

# 80 A SUB-KM-GRID ENSEMBLE FOR REPRESENTING MESO-GAMMA HAZARD-PREDICTION UNCERTAINTY IN COMPLEX TERRAIN

Eric B. Wendoloski\*, David R. Stauffer, Astrid Suarez, Glenn K. Hunter  
*Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania*

## 1. INTRODUCTION

Mountain wave activity within the stable boundary layer (SBL) has a large impact on circulations that directly influence the outcome of atmospheric transport and dispersion (AT&D) in the event of a release. Mesoscale terrain-induced mountain wave activity is highly sensitive to changes in background winds or static stability. Varying initial conditions and parameterizations of surface layer and planetary boundary layer (PBL) physics within weather models leads to unique patterns in modeled mountain wave activity and thus uncertainty in AT&D predictions. A twelve member sub-km-grid ensemble for evaluating this AT&D uncertainty within the SBL over complex terrain is described. Ensemble spread is achieved through diversity in initial conditions and PBL/surface layer physics within the Weather Research and Forecasting (WRF) model (Skamarock et al. 2008). Ensemble member output is used as input to the Lagrangian transport and diffusion Second Order Closure Integrated Puff (SCIPUFF) model for hazard prediction (Sykes et al. 2006). In addition, the use of the ensemble mean or a single best member (the member whose vector wind difference against the ensemble mean or observations is smallest) capable of utilizing ensemble wind field uncertainty statistics (SCIPUFF hazard mode) is explored as a less computationally costly approach to quantify AT&D uncertainty.

Ensemble performance is evaluated for a case study using probabilistic verification techniques including the Continuous Ranked Probability Score (CRPS) and rank histograms (Wilks 2011). Performance statistics are calculated using observations from the Rock Springs, PA network along Tussey Ridge within the Nittany Valley of central Pennsylvania (Hoover et al. 2014). Evaluation results from ensemble configurations using 1.333-km horizontal grid spacing (referred to as the 1.3-km ensemble) and 0.444-km horizontal grid spacing (referred to as the 0.4-km ensemble) are compared. Confidence intervals are determined using a bootstrap technique in order to test for significant differences in performance between the two ensemble configurations (Efron 1979; DiCiccio and Efron 1996; Candille et al. 2007).

Two additional experiments are conducted to examine ensemble performance for multi-initialization versus multi-physics ensemble configurations. CRPS values of four three-member multi-initialization ensembles holding PBL/surface layer physics constant are compared and, the same type of comparison is performed for three four-member multi-physics ensembles holding

initial conditions constant.

## 2. MODEL DESCRIPTION AND ENSEMBLE CONFIGURATION

The advanced research WRF (WRF-ARW; Skamarock et al. 2008) version 3.6 is configured with four one-way nested domains with 12-km, 4-km, 1.333-km, and 0.444-km horizontal grid spacing. Because the SBL is often only a few tens of meters deep, fine vertical resolution is necessary to better resolve low-level nocturnal wind speeds and valley cold pool development (Seaman et al. 2012). Thus, each nest includes 44 vertical levels with 10 levels in the lowest 50 m and 2-m spacing in the lowest 10 m. Initial and lateral boundary conditions are provided every 6 h by NCEP 0.5°x0.5° Global Forecast System (GFS) analyses and forecasts. The Rapid Radiative Transfer Model (RRTM; Mlawer et al. 1997) longwave and Dudhia shortwave (Dudhia 1989) atmospheric radiation physics parameterizations are used. The Kain-Fritsch cumulus parameterization (12-km domain only; Kain 2004), 3-class simple ice microphysics (Hong et al. 2004), and Noah land surface model (Chen and Dudhia 2001) with MODIS land use are also employed (Friedl et al. 2002).

Ensemble diversity is created by using different initial conditions and PBL physics schemes with corresponding surface layer physics schemes. Initial condition diversity is given by Control (CTRL), a 12-h cold-start forecast, Baseline (BSL), a 24-h cold-start forecast initialized 12 h prior to the forecast period of interest, and Four Dimensional Data Assimilation (FDDA), a 12-h forecast based on a 12-h pre-forecast which utilizes analysis and observation nudging (e.g., Rogers et al. 2013; Lei et al. 2012). PBL/surface layer scheme diversity includes the Mellor-Yamada-Janjic (MYJ) with Eta surface layer (Janjic 1994), the Mellor-Yamada Nakanishi Niino (MYNN; Nakanishi and Niino 2004) with Eta surface layer, the Quasi-Normal Scale Elimination (QNSE; Sukoriansky et al. 2005) with QNSE surface layer, and the Yonsei University (YSU; Hong et al. 2006) with revised MM5 surface layer (Jimenez et al. 2012). Each PBL scheme employs different treatments of turbulence closure and PBL mixing. Table 1 presents the pairings of initial conditions and PBL/surface layer physics providing a total of 12 ensemble members.

