10.4 "SUPERPARAMETERIZATION" AND STATISTICAL EMULATION IN THE LORENZ '96 SYSTEM

Daniel S. Wilks
Cornell University, Ithaca NY
dsw5@cornell.edu

1. INTRODUCTION

Computationally practical numerical integration of dynamical weather-forecast and climate models requires truncation of the smaller space and time scales. Typically the effects of these unresolved scales on the explicitly resolved prognostic variables are represented through deterministic parameterizations, which are relatively simple (and therefore computationally fast) specifications, formulated as functions of the resolved prognostic variables themselves. However, a deterministic parameterization specifies at best a mean influence of the unresolved scales, conditional on the resolved prognostic variables, whereas nature provides an unpredictable realization more-or-less near the parameterized value (Palmer *et al.* 2005, Wilks 2005). It is therefore natural to extend parameterizations to include explicitly stochastic influences, i.e. random realizations from a probability distribution, possibly centered in some way at a deterministic parameterization.

A different approach to the parameterization problem involves nesting a spatially and dynamically explicit, yet dimensionally truncated, model for important small-scale processes within each large-scale grid element, which is known as "superparameterization" (e.g., Grabowski 2001, 2004; Grabowski and Smolarkiewicz 1999; Khairoutdinov et al. 2005; Khairoutdinov and Randall 2001). Randall *et al*. (2003) provide an introduction to and review of this approach, which has to date focused on two-dimensional (e.g., longitude-height) cloud-resolving models that replace conventional parameterized tendencies with dynamically computed cloud influences on the larger scales. Although they are attractive conceptually, superparameterizations are very expensive computationally, yielding full-model integration times that are two to three orders of magnitude slower than with conventional parameterizations (Randall *et al*. 2003), even though three-dimensional small-scale cloud fields are idealized to one-dimensional arrays of atmospheric columns that do not communicate across the large-scale grid cells.

This paper considers the possibility of extracting a substantial part of the information provided by a superparameterization through the use of statistical models ("emulators") of the superparameterization, requiring substantially less computation. This idea is pursued in the setting of the simple Lorenz '96 model (Lorenz 1996, 2006), the two-scale structure of which is particularly well suited to study of parameterization issues (Crommelin and Vanden-Eijnden 2008, Horenko 2010, Wilks 2005).

2. THE LORENZ '96 SYSTEM AND A SUPERPARAMETERIZATION FOR ITS SMALL-SCALE DYNAMICS

The Lorenz '96 system (Lorenz 1996, 2006) is defined by the coupled equations

$$\frac{\mathrm{d}X_k}{\mathrm{d}t} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - \frac{hc}{b}\sum_{j=1}^{J} Y_{j,k} \quad , \; k = 1,\cdots, K \tag{1a}$$

and

$$\frac{\mathrm{d}Y_{j,k}}{\mathrm{d}t} = -cb\,Y_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - c\,Y_{j,k} + \frac{hc}{b}X_k \; ; \; j = 1,\cdots, J; \; k = 1,\cdots, K \tag{1b}$$

As illustrated in Figure 1, the K large-scale ("resolved") X variables whose tendencies are defined in Eq. (1a) are arranged in a cyclic domain that might be imagined as a latitude circle, so that $X_{K+1} = X_1$ and $X_0 = X_K$. Each of the X variables has associated with it J small-scale ("unresolved") Y variables, whose tendencies are defined in Eq. (1b). The Y variables are cyclic also and their dynamics communicate across the boundaries of the large-scale variables, so that $Y_{J+1,\,k} = Y_{1,\,k+1}$, and $Y_{0,k} = Y_{J,k-1}$. Here the parameters have been chosen as in Wilks (2005), with $K = 8$ large-scale variables, $J = 32$ small-scale variables associated with each large-scale variable, $F = 20$, $h = 1$, and $b = c = 10$. Integrations of Eq. (1) are carried out using a fourth-order Runge-Kutta algorithm with time step $\Delta t = 0.0001$, the results of which are regarded as true realizations from nature. The effects of the unresolved Y variables on the larger scale are felt through the last term in Eq. (1a), which will require parameterization or representation through a superparameterization of Eq. (1b), in model approximations to Eq. (1a).

