7B.2 A SIMPLE METHOD FOR CALIBRATING ENSEMBLE VARIABILITY TO REPRESENT METEOROLOGICAL MODEL UNCERTAINTY

Walter C. Kolczynski, Jr.*
David R. Stauffer
Sue Ellen Haupt
Pennsylvania State University, State College, PA

1. INTRODUCTION

The uncertainty of meteorological predictions can have a large impact on public safety and monetary interests. Accurate estimates of meteorological forecast uncertainty are important for many wide-ranging applications including: 1) computing hazardous areas using an atmospheric transport and dispersion (AT&D) model following the release of a dangerous airborne substance, 2) performing risk assessment for potential catastrophic weather events, 3) defining confidence intervals for daily temperature forecasts for energy traders. It is vital that models are able to supply / utilize this uncertainty information along with their best predictions (e.g., Tennekes *et al*. 1987, NRC Panel 2003). This paper concentrates on the meteorological uncertainty estimates needed for hazard prediction and consequence assessment (e.g., Stauffer *et al*. 2007), but the techniques are relevant for many applications of meteorological forecasts.

In meteorology, we recognize that our models have finite limits on predictability, due to the nonlinearity of the earth-atmosphere system, our assumptions for model physics and numerics, and the sensitivity to the "under-determinacy" of the initial conditions due to an incompletely observed atmosphere. To replicate this uncertainty, we often create an ensemble meteorological model by slightly changing the initial conditions and/or the model physics with each run. The goal of the ensemble is to span the space of all possible outcomes given the uncertainties in the model and its inputs. In this way, we also try to account for the likelihood of each outcome. An example of such an ensemble on the regional scale is the NCEP/NOAA Short-Range Ensemble Forecast (SREF), which is used in this investigation (McQueen *et al.* 2005).

* *Corresponding Author Address:* Walter C. Kolczynski, Jr. Dept. of Meteorology, Pennsylvania State University, 503 Walker Bldg. State College, PA 16802; wck122@psu.edu.

Although we would like to have a model ensemble large enough to represent completely the probability density function (PDF) of possible forecasts, this is not practical with current computing resources. Therefore, meteorological model ensembles are only a sampling of the full PDF, and any ensemble statistics we generate (such as variance) should be calibrated to more accurately represent the full PDF rather than just the limited set of ensemble members.

Most AT&D models are designed to use outputs from meteorological models to drive their sophisticated dispersion calculations. Ideally, we would like separate AT&D model runs based on the output from each meteorological ensemble member to create an ensemble of dispersion models on which to base our hazard predictions. Unfortunately, in a crisis, where decisions need to be made quickly in order to coordinate emergency personnel response, this is not always possible and only a single run of an AT&D model may be performed. We would, however, still like to use any uncertainty information we can deduce from the meteorological ensemble to help determine our confidence in the output from the AT&D model. Therefore, we need a simple but robust way to translate uncertainty information from the meteorological ensemble into a single AT&D run. This paper explores the applicability of a simple linear calibration to calculate an actual uncertainty based on uncertainty information derived from the meteorological model ensemble. Because the Defense Threat Reduction Agency (DTRA) uses the Second-Order Closure Integrated Puff Model (SCIPUFF, e.g., Deng et al. 2004) in their Hazard Prediction and Assessment Capability tool kit (HPAC) as its primary AT&D model for responding to real and potential threats, we focus on methods that are applicable to this HPAC/SCIPUFF system.

The rest of this paper is divided into five sections. Section 2 describes the details of the model ensemble used in this study and the basics of our linear calibration technique, including our binning method. Section 3 presents and discusses the results when our technique is applied to single-point variances.

The results when the technique is used on two-point covariance information are reviewed in Section 4. Section 5 introduces the use of covariance information to derive SLE, a length scale parameter used by SCIPUFF that is related to the Lagrangian time scale. Section 6 draws some conclusions and outlines future work.