## 3. NITTANY VALLEY AND ROCK SPRINGS OBSERVATION NETWORK

The 0.4-km nested domain is centered over Nittany Valley in central PA. The valley is approximately 20 km wide and is bordered by the Allegheny Mts. (~350 m AGL) to the northwest and Tussey Ridge (~300 m AGL) to the southeast (Fig. 1a). A special observing network

---

\*Corresponding author address: Eric B. Wendoloski, 623 Walker Building; University Park, PA 16802-5013. Email: ebw126@psu.edu

located in Rock Springs, PA is situated along Tussey Ridge (Fig. 1b). The 2-m temperature and wind observations from sites 3, 6, 7, 8, 9, and 12 are utilized for ensemble performance evaluation. In addition to thermistors and 2-D or 3-D sonic anemometers at the sites, a sonic detection and ranging instrument (SODAR) is located within the network near the base of Tussey Ridge.

#### 4. CASE STUDY: 16 SEPTEMBER 2011

A case study is performed for the night of 16 September 2011 from 0000-1200 UTC (SEP16). SEP16 is characterized by weakening pressure gradients and low-level wind speeds as an area of high pressure moved in from the west. Skies were clear with northwesterly surface wind (perpendicular to the Allegheny Mts.) over western NY and central PA. Trapped-lee waves excited by the Allegheny Mts. due to wind ducting regions persisted throughout the night. SODAR 2027 observations from 0600-0800 UTC demonstrate northwesterly to southeasterly reversal flow from ~160 m to ~30 m AGL and minimum wind speed values between the regions of reversed flow (Fig. 2). This flow pattern indicates the presence of a rotor circulation with reversal flow below 100 m and perpendicular to Tussey Ridge into Nittany Valley.

The SEP16 case study focuses on a simulated release from Site 9 in the Rock Springs network from 0600-0800 UTC to demonstrate the broad range of AT&D outcomes exhibited by explicit ensemble members during this time frame. SCIPUFF surface dosage plumes for a Site 9, 3-m, 12-min continuous release from 0600-0800 UTC of the passive tracer  $C_7F_{14}$  for each explicit member are plotted in Fig. 3. In most members, the dosage plumes move southeast across Tussey Ridge; however the dosage plumes of several members deviate against the mean flow from the northwest and track along and into Nittany Valley. Corresponding vertical trajectory cross sections are plotted in Fig. 4 for a 3-m AGL release from nine grid cells (a 3x3 cell region surrounding Site 9). The cross section location is depicted by the dashed line in Fig. 1a. In several members, it is clear that some of the particle trajectories are influenced by a circulation along Tussey Ridge and within Nittany Valley (e.g., QNSECTRL). Surface dosage values and the spatial extent of the surface dosage plumes are closely related to the variations in transport associated with the rotor-like circulation resolved by the model. From a hazard prediction standpoint, wind reversal regions and along valley flow associated with rotor circulations are both very important for surface dosage prediction. Predicting the correct location and timing of these rotor motions is especially difficult. Therefore, a 12-member ensemble appears capable of providing reasonable spread (among AT&D outcomes) important for quantifying forecast uncertainty for a chem-bio release.

Explicit ensemble AT&D surface dosage probabilities (an aggregate of dispersion results from the explicit members) are compared to those derived from a single-member approach (ensemble mean or best member)

that is capable of utilizing ensemble-based wind field uncertainty in SCIPUFF hazard mode. Hazard mode employs the ensemble-based wind field uncertainty, single-point U wind variance, V wind variance, and UV wind covariance to compute plume dispersion. Figure 5a depicts explicit ensemble surface dosage probability for a threshold value of  $10^{-9} \text{ m}^3 \cdot \text{sm}^{-3}$  following a Site 9 release integrated from 0600-0800 UTC. Figure 5b depicts the corresponding probabilities derived from a clipped-normal distribution for the ensemble mean. It is noted that the single-member approach missed the western extent of the 5% probability contour in Fig. 5a. Only one member, the YSU-FDDA, had spatial plume extent in the western half of the domain. The single-member surface dosage probabilities do not exactly match those of the explicit ensemble, although the single-member hazard-mode approach gives a clear indication of the overall range of potential outcomes and the spatial area most likely to be impacted by a chem-bio release at a much reduced cost. Although the explicit ensemble probabilities provide greater detail, the computationally efficient single-member approach could prove valuable in a time sensitive situation.

#### 5. ENSEMBLE PERFORMANCE AND EVALUATION TECHNIQUES

Ensemble performance is evaluated using probabilistic verification techniques including the rank histogram and the Continuous Ranked Probability Score (CRPS; Wilks 2011). A bootstrap technique is employed to test for significant differences between two ensemble configurations (Efron 1979; DiCiccio and Efron 1996; Candille et al. 2007).

##### 5.1 RANK HISTOGRAMS

Rank histograms are used to quantify the reliability and spread of an ensemble prediction system. A reliable ensemble forecast probability is equal to the true probability of the event. If an ensemble forecast is reliable, the observation should be equally likely to fall in 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 a forecast of a variable over some spatial and temporal interval are combined. 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 values in the middle of a rank histogram. This behavior is caused by verification values that are infrequently extreme in comparison to the 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 (the CRPS in this case; Hamill 2001; Wilks 2011).