An idealized superparameterization for the dynamics represented in Eq. (1b) has been constructed by assuming that basic dynamics of the Y variables are known, but that the parameter values are only approximately correct. It is further assumed that the small-scale dynamics are entirely local (no communication between Y variables associated with different X variables), as is typical for cloud superparameterizations (Randall 2003). Specifically,

this idealized superparameterization replaces Eq. (1b) with

$$\frac{dY_{j,k}}{dt} = -c^* b^* \; Y_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - c^* \; Y_{j,k} + \frac{h^* c^*}{b^*} X_k \quad , \; j = 1, \cdots, J^* \tag{2}$$

where the starred quantities are the incorrect parameter estimates $h^* = 0.4$, $c^* = 18$, $b^* = 7$, and $J^* = 34$, and locality is enforced by introducing the constant boundary values $Y_{0,k} = Y_{J^*+1,k} = Y_{J^*+2,k} = 0.16$. These five parameter values were obtained by optimizing the superparameterization to yield minimum forecast mean-squared error (MSE) for lead times of one to two time units, using second-order Runge-Kutta integration with time step $\Delta t = 0.005$.

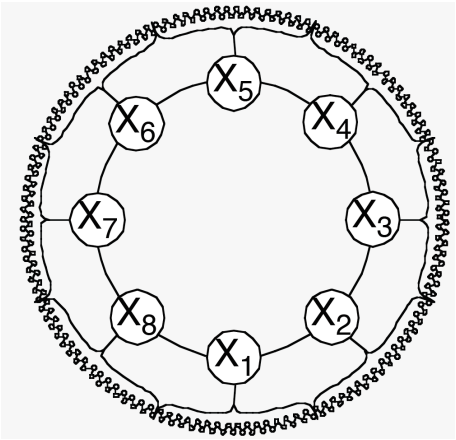

Figure 1. Schematic illustration of the Lorenz '96 system (Eq. 1), showing $K = 8$ resolved X variables, each with $J = 32$ unresolved Y variables; both of which are cyclic. Eq. (2) approximates the dynamics of Eq. (1b), using $J^* = 34$ unresolved variables for each X, which do not communicate with those of adjacent X's. From Wilks (2005).

The differences between the starred superparameterization parameter values and their "true" counterparts in Eq. (1b) represent the structural errors inherent in abstracting more complex smaller-scale behavior to a simpler superparameterization. Even though these values differ appreciably from their true counterparts, the coupling factor $h^*c^*/b^* = 1.029$ is nearly the same as the true ratio $hc/b = 1.000$. The Y values evolved according to Eq. (2) enter the final term of Eq. (1a) by substituting h^*c^*/b^* for hc/b, and summing over all $J^* = 34$ superparameterized small-scale variables.

3. EMULATION OF THE SUPERPARAMETERIZATION

A statistical emulator is a computationally fast statistical model summarizing the behavior, including the uncertainty, of a computationally demanding simulation model, which in the present context is the superparameterization of Eq. (2). That is, emulators are meant to provide fast probabilistic predictions of the behavior of the full simulation model output, for a given input.

Both regression and Bayesian emulators will be constructed and compared in the following. The present setting requires minimal complexity for the emulators, which will have only a single scalar input, X, and yield univariate probability distributions for the output, ΣY. The notation ΣY indicates specification by an emulator of the sum of the $J^* = 34$ small-scale superparameterization variables $Y_{j,k}$ in Eq.(2), which affects the large-scale X_k through the final term in Eq. (1). Because the point of statistical emulation is to represent the behavior of a computationally expensive model, training data will be very limited in general for real settings, so only $n = 20$ (X, ΣY) pairs from the outputs of Eqs. (1a) and (2) are used to train each emulator. These have been taken from a single integration of length 60 time units, with each training-data pair separated from the next by three time units, so that successive pairs are nearly independent.

All the emulator formulations described below are implemented by integrating Eq. (1a) using second-order Runge-Kutta integration with time step $\Delta t = 0.005$.

3.1 Regression emulators

The basic regression emulator is constructed simply as a polynomial regression representing the $n = 20$ training points,

$$\Sigma Y = -0.04328 + 1.1648\, X - 0.002656\, X^3 \quad . \tag{3}$$

This regression exhibits an R^2 of 91.4%, and has an estimated residual variance of $s^2_e = 1.591^2$. It is similar to the polynomial parameterization regressions in Wilks (2005), although there full fourth-order polynomial regressions were justified by the $n = 2000$ training-data samples taken from the "true" system (Eq. 1) rather than from output of the superparameterization (Eqs. 1a and 2).