2. METHODOLOGY

For this study, we use output from the SREF for a 22-day period from 25 August to 15 September 2004. At that time, ensemble forecasts were issued twice daily at 09 UTC and 21 UTC with a maximum forecast length of 63 hours. Only the ETA model members of the SREF ensemble are used here to retain model consistency. This choice gives us 44 63–hour forecast sequences of a ten-member ensemble. For verification, we use the 0-hour forecast (analysis) from a control member of the ensemble (ETA-ctl1). This yields five different forecast/verification pairs (every 12 hours from 12-60 hours) for each ensemble set.

Because the primary goal of this study is to produce a practical calibration technique to be used for meteorological data input into SCIPUFF, we examine boundary-layer winds in this analysis. We do not use winds calculated on constant pressure surfaces in order to avoid using fictitious data from underground points when pressure levels intersect terrain. Instead, we use the U- and V-components of wind from the first 30hPa above ground level (AGL). These data are reported in the SREF-ETA data files as 15hPa AGL winds, and are abbreviated as 15hPa AGL U in this paper. These winds represent the average values for a layer of ~300m in depth.

Variances and covariances are used to represent the uncertainty of a meteorological forecast. We use the standard definition of these quantities, so that the ensemble variance *EVar* of a scalar quantity *s* is given by:

$$EVar\ \ s(ij)\ \ =\frac{1}{N}\sum_{m=1}^{N}\ s_m(ij)-\overline{s(ij)}\ ^2 \qquad (1)$$

where *N* is the number of ensemble members, *ij* denotes the value at a single point (i,j), subscript *m* denotes a single ensemble member, and an overbar denotes an average over all ensemble members. Thus *EVar*(*s*(*ij*)) is the ensemble variance of scalar *s* at point (i,j), and $s_m(ij)$ is the value of scalar *s* at point (i,j) in ensemble member *m*. The actual error variance *AVar* is defined as:

$$AVar\ \ s(ij)\ \ =\ \ s_v(ij)-\overline{s(ij)}\ ^2 \qquad (2)$$

where s_v denotes the analysis or verification value of scalar *s* at the verification time (i.e., the true value). Similarly, the ensemble covariance and actual error covariance are:

$$ECov\ \ s(ij),s(kl)\ \ = \frac{1}{N}\sum_{m=1}^{N}\Big[\ \ s_m(ij)-\overline{s(ij)}\ \ \ s_m(kl)-\overline{s(kl)}\ \ \Big] \qquad (3)$$

$$ACov\ \ s(ij),s(kl)\ \ = \ \ s_v(ij)-\overline{s(ij)}\ \ \ s_v(kl)-\overline{s(kl)} \qquad (4)$$

with *kl* representing a second point (k,l) and all other parameters are defined as before, with $s_m(kl)$ representing the value of scalar *s* at point (k,l) for ensemble member *m.*

We also define ensemble correlation *ECor* as:

$$ECor\ \ s(ij),s(kl)\ \ = \frac{ECov\ \ s(ij),s(kl)}{\sqrt{EVar\ \ s(ij)}\sqrt{EVar\ \ s(kl)}} \qquad (5)$$

Because the resulting matrices are quite large, especially for covariances, we use bootstrap sampling (Wilks 2006) to obtain a representative sample rather than computing the value for all points and point combinations. We then apply the binning technique described in Roulston (2005) to extract meaningful information from the resulting sample. This binning technique ranks the samples based on the values of the predictor. These values are then sorted into bins of a specified number of points and averaged. For this study, all of our bins represent 1000 points, so that the 1000 points with the lowest value of the predictor would go into the first bin, and the next 1000 points into the second bin, etc.