## 5.2 CONTINUOUS RANKED PROBABILITY SCORE

The Continuous Ranked Probability Score (CRPS) is a probabilistic score that quantifies the difference between the ensemble forecast cumulative distribution function (CDF) and the CDF of the corresponding observations. A perfect CRPS of zero is achieved when every ensemble forecast value is equal to the observed value. As shown by Hersbach (2000), the CRPS can be decomposed into three components so that:

$$CRPS = Reliability - Resolution + Uncertainty .$$

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 event forecast, and the value of uncertainty is proportional to the standard deviation of the observations. The potential CRPS is a value that combines the resolution and uncertainty terms into a single term where smaller values are optimal. Ideally, the ensemble forecast system strikes an optimal balance between reliability and resolution (e.g., an ensemble with a high value of resolution but a large reliability value refers to a specific (or sharp) forecast but for the wrong value).

## 5.3 BOOTSTRAPPING AND CONFIDENCE INTERVALS

Bootstrapping is a technique used to create a substitute population that approximately represents the sampling distribution of a summary statistic of interest. Data is resampled with replacement and the statistic of interest for the new sample is then calculated. This process is generally repeated thousands of times to create a distribution of sample statistics known as a bootstrap distribution (Efron 1979). Percentiles of the bootstrap distribution are used to define confidence intervals that correspond to those of the unknown parent population. This is known as the percentile method. The bias corrected and accelerated (BCa) correction to the percentile method adjusts percentiles of interest based on the bias and skew of the bootstrap distribution (DiCiccio and Efron 1996). This technique is used in the next section to calculate significance of the differences in CRPS, reliability, and potential CRPS between two ensemble configurations (Candille et al. 2007).

The methodology behind this approach is based on Candille et al. (2007). Data is resampled 10,000 times to create 10,000 bootstrap samples. CRPS is calculated for each of these new samples for two separate ensembles. CRPS from ensemble B is subtracted from ensemble A where a positive value indicates an improvement by ensemble B. A distribution of CRPS differences is created. 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 this comparison of two ensembles 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.

## 6. PERFORMANCE RESULTS

The 12-member configurations of the 1.3-km and 0.4-km ensembles are evaluated and compared. The low-level temperature and wind forecasts at 2 m AGL are evaluated over six Rock Springs sites as proxy to actual AT&D surface tracer measurements. U and V wind components have been rotated to an “along-valley” component parallel to Tussey Ridge and an “across-valley” component perpendicular to Tussey Ridge and parallel to rotor circulation reversal flow. Figure 6a depicts across-valley wind CRPS and the corresponding reliability component calculated every 12 min from 0000-1200 UTC for SEP16 for the 1.3-km ensemble. Observed, ensemble mean, and best member values are also plotted over the same time period. Large spikes in 12-min CRPS and reliability are apparent in Fig. 6a where ensemble values deviate largely from observed (e.g., ~ 0630 UTC). Additionally, ensemble values are greater than observed throughout the night indicating a strong positive across-valley wind bias in the 1.3-km ensemble forecasts. Figure 6b depicts the same information but for the 0.4-km ensemble. Examination of Fig. 6b reveals that the magnitude of spikes in 12-min CRPS is reduced and reliability values tend closer to zero throughout the night. Ensemble values produced by the sub-km grid are much closer to those observed indicating that the positive bias has decreased.

Table 2 summarizes “nightly” CRPS, reliability, and potential CRPS values calculated over the full nighttime period (0000-1200 UTC) for each variable of interest along with percent improvement (or skill score) where a positive percent improvement indicates better performance by the 0.4-km ensemble when compared to the 1.3-km ensemble. The significance of each result is also reported. Significant improvements in nightly CRPS and reliability are noted for 2-m temperature with no significant change in resolution (potential CRPS). Nightly CRPS, reliability, and resolution all show significant improvements for both 2-m wind speed and across-valley wind forecasts. The significant improvement in across-valley wind forecasts (Fig. 7) is mainly due to a large significant improvement in reliability due to the large reduction in bias discussed previously. This result is evidenced in the rank histograms for the 1.3-km ensemble (Fig. 8a) and the 0.4-km ensemble (Fig. 8b). The large frequency

noted in rank “1” (Fig. 8a) indicates that all across-valley wind ensemble forecast values were greater than observed over 40% of the time. That bias is dramatically reduced in Fig. 8b which tends more towards rank uniformity indicated by the dashed line. Results for along-valley wind show a statistically significant degradation in CRPS and resolution. This degradation is likely due to phase errors in the 0.4-km ensemble better capturing reversals in along-valley flow related to the rotor circulation, a phenomenon important to AT&D outcomes largely unresolved by the 1.3-km ensemble (time series not shown). Finer horizontal grid spacing appears necessary to better resolve fluctuations in surface temperature, wind speed, and cross-valley surface wind components related to variability associated with the rotor circulation.

Four three-member multi-initialization ensembles each holding separate PBL/surface layer physics constant are compared. Calculating CRPS for temperature, wind speed, across-valley, and along-valley wind using each of the four ensemble configurations did not reveal advantageous performance for any one PBL/surface layer physics parameterization (not shown). The same comparison is performed using three four-member multi-physics ensembles that hold the initialization strategy constant. Figure 9 reveals CRPS values for multi-physics ensembles holding CTRL, BSL, and FDDA initial conditions constant. The FDDA multi-physics ensemble demonstrates superior probabilistic forecast skill for temperature, wind speed, and, most notably, across-valley wind (recall lower values of CRPS depict better performance). Along-valley wind results were comparable for the three ensemble configurations. This result indicates that during the nighttime period (0000 – 1200 UTC) of SEP16, PBL/surface layer physics diversity benefits from data assimilation performed over the preceding daytime period.

## 7. SUMMARY

For the SEP16 case study, a 12-member WRF ensemble with diversity in initial conditions and planetary boundary layer/surface layer physics appears capable of producing reasonable spread in the AT&D outcomes. A single-member approach using SCIPUFF hazard mode is shown to generally cover explicit ensemble spatial spread and capture the value of explicit-member dosage probability. In this case, positive percent improvement of nightly CRPS over six Rock Springs sites for 2-m temperature, wind speed, and across-valley wind demonstrates improved performance for the 0.4-km ensemble when compared to the 1.3-km ensemble. The largest percent improvements are evident for wind speed and across-valley wind. The largest improvement in across-valley wind nightly CRPS is due to a large bias correction that improves the reliability component of the CRPS and causes the corresponding rank histogram to tend much closer to rank uniformity. Sub-km horizontal grid spacing along with fine vertical resolution appears necessary to better resolve across-valley surface wind components related to rotor circulations important for hazard prediction in this case. Additionally, the probabilistic skill of an FDDA multi-

physics ensemble configuration is superior when compared to that of a 12-h cold-start forecast and that of a 24-h cold-start forecast initialized 12 h prior to the nighttime period for SEP16.

## ACKNOWLEDGMENTS

This research is funded by the Defense Threat Reduction Agency Grant No. HDTRA1-10-1-0033 under the supervision of John Hannan. Under the supervision of Walter Bach, SODAR funding was provided by the US Army Research Office under DURIP Grant No. W991NF-10-1-0238.

## 8. REFERENCES

- 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.
- 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.
- DiCiccio, T.J. and B. Efron, 1996: Bootstrap confidence intervals. *Statistical Science*, **11**, 189-228.
- Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077-3107.
- Efron, B., 1979: Bootstrap method: another look at the jackknife. *The Annals of Statistics*, **7**, 1-26.
- 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.
- Hamill, 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 predictions systems. *Wea. Forecasting*, **15**, 559-570.
- 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.
- Hong, 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.

Hoover, J., D. Stauffer, S. Richardson, L. Mahrt, B. Gaudet, and A. Suarez, 2014: Submeso motions within the stable boundary layer and their relationship to local indicators and synoptic regime in moderately complex terrain. *Conditionally accepted to J. of Appl. Meteor. and Climatol.*

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.

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.

Kain, J., 2004: The Kain–Fritsch convective parameterization: an update. *Journal of Applied Meteorology*, **43**, 170-181.

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

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

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.

Rogers, E.R., A. Deng, D.R. Stauffer, B.J. Gaudet, Y.J. Su-Tzai Soong, and S. Tanrikulu, 2013: Application of the Weather Research and Forecasting model for air quality modeling in the San Francisco Bay area. *J. Appl. Meteor. Climatol.*, **52**, 1953-1973.

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. Review*, **140**, 956-997.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2008: A description of the advanced research WRF version 3. *NCAR Tech. Note*. NCAR/TN-475+STR, 125 pp.

Sukoriansky, S., B. Galperin, 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.

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

Wilks, D.S., 2011: *Statistical Methods in the Atmospheric Sciences*, Academic Press, 704 pp.

## 9. TABLES & FIGURES

Table 1. The 12 ensemble members given by initial conditions (IC) and planetary boundary layer (PBL) / 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

Table 2. Summary of nightly 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 0.4-km ensemble (0.4-km) along with percent improvement (%) of the 0.4-km ensemble over the 1.3-km ensemble and significance (Sig.) at the 95% confidence level.

	1.3-km	0.4-km	%	Sig.
<b>CRPS</b>				
Temperature	0.807	0.711	11.9	Y
Wind Speed	0.386	0.251	35.0	Y
Across-Valley	0.752	0.398	47.1	Y
Along-Valley	0.231	0.258	-11.7	Y
<b>Reliability</b>				
Temperature	0.171	0.084	50.9	Y
Wind Speed	0.094	0.025	73.4	Y
Across-Valley	0.339	0.036	89.4	Y
Along-Valley	0.006	0.009	-50.0	N
<b>Potential CRPS</b>				
Temperature	0.635	0.628	1.1	N
Wind Speed	0.293	0.226	22.9	Y
Across Valley	0.413	0.362	12.3	Y
Along Valley	0.225	0.248	-10.2	N

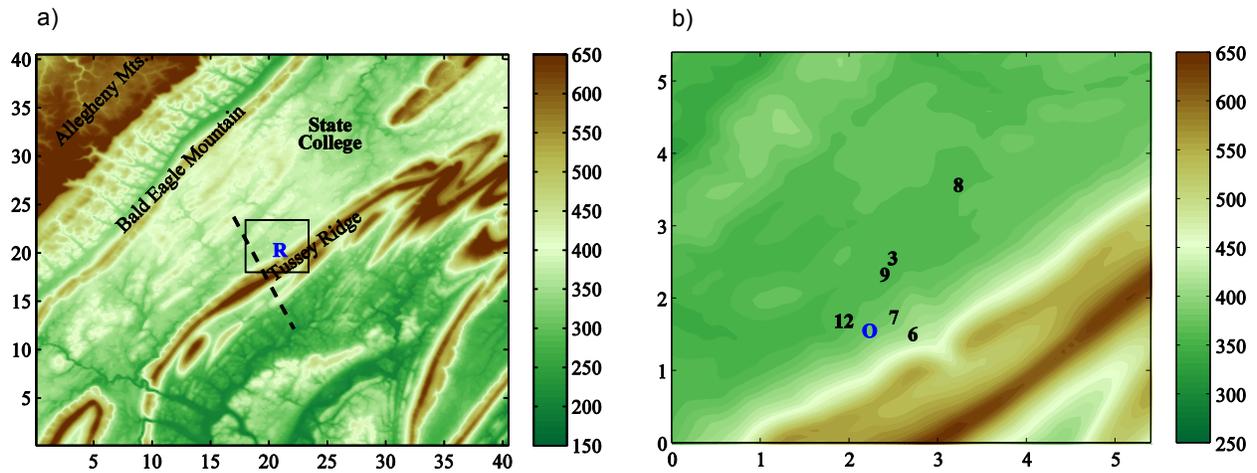


Figure 1. 90-m resolution terrain (MSL) for a) a 40 km by 40 km region containing the Rock Springs network (R) and major topographical features and b) a 5 km by 5 km region denoted by the black square in a) showing the distribution of instrumented sites and SODARs within the observation network. O represents the location of SODAR 2027. The dashed line in a) corresponds to the location of the cross sections in Fig. 4.

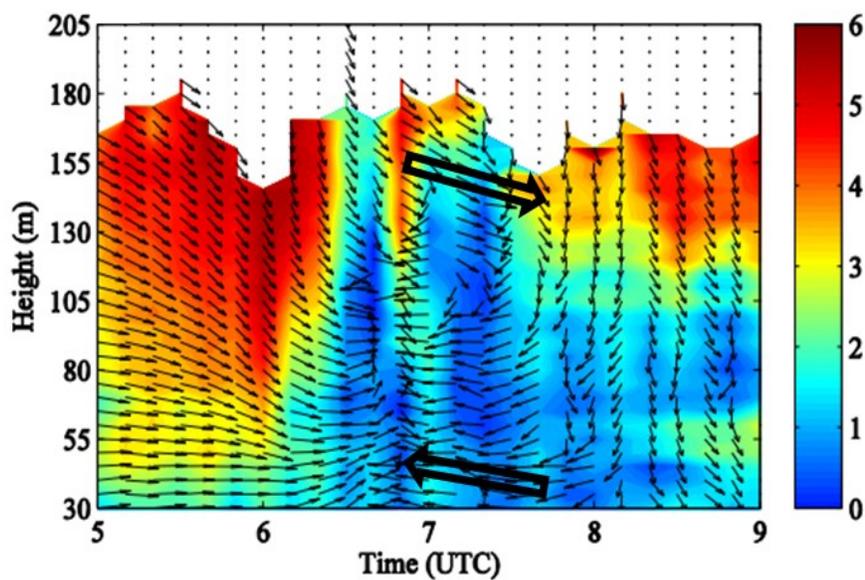


Figure 2. SEP16 SODAR 2027 observed wind speed (color-filled contours) and horizontal wind direction (plan view wind vectors) with height (AGL) from 0500-0900 UTC.