The regression emulator implies a specification for ΣY that is a probability distribution, which is assumed here to be Gaussian, centered on Eq. (3) for a given input X. In implementation this first regression emulator operates by drawing an independent realization from this distribution, with standard deviation $s_e = 1.591$, and substituting this into the last term of Eq. (1a) together with the estimated coupling factor $h^*c^*/b^* = 1.029$ on each $\Delta t = .005$ time step. This procedure will be called the independent regression emulator in the following.

In previous work on stochastic parameterizations, both with the Lorenz '96 system (Wilks 2005) and in other settings (Berner *et al*. 2009; Buizza *et al*. 1999; Lin and Neelin 2000, 2003), it has been found that serially independent random numbers are relatively ineffective at improving ensemble forecast performance, and that serially dependent (i.e. temporally coherent) random innovations yield better forecast ensembles. Therefore an autoregressive regression emulator is also investigated here, which is

built on realizations of serially correlated Gaussian time series generated using the simple first-order autoregression

$$e_t = \phi\, e_{t-\Delta t} + \sigma_e (1 - \phi^2)\, z_t \quad . \qquad (4)$$

Here e_t and $e_{t-\Delta t}$ are successive values of the random forcing for the emulator separated by a single time step, z_t is an independent random draw from the standard Gaussian distribution, and for the autoregressive regression emulator s_e = 1.591 and ϕ = 0.995. The autoregressive parameter ϕ has been chosen on the basis of the result in Wilks (2005) that it should be a large value (reflecting strong serial correlation) when s_e is set equal to the root-mean-squared error (RMSE) of the regression. It has not resulted from a tuning exercise, and so could almost certainly be improved upon.

3.2 Bayesian Gaussian process emulators

The Bayesian emulators investigated here are based on the conventional formulation for a Bayesian Gaussian process emulator, using a linear mean function and weak priors for the parameters (Kennedy and O'Hagan 2001; O'Hagan 1992, 2006; see also the excellent treatment at http://www.mucm.ac.uk/toolkit).

Conditional on the single scalar input X, the result for the expected value of the output from the superparameterization ΣY is

$$\mathrm{E}\left[\Sigma Y\right] = b_0 + b_1 X + M(X, \mathbf{X}', \delta) \quad , \qquad (5)$$

and the corresponding (approximately Gaussian) variance is

$$\mathrm{Var}\left[\Sigma Y\right] = \hat{\sigma}^2\, V(X, \mathbf{X}', \delta) \quad . \qquad (6)$$

Eq. (5) estimates the mean output from the superparameterization as a linear function of the resolved variable X, plus an adjustment M that depends on X and its distance from the n- (=20) member training data vector $\boldsymbol{X}'$ in terms of a decorrelation scale δ. Eq. (6) quantifies uncertainty about the superparameterization output in terms of the variance parameter $\hat{\sigma}^2$ and a variance adjustment V that also depends on X, $\boldsymbol{X}'$, and δ. The mathematical details of parameter estimation (yielding b_0 = 1.216, b_1 = 0.7891, $\hat{\sigma}^2$= 1.861^2, and δ = 0.01156), and definition of the adjustment functions M and V, are presented in Wilks (2012).

Analogously to the independent regression emulator, the independent Bayesian emulator operates by calculating a conditional mean using Eq. (5) at each time step, adding to that an independent random draw from the Gaussian distribution with zero mean and the variance specified by Eq. (6), and substituting this independent realization for ΣY in the last term of Eq. (1a), together with the estimated coupling factor hc/b.

Also considered in the following is an ad-hoc variant of the independent Bayesian emulator, which is based on the mean function in Eq. (5), but with the random excursion around the mean at each time step generated with the autoregression in Eq. (4). Again somewhat arbitrarily and without tuning, ϕ = 0.995 and $s_e = \sqrt{\hat{\sigma}^2}$ = 1.861 are assumed. This formulation will be referred to as the autoregressive Bayesian emulator, even though the autoregressive component is an add-on that is unrelated to the Bayesian structure that yields Eqs. (5) and (6).