3. VARIANCE CALIBRATION

SCIPUFF ingests uncertainty information in the wind field via the UUE, VVE, and UVE parameters, which represent the single-point variance in U, variance in V, and covariance of U and V, respectively, in three space dimensions, and new arrays are read every meteorological model output time. Ideally, we would like to use the actual variance of these fields as inputs for SCIPUFF. The ensemble variance for each of these variables is calculated readily from the model ensemble outputs. Because, as we discuss above, this estimate is not necessarily the same as the actual variance, we would like a way to calibrate our ensemble variances to represent more accurately the actual variances. Toward this end, we apply the bootstrap sampling and binning technique described in Section 2, with ensemble variance as the predictor quantity and actual error variance as the predicted quantity.

Figure 1 shows the result when this is applied to the 15hPa AGL U component of the wind field of the 12-hour forecast, and Figures 2-5 shows the binned scatterplot for progressively longer forecasts, every 12 hours from 24 hours to 60 hours. All of the plots show a strong functional relationship between ensemble variance and actual error variance. This means that, not only is ensemble uncertainty a good predictor of the actual uncertainty, but also that we can often apply a simple, computationally inexpensive, linear calibration based on the least-squares fit line to the ensemble variances to derive expected actual variances. While some of the plots seem to suggest that a quadratic fit line would provide a better representation, the calibration using a quadratic fit line is still relatively inexpensive, and for simplicity, we will continue to discuss only the linear fit here, acknowledging there may be other functions that are useful.

Another important thing to note from the figures is that the slope of the line, which changes depending on the duration of the forecast, becomes steeper as the forecast period gets longer. This line represents the variance of the model compared to that of the actual atmosphere, with a 1:1 line representing an ideal ensemble, steeper lines an under-dispersive ensemble and shallower lines an over-dispersive ensemble. The steepening is indicative of the ensemble becoming less over-dispersive with increased integration time. This steepening is consistent with other studies of ensemble spread, which show that ensembles are start over-dispersive at short forecast times and gradually become under-dispersive at longer forecast times (e.g. Du *et al* 2003, Whitaker and Loughe 1998). This result is important to our calibration, because it means we must apply a different calibration to the ensemble at each forecast time for which we wish to produce actual variances.

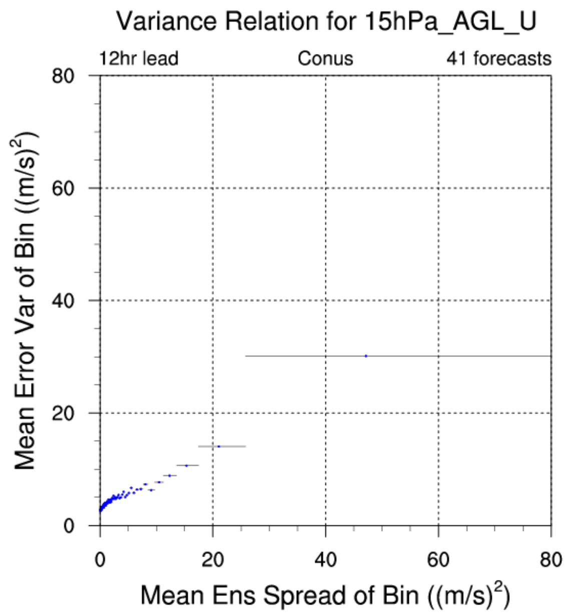

Figure 1 - Relationship of ensemble variance (abscissa) to the actual error variance (ordinate) for 12-hour SREF forecasts of 15hPa AGL U made during the study period 25 August – 15 September 2004. Horizontal lines represent the width of the 1000-point bin.

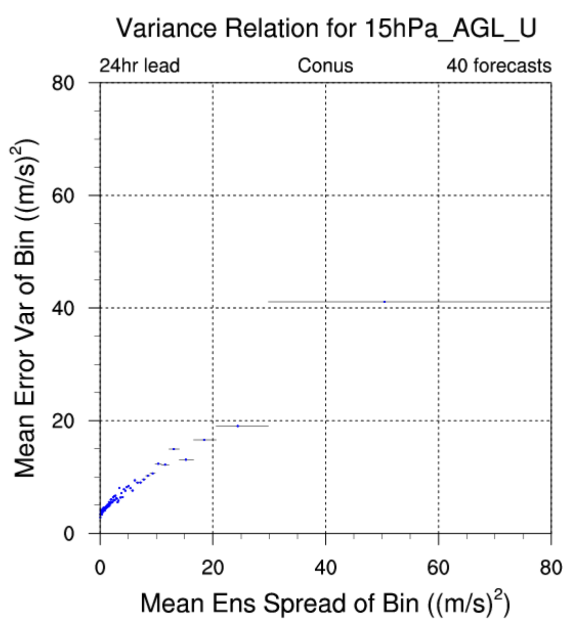

Figure 2 – As in Figure 1, except for 24-hour forecasts.

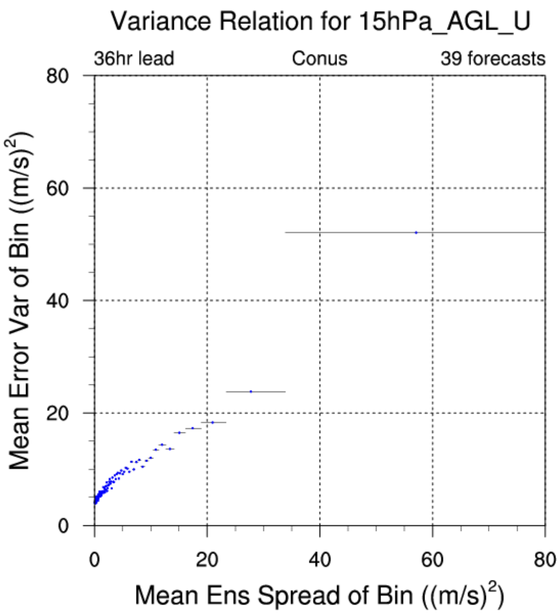

Figure 3 - As in Figure 1, except for 36-hour forecasts.

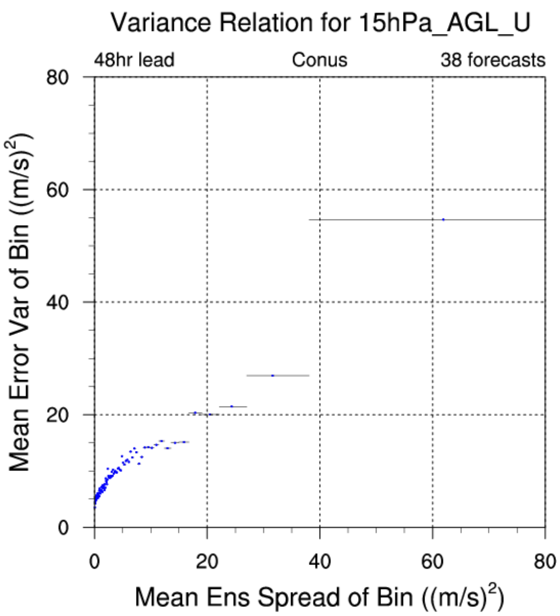

Figure 4 – As in Figure 1, except for 48-hour forecasts.

4. COVARIANCE CALIBRATION

Because spatially correlated errors in the wind field cause much larger errors when integrated over the path of a particle than do uncorrelated errors, it would also be useful to have information about the covariance of errors in the wind field at two separate points. We easily can apply the same technique that we

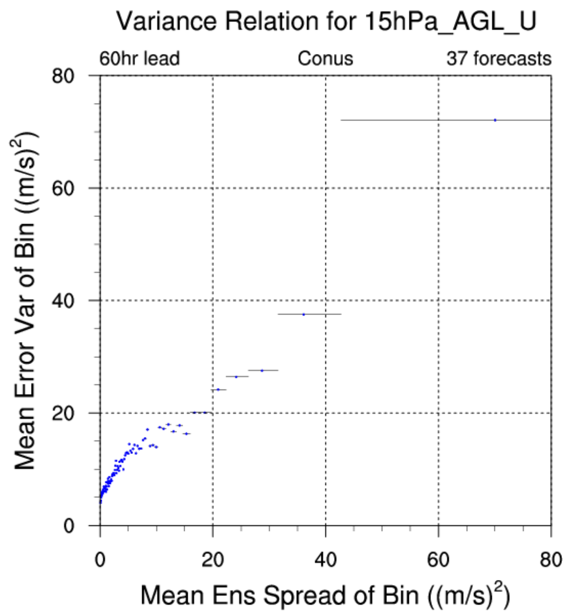

Figure 5 - As in Figure 1, except for 60-hour forecasts.

used on variances in Section 3 to two-point covariances to derive a calibration of ensemble covariances to actual error covariances.

Figure 6 demonstrates the relationship between ensemble covariance and actual error covariance for 48-hour forecasts. Because two points separated in space can have a negative correlation, the covariances in this plot are both positive and negative, as opposed to the variance plots, which are positive definite. The linear relationship does not look as good as that of the variance relations, but this is not surprising, as there is no reason to expect errors in the ensemble mean field to be spatially correlated, as they may be in individual model runs. Additionally, note the smaller scale for the actual error covariance compared to the ensemble covariance.

Recognizing the problem with using the ensemble mean field for covariance, we also explored using a control member of the ensemble, as the basis for computing our covariances by replacing the overbar terms in (3) and (4) with the value of ensemble member ETA-ctl1 at that point and removing that control member from the sum. These results for the same variable and forecast time as Figure 6 are shown in Figure 7. The linear relationship appears much stronger for this formulation of covariance.

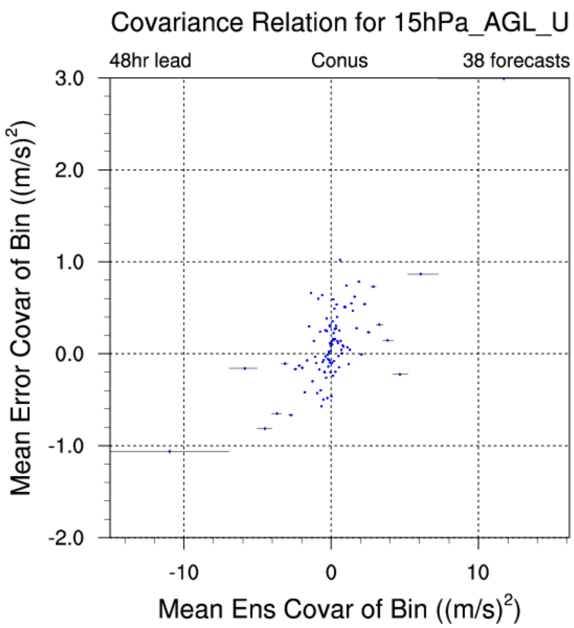

Figure 6 - The relationship of ensemble covariance (abscissa) to actual error covariance (ordinate) for 48-hour SREF forecasts of 15hPa AGL U made during the study period. Note the much smaller scale along the ordinate axis compared to the abscissa axis.

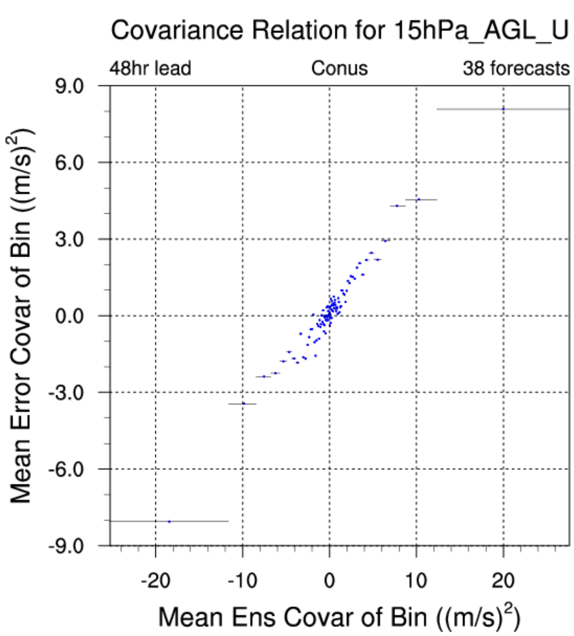

Figure 7 – Same as in Figure 6, except using a control member (ETA-ctl1) rather than the ensemble average for the overbar terms in calculating the covariances (using (3) and (4)).

5. DISTANCE CORRELATION

Although the work presented in the previous section provides a useful calibration for the ensemble covariance, SCIPUFF does not input Eulerian covariance information directly. Instead, information about the spatial error correlation is provided via the parameter SLE, which is a length scale related to the Lagrangian time scale (Peltier *et al.* 2007).

However, this Lagrangian length scale is likely related to the length scale of the Eulerian covariance in some way. Using this reasoning, we compare the correlation of ensemble spread between two points, which is just a normalized form of the covariance, as a function of the distance separating the two points. This relationship between distance and correlation of ensemble spread is shown in Figure 8, which demonstrates that there is a significant correlation between that at two points close together but almost no correlation for points at much greater separation distances. This result is consistent with the common practice of weighting observations in a data assimilation scheme based on background model error covariances such that they have less influence with increasing distance from the observation site (Daley 1991). This error correlation distance scale generally increases for observations located above the surface and boundary layer where the atmosphere contains much larger scale energy (e.g., Stauffer and Seaman 1994).

The exact length at which ensemble spread or errors are uncorrelated depends on many parameters including the variable type, forecast length, vertical level and the choice of threshold. Clearly, beyond about 1500km the correlations for low-level wind in this case are all within the noise level of the plot and are thus uncorrelated. Choosing 0.2 as the correlation cutoff yields an uncalibrated "length scale" of ~600km. This distance correlation for other variables and levels (Figures 9 and 10) shows significantly different correlation lengths as expected: 500hPa geopotential height, which is dominated by large-scale processes, appears to have a much longer correlation length, and 2m temperature, which has significant local effects, indicates a much shorter correlation length. Again using the 0.2 cut-off, the uncalibrated "length scale" is ~2000km for 500hPa geopotential height and <500km for 2m temperatures. These results are encouraging since they suggest that these distance

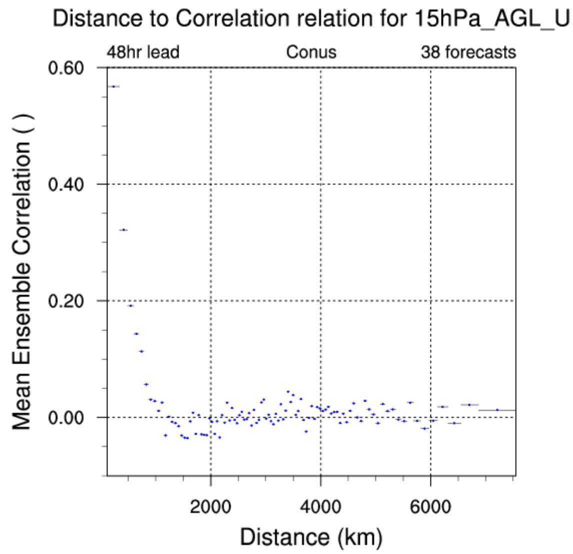

Figure 8 – The relationship of ensemble covariance of 15hPa AGL U to the distance between the two points for 48-hour SREF forecasts.

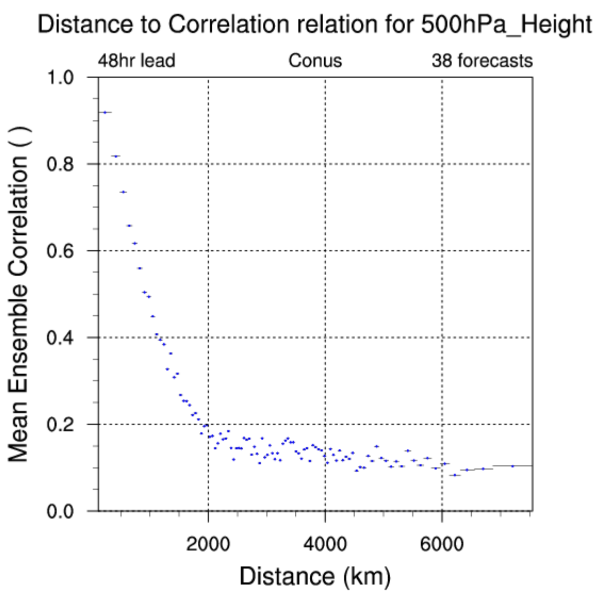

Figure 9 – As Figure 8, but for 500hPa geopotential heights.

correlations of ensemble spread may be useful for determining SLE.

6. CONCLUSIONS

In this study, we examine the utility of using the Roulston (2005) binning method to extract useful information and relate measures of model

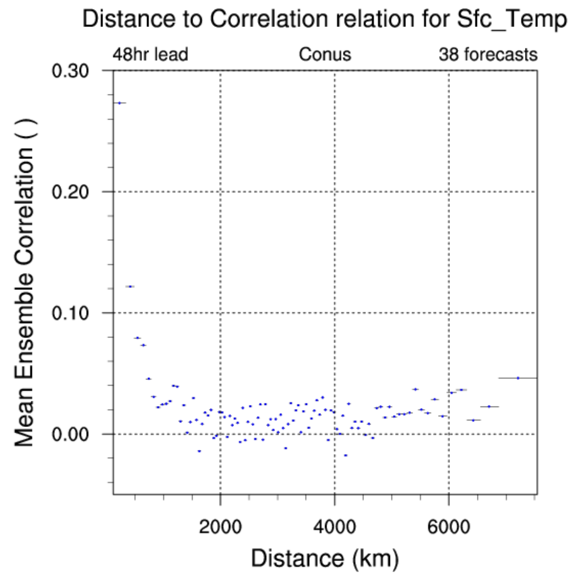

Figure 10 - As Figure 8, but for 2m Temperature.

ensemble uncertainty to actual uncertainty. This technique reveals a strong relationship between the ensemble variance and actual error variance in the 15hPa AGL wind field that can be fit well by a simple function (linear or perhaps quadratic). This function provides a computationally inexpensive way to calibrate ensemble variances in model output for use in the HPAC/SCIPUFF dispersion software used by DTRA or other applications discussed in Section 1. Use of this additional meteorological uncertainty information should produce better probabilistic forecasts of hazardous concentrations in the event of a toxic chemical release.

We also examine using the same technique to calibrate ensemble covariance. We hypothesize that because of the dynamically unbalanced nature of the ensemble mean field, results calculating the covariance based on the mean can fail to show a strong relationship between ensemble covariance and actual error covariance. However, when covariance was calculated using a single member (thus retaining continuity and dynamic consistency), the binning method in this application revealed a strong relationship between ensemble covariance and actual error covariance. Although there is currently no direct method for ingesting this sort of data into the DTRA HPAC/SCIPUFF system, it suggests that the ensemble covariance represents something physical, and thus useful in calculating correlation distances using the ensemble.

This relationship between distance and ensemble correlation is further explored using the binning technique. The plots reveal a strong relationship between distance and ensemble spread, producing an expected pattern of higher correlations at closer distances. The two other variable fields examined here, the 500hPa geopotential height and 2m temperature, have ensemble-spread distance-correlation relations that vary as expected from the low-level wind field: the spread in 500hPa height remained correlated over a much longer distance and the 2m temperature over a much shorter distance.

This area of research is still in its infancy and there are several areas where continued investigation is needed. Of primary interest is the sensitivity and added value of HPAC/SCIPUFF predictions to this uncertainty information, and whether these relationships are maintained for other variables and vertical levels, different size and resolution domains, varying proportions of land versus water in the domains, different seasons and locales, and especially when applied to model ensembles specifically designed to address the variability of the boundary layer, rather than those focused on mesoscale to synoptic-scale features and predicting precipitation.

7. ACKOWLEDGEMENTS

We would like to acknowledge Mark Roulston for stressing the importance of covariance information and starting us down the path of using a linear calibration via binning. Our many discussions with Ian Sykes about parameterizing meteorological uncertainty in SCIPUFF are much appreciated. We thank Jeff McQueen and the rest of our NCEP team members for providing full ensemble data for SREF. Finally, we thank our fellow Penn State team members Aijun Deng, Joel Peltier, and John Wyngaard for many useful discussions about this Lagrangian - Eulerian problem of quantifying uncertainty. This work is supported by L3-Titan/DTRA contract DTRA01-03-0013.

8. REFERENCES

Daley, R., 1991: Atmospheric Data Analysis. Cambridge University Press, New York, NY, 457 pp.

Deng, Aijun, Nelson L. Seaman, Glenn K. Hunter and David R. Stauffer, 2004: Evaluation of Interregional Transport Using the MM5-SCIPUFF System. *Journal of Applied Meteorology,* **43**, 1864-1886.

Du, J. G. DiMego, M. S. Tracton and B. Zhou, 2003: NCEP Short-range Ensemble Forecasting System: multi-IC, multi-model and multi-physics approach. Research Activities in Atmospheric and Oceanic Modelling (edited by J. Cote), Report 33, CAS/JSC Working Group Numerical Experimentation (WGNE), WMO/TD-No. 1161, 5.09-5.10.

McQueen, J., J. Du, B. Zhou, G. Manikin, B. Ferrier, H-Y. Chuang, G. DiMego, and Z. Toth, 2005: Recent Upgrades to the NCEP Short-Range Ensemble Forecasting System (SREF) and Future Plans. *Preprints, 17th Conference on Numerical Weather Prediction/21st Conference on Weather Analysis and Forecasting*, Washington D.C., Aug. 1-5, 2005, Amer. Meteor. Soc. (paper 11.2)

National Research Council, *Tracking and Predicting the Atmospheric Dispersion of Hazardous Material Releases. Implications for Homeland Security*, The National Academies Press, Washington, D.C., 2003.

Peltier, L. J., S.E. Haupt, and J.C. Wyngaard, D.R. Stauffer, and A. Deng, 2007: Assessing the Impact of Meteorological Data Uncertainty on SCIPUFF Concentration Predictions. Chemical and Biological Information Systems Conference, Austin, TX, Jan. 8-12.

Roulston, M. S., 2005: A comparison of predictors of the error of weather forecasts. *Nonlinear Processes in Geophysics*, **12**, 1021-1032.

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

Stauffer, D.R., G.K. Hunter, A. Deng, J.R. Zielonka, K. Tinklepaugh, P. Hayes, and C. Kiley, 2007: On the role of atmospheric data assimilation and model resolution on model forecast accuracy for the Torino Winter Olympics, 22nd Conference on Weather Analysis and Forecasting/18th Conference on Numerical Weather Prediction, June 25-29, Park City, UT.

Tennekes, H., Baede, A., and Opsteegh, J., 1987: Proc. ECMWF Workshop on predictability, chap. Forecasting forecast skill, ECMWF, Reading, UK.

Wilks, D.S., 2006: *Statistical Methods in the Atmospheric Sciences, Second Ed*., Academic Press, Burlington, MA, 627 pp.

Whitaker, Jeffery S. and Andrew F. Loughe, 1998: The Relationship between Ensemble Spread and Ensemble Mean Skill. Monthly Weather Review 126, 3292-3302.