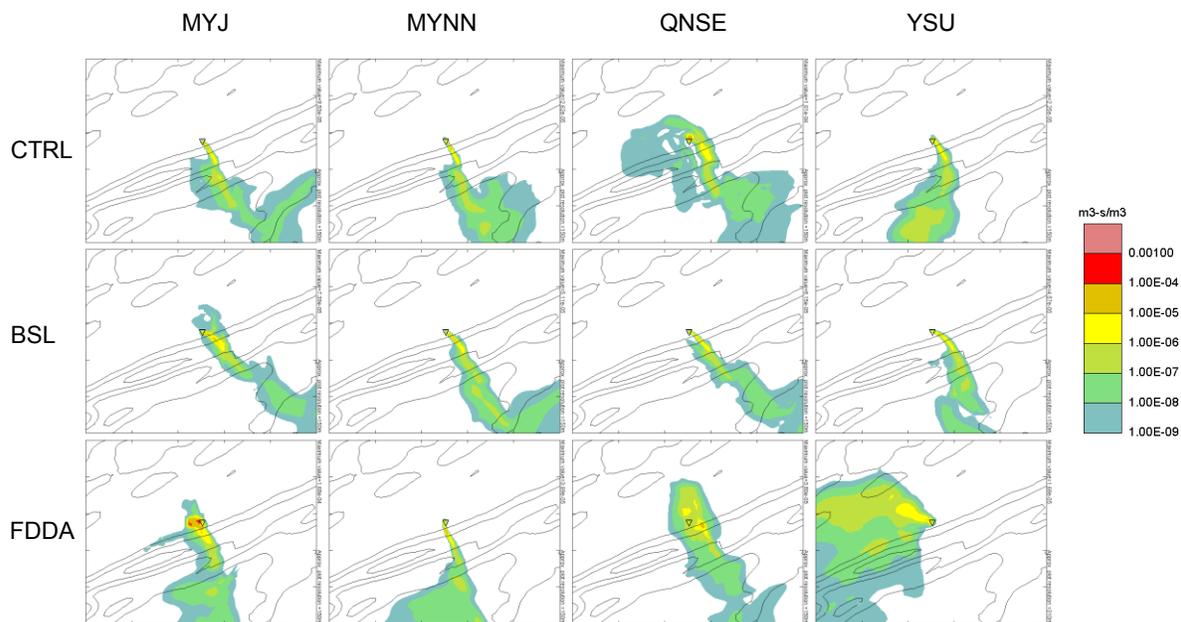


Figure 3. SCIPUFF surface dosages for a 3-m AGL Site 9 release of the passive tracer  $C_7F_{14}$  from 0600-0800 UTC for each of the ensemble members summarized in Table 1.

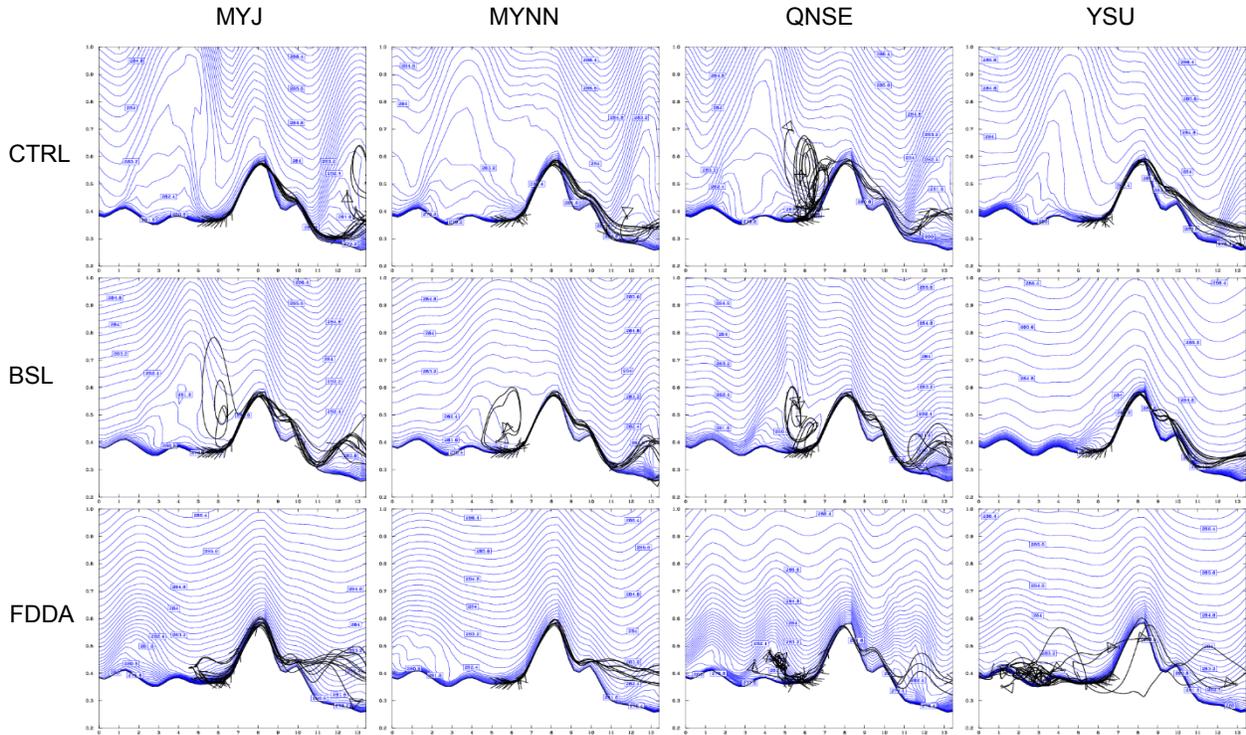


Figure 4. Vertical trajectory cross sections for a 3-m AGL, Site 9 release from 0600-0800 UTC for each of the ensemble members summarized in Table 1. The region of enhanced topography is Tussey Ridge. Contours depict potential temperature (K) as 0800 UTC.

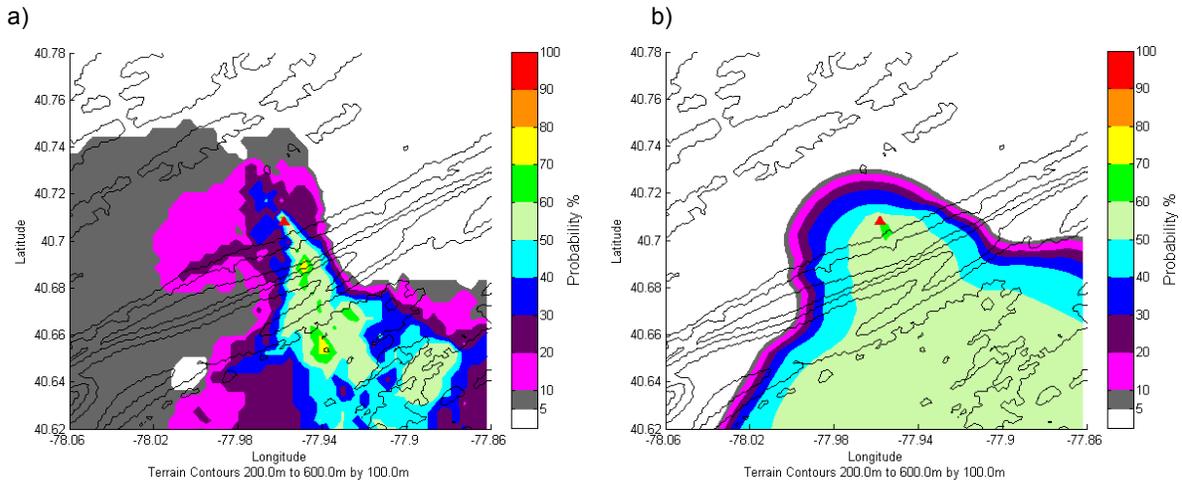


Figure 5. Surface dosage probabilities for a 3-m AGL Site 9 release of the passive tracer  $C_7F_{14}$  for a threshold value of  $10^{-9} \text{ m}^3\text{-sm}^{-3}$  for the a) explicit ensemble and the b) ensemble mean single-member SCIPUFF hazard mode approach.

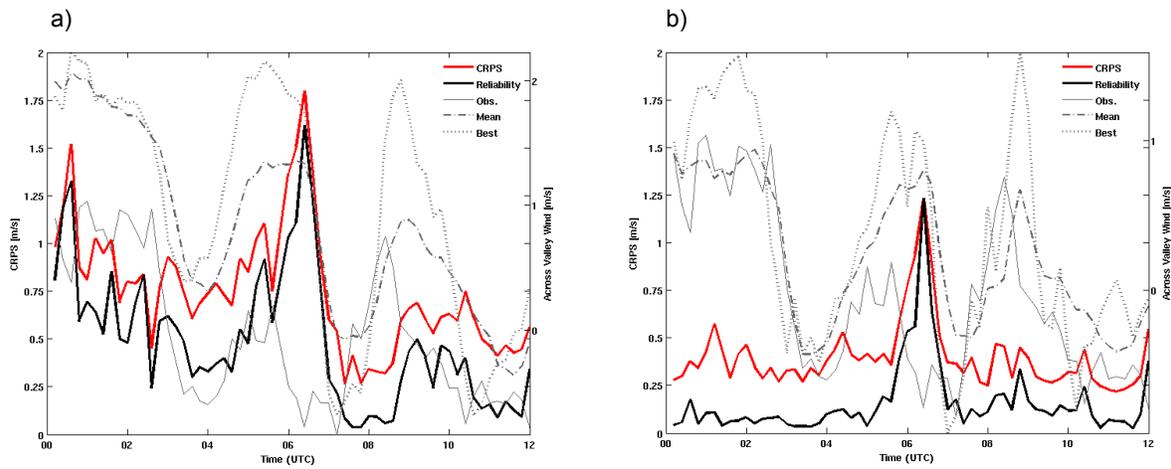


Figure 6. SEP16 0000-1200 UTC 12-min CRPS values for across-valley wind and corresponding observations along with ensemble mean and best member model predictions for the a) 1.3-km ensemble and the b) 0.4-km ensemble.

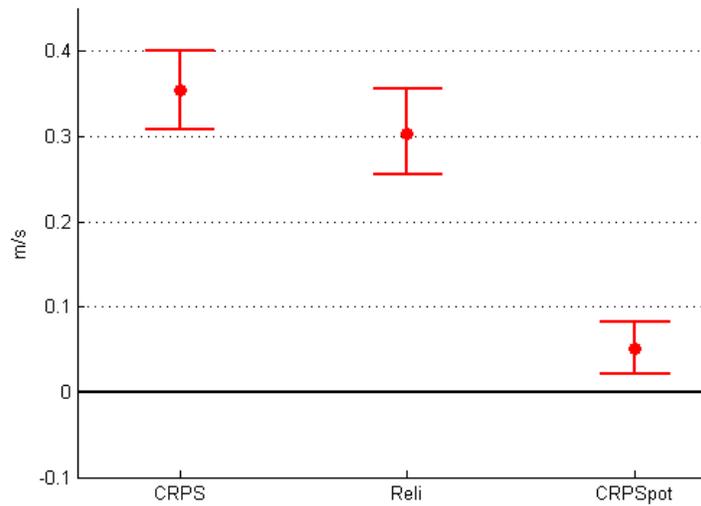


Figure 7. SEP16 difference (1.3-km – 0.4-km ensemble) in across-valley wind nightly CRPS, reliability (Reli), and potential CRPS (CRPSpot) with 95% confidence intervals. Results interpreted as in section 5.3.

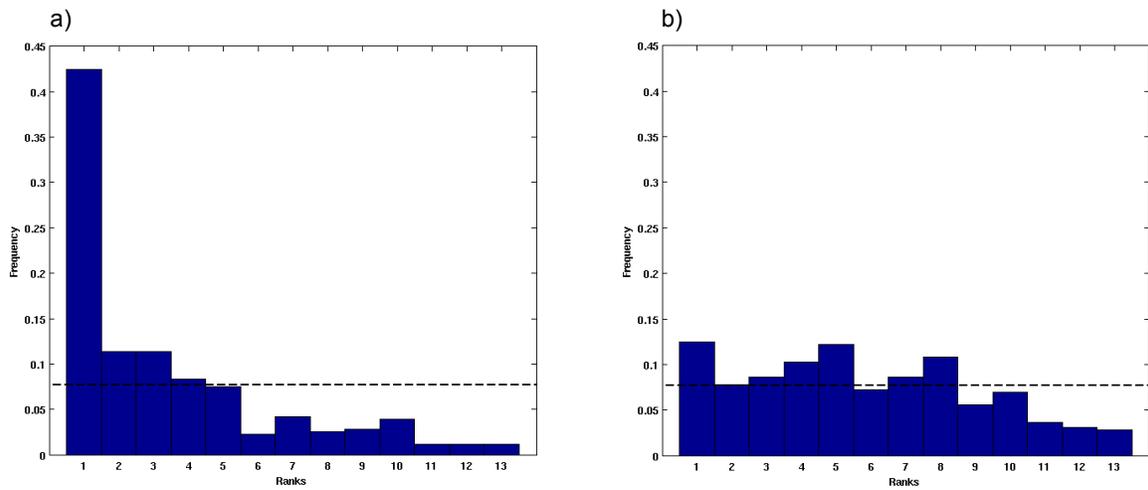


Figure 8. SEP16 nightly (0000-1200 UTC) across-valley wind rank histograms for the a) 1.3-km ensemble and the b) 0.4-km ensemble. A dashed line indicating the frequency required for rank uniformity is plotted for reference.

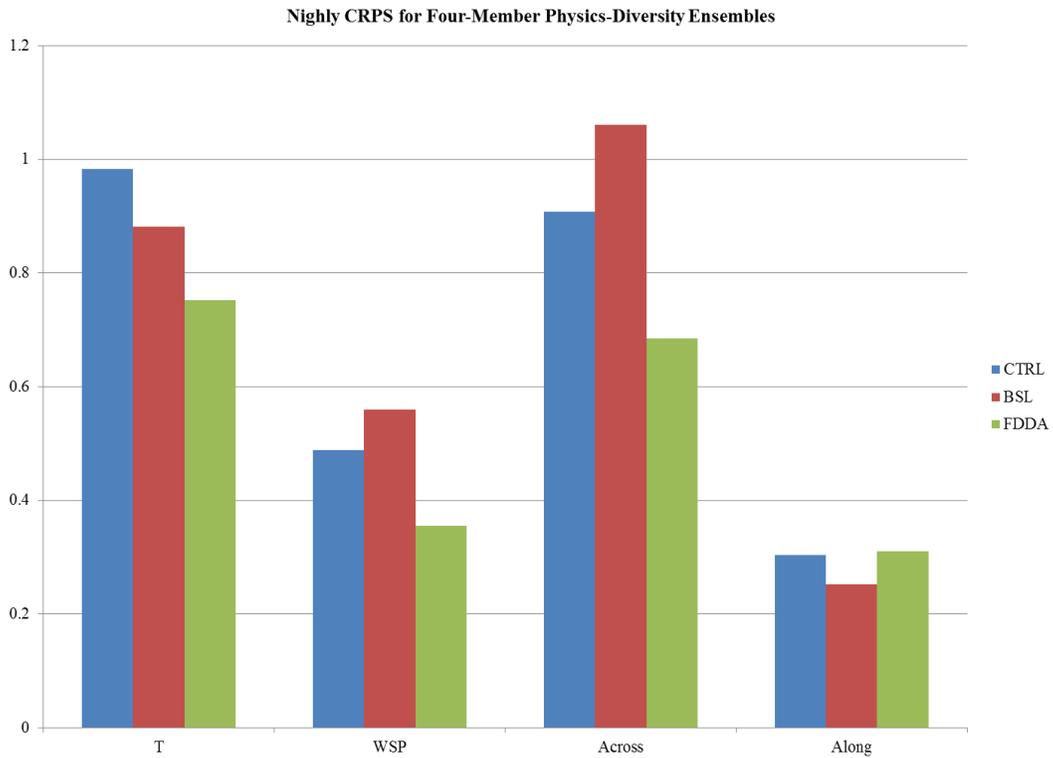


Figure 9. SEP16 nightly (0000-1200 UTC) CRPS for the CTRL multi-physics ensemble (blue), the BSL multi-physics ensemble (red), and the FDDA multi-physics ensemble (green) for 2-m temperature (T), wind speed (WSP), across-valley wind component (Across), and along-valley wind component (Along).