4. CONSTRUCTION OF FORECAST ENSEMBLES

Forecast ensembles are initialized using the same algorithm, and based on the same set of 10,000 initial points, as used in Wilks (2005). These initial points are taken from a long integration of the "true" dynamics (Eq. 1), with each K-member set of X variables separated by 50 time units. Initial ensemble distributions around these points are generated in a way that approximates the local attractor shape, as would a data assimilation procedure.

Superparameterization integrations of Eq. (2) require in addition an initialization for the J* = 34 unresolved Y variables. These initializations are achieved through a (35-dimensional) multivariate normal distribution for the joint distribution of the Y variables and their associated X variable, fit to results from a sample integration of Eqs. (1a) and (2). Each superparameterization ensemble member is then initialized with a different realization from the joint distribution for each set of Y variables, conditional on the initialization for its X variable (e.g., Wilks 2011, Eq. 11.8). In effect this procedure amounts to an initialization of random turbulence as in Khairoutdinov *et al.* (2005), without which the superparameterization integrations are slow to spin up fully, and yield inferior forecasts.

5. ENSEMBLE FORECASTING RESULTS

The performance of forecast ensembles based on the four statistical emulators (independent and autoregressive regression, and independent and autoregressive Bayesian) are compared in this section, in relation to two superparameterization-based ensembles. The first of these is the deterministic superparameterization described in Section 2, whose only random input is in the initialization of the small-scale Y variables. In addition, a stochastic superparameterization is also investigated, which adds autoregressive noise (Eq. 4), with ϕ = 0.995 and standard deviation equal to half of $\sqrt{\hat{\sigma}^2}$ from the Bayesian emulator, to the superparameterization output of ΣY at each time step. This amplitude for the random forcing is reduced relative to its counterparts in the autoregressive emulators because the internal dynamics of the superparameterization (Eq. 2) themselves induce some variance and autocorrelation. Again, this parameter choice could almost certainly be improved upon through a tuning exercise.

Only results for ensemble size 20 will be presented, as they are representative of those for other choices, although as would be expected smaller ensembles yield generally poorer forecasts and larger ensembles yield generally better forecasts, for all six formulations.

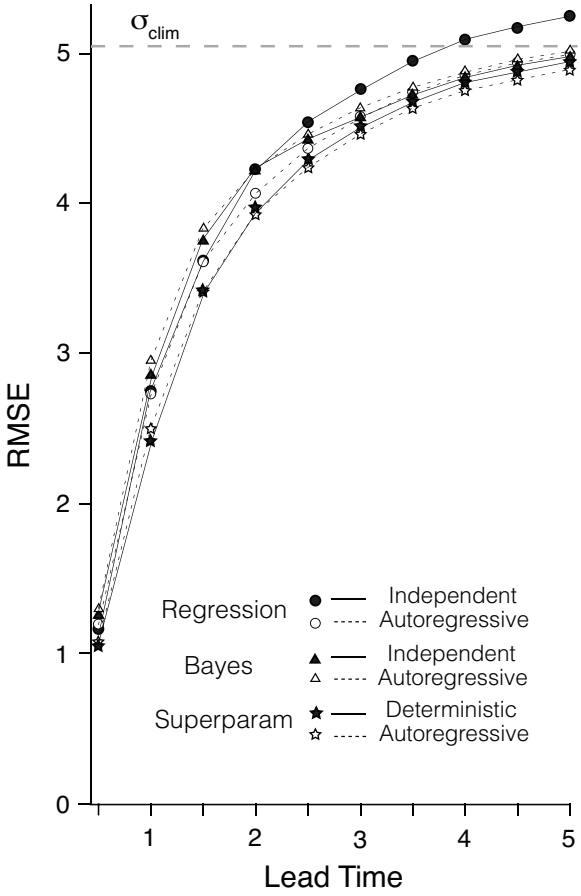

Figure 2. Average root-mean squared error for ensemble-mean forecasts from regression emulators (circles), Bayesian emulators (triangles), and superparameterizations (stars). Independent emulators and deterministic superparameterization are indicated by filled symbols connected by solid lines. Autoregressive emulators and autoregressive stochastic superparameterization are indicated by open symbols connected by dashed lines.

Figure 2 shows RMSE for ensemble mean forecasts through lead times of 5 units, at which point forecast RMSE is comparable to the climatological standard deviation. As would be expected, superparameterization forecast errