithms and early results. *Remote Sensing of Environment*, **83**, 287-302.

Gneiting, T., A.E. Raftery, A.H. Westveld, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, **133**, 1098-1118.

___, and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. *J. Amer. Stat. Assoc.*, **102**, 359-378.

Grimit, E.P., and C.F. Mass, 2002: Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Wea. Forecasting.*, **17**, 192-205.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155-167.

___, T.M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550-560.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559-570.

Hertenstein, R.F., and J.P. Kuettner, 2005: Rotor types associated with steep lee topography: influence of the wind profile. *Tellus, **57A**, 117-135.*

Hong, S.Y., J. Dudhia, and S.H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.*, **132**, 103-120.

___, S.Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318-2341.

___, S.Y., 2010: A new stable boundary-layer mixing scheme and its impact on the simulated East Asian summer monsoon. *Quart. J. Roy. Meteor. Soc.*, **136**, 1481-1495.

Hoover, J., D. Stauffer, S. Richardson, L. Mahrt, B. Gaudet, and A. Suarez, 2015: Submeso motions within the stable boundary layer and their relationship to local indicators and synoptic regime in moderately complex terrain. *J. of Appl. Meteor. and Climatol.*, DOI: 10.1175/JAMC-D-14-0128.1

Houtekamer, P.L., and H. Mitchell, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **131**, 3269-3289.

___, and J. Derome, 1995: Methods for ensemble prediction. *Mon. Wea. Rev.*, **123**, 2181-2196.

Hu, X., 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. *Atmospheres,* **118**, 10490-10505.

Janjic, Z.I., 1994: The step-mountain eta coordinate model: Further developments of the

convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927-945.

___, 1996: The surface layer in the NCEP Eta Model, Eleventh Conference on Numerical Weather Prediction, Norfolk, VA, 19–23 August; Amer. Meteor. Soc., Boston, MA, 354–355.

Jimenez, P.A., J. Dudhia, J.F. Gonzalez-Rouco, J. Navarro, J.P. Montalvez, and E. Garcia-Bustamante, 2012: A revised scheme for the WRF surface layer formulation. *Mon. Wea. Rev.*, DOI:10.1175/MWR-D-11-00056.1.

Jimenez-Guerrero, P., S. Jerez, J.J. Gomez-Navarro, R. Lorente, and J.P. Montavez, 2014: Future air pollution in Europe from a multi-physics ensemble of climate change air quality projections. *Handbook of NATO Science for Peace and Security Series C: Environmental Security*, D.G. Steyn, P.J.H. Builtjes, and R.M.A. Timmermans, Eds., Springer Science+Business Media Dordrecht, 3-7.

Jirak, I.L., S.J. Weiss, and C.J. Melick, 2012: The SPC storm-scale ensemble of opportunity: Overview and results from the 2012 hazardous weather testbed spring forecasting experiment. *26th Conf. on Severe Local Storms*, Nashville, TN, Amer. Meteor. Soc., P9.137. [Available online at https://ams.confex.com/ams/26SLS/webprogram/Paper211729.html]

Kain, J., 2004: The Kain–Fritsch convective parameterization: an update. *J. of Appl. Meteor.*, **43**, 170-181.

Kolczynski, W.C., D.R. Stauffer, S.E. Haupt, and A. Deng, 2009: Ensemble variance calibration for representing meteorological uncertainty for atmospheric transport and dispersion modeling. *J. Appl. Meteor. Climatol.*, **48**, 2001-2021.

___, D.R. Stauffer, and S.E. Haupt, 2011: Investigation of ensemble variance as a measure of true forecast variance. *Mon. Wea. Review*, **139**, 3954-3963.

Kuettner, J., 1959: The rotor flow in the lee of mountains. Geophysics Research Directorate (GRD) Research Notes 6, AFCRC-TN-58-626, Air Force Cambridge Research Center, USA, 20 pp.

Lee, J.A., L.J. Peltier, S.E. Haupt, J.C. Wyngaard, D.R. Stauffer, and A. Deng, 2009: Improving SCIPUFF dispersion forecasts with NWP ensembles. *J. Appl. Meteor. Climatol.*, **48**, 2305-2319.

___, J.A., W.C. Kolczynski, T.C. McCandless, and S.E. Haupt, 2012: An objective methodology for configuring and down-selecting an NWP ensemble for low-level wind prediction. *Mon. Wea. Rev.*, **140**, 2270-2286.

Lei, L., D.R. Stauffer, and A. Deng, 2012: A hybrid nudging-ensemble Kalman filter approach to data assimilation in WRF/DART. *Quart. J. Roy. Meteor. Soc.*, **138**, 2066-2078.

Leith, C.E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409-418.

Leutbecher, M., and T.N. Palmer, 2007: Ensemble forecasting. *Journal of Computational Physics*, **227**, 3515-3539.

Lewellen, W.S., and R.I. Sykes, 1989: Meteorological data needs for modeling air quality uncertainties. *J. Atmos. Oceanic Technol.*, **6**, 759-768.

Ma, J., Y. Zhu, D. Hou, X. Zhou, and M. Peña, 2014: Ensemble transform with 3D rescaling initialization method. *Mon. Wea. Rev*. doi:10.1175/MWR-D-13-00367.1, in press.

Mahrt, L., 1999: Stratified atmospheric boundary layers. *Boundary-Layer Meteorology*, **90**, 375-396.

___, and D. Vickers, 2002: Contrasting vertical structures of nocturnal boundary layers. *Boundary-Layer Meteorology*, **105**, 351-363.

___, 2007: Weak-wind mesoscale meandering in the nocturnal boundary layer. *Environmental Fluid Mechanics*, **7**, 331-347.

___, 2014: Stably stratified atmospheric boundary layers. *Annual Review of Fluid Mechanics*, **46**, 23-45.

Markowski, P., and Y. Richardson 2010: Mountain Waves and Downslope Windstorms. *Mesoscale Meteorology in Midlatitudes*, P. Inness and W. Beasley, Eds., Wiley-Blackwell, 327-342.

Marsigli, C., A. Montani, F. Nerozzi, T. Paccagnella, S. Tibaldi, F. Molteni, and R. Buizza, 2001: A strategy for high-resolution ensemble prediction II: Limited-area experiments in four Alpine flood events. *Quart. J. Roy. Meteor. Soc.*, **127**, 2095-2115.

Mass, C., and Coauthors, 2009: PROBCAST: A web-based portal to mesoscale probabilistic forecasts. *Bull. Amer. Meteor. Soc.*, **90**, 1009-1014.

Mass, C.F., D. Ovens, K. Westrick, and B.A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts?. *Bull. Amer. Meteor. Soc.*, **83**, 407–430.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343-360.

Mlawer, E.J., S.J. Taubman, P.D. Brown, M.J. Iacono, and S.A. Clough, 1997: Radiative transfer for in-homogeneous atmosphere: RRTM, a validate correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16663-16682.

Molteni, F., R. Buizza, C. Marsigli, A. Montani, F. Nerozzi, and T. Paccagnella, 2001: A strategy for high-resolution ensemble prediction I: Definition of representative members and global model experiments. *Quart. J. Roy. Meteor. Soc.*, **127**, 2069-2094.

___, R. Buizza, T.N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73-119.

Munsell, E.B., and F. Zhang, 2014: Prediction and uncertainty of Hurricane Sandy (2012) explored through a real-time cloud-permitting ensemble analysis and forecast system assimilating airborne Doppler radar observations. *Journal of Advances in Modeling Earth Systems*, **6**, 1-20.

Murphy, J.M., 1988: The impact of ensemble forecasts on predictability. *Quart. J. Roy. Meteor. Soc.*, **114**, 463-493.

Mylne, K., 2013: Scientific framework for the ensemble prediction system for the UKV. Preprints, *MOSAC and SRG Meeting 2013*, Exeter, Devon, UK, Met Office, 18.6. [Available online at http://www.metoffice.gov.uk/media/pdf/q/0/MOSAC_18.6_Mylne.pdf]

Nakanishi, M. and H. Niino, 2004: An improved Mellor-Yamada level-3 model with condensation physics: Its design and verification. *Bound.-Layer Meteor.*, **112**, 1-31.

Nappo, C., 1991: Sporadic breakdowns of stability in the PBL over simple and complex terrain. *Boundary-Layer Meteorology*, **54**, 69-87.

Raftery A.E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea, Rev.*, **133**, 1155–1174.

Rao, K. S., 2005: Uncertainty analysis in atmospheric dispersion modeling. *Pure Appl. Geophys.*, **162**, 1893–1917.

Romine, G.S., C.S. Schwartz, J. Berner, K.R. Fossell, C. Snyder, J.L. Anderson, and M.L. Weisman, 2014: Representing forecast error in a convection-permitting ensemble system. *Mon. Wea. Rev.*, **342**, 4519-4541.

Salmond, J.A., and I.G. McKendry, 2005: A review of turbulence in the very stable nocturnal boundary layer and its implications for air quality. *Progress in Physical Geography*, **29**, 171-188.

Scorer, R.S., 1949: Theory of waves in the lee of mountains. *Quart. J. Roy. Meteor. Soc.*, **75**, 41-56.

Seaman, N.L., B.J. Gaudet, D. R. Stauffer, L. Mahrt, S.J. Richardson, J.R. Zielonka, and J.C. Wyngaard, 2012: Numerical prediction of the submesoscale flow in the nocturnal boundary layer over complex terrain. *Mon. Wea. Rev.*, **140**, 956-997.

Shin, H. H., and S. Hong, 2011: Intercomparison of planetary boundary-layer parameterizations in the WRF model for a single day from CASES-99. *Boundary Layer Meteorology*, **139**, 261–281.

Skamarock, W. C., and Coauthors, 2008: A description of the advanced research WRF version 3. *NCAR Tech. Note.* NCAR/TN-475+STR, 113 pp.

Stauffer, D.R., 2013: Uncertainty in Environmental NWP Modeling. *Handbook of Environmental Fluid Dynamics, Volume 2*, H.J.S. Fernando, Ed., CRC Press/Taylor & Francis Group, 411-424.

___, D.R., and N.L. Seaman, 1994: Multiscale four-dimensional data assimilation. *J. Appl. Meteor*., **33**, 416-434.

Stensrud, D.J., J. Bao, and T.T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077-2107.

___, and J.M. Fritsch, 1994: Mesoscale convective systems in weakly forced large-scale environments. Part III: Numerical simulations and implications for operational forecasting. *Mon. Wea. Rev.*, **122**, 2084-2104.

Stull, R.B., 1988: *An Introduction to Boundary Layer Meteorology*, Kluwer Academic Publishers, 666 pp.

Sukoriansky, S., B. Galpering, and V. Perov, 2005: Application of a new spectral theory of stable stratified turbulence to the atmospheric boundary layer over sea ice. *Bound.-Layer Meteor*., **117**, 231-257.

Sukoriansky, S., 2008: Implementation of the Quasi-Normal Scale Elimination (QNSE) model of stably stratified turbulence in WRF. *Report on WRF-DTC Visit*, Developmental Testbed Center, [Available online at http://www.dtcenter.org/visitors/reports_07/Sukoriansky_report.pdf]

Sun, J., and Coauthors, 2015: Review of wave-turbulence interactions in the stable boundary layer. *Submitted to Reviews of Geophysics*.

Sykes, I. R., S. F. Parker, D. S. Henn, and B. Chowdhury, 2006: SCIPUFF Version 2.2 Technical Documentation, 318 pp.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.

Wang, X., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble-variational hybrid data assimilation for NCEP Global Forecast System: Single resolution experiments. *Mon. Wea. Rev.,* **141**, 4098-4117.

Warner, T.T., R. Sheu, J.F. Bowers, R.I. Sykes, G.C. Dodd, and D.S. Henn, 2002: Ensemble simulations with coupled atmospheric dynamic and dispersion models: Illustrating uncertainties in dosage simulations. *J. of Appl. Meteor.,* **41**, 488-504.

Wilks, D.S., 2011a: On the reliability of the rank histogram. *Mon. Wea. Rev.*, **139**, 311-316.

___, 2011b: *Statistical Methods in the Atmospheric Sciences*, Academic Press, 704 pp.

Xue, M., F. Kong, K.W. Thomas, J. Gao, Y. Wang, K. Brewster, and K.K. Droegemeier, 2013: Prediction of convective storms at convective-resolving 1km resolution over continental United States with radar data assimilation: An example case of 26 May 2008 and precipitation forecasts from spring 2009. *Advances in Meteorology,* **2013**, dx.doi.org/10.1155/2013/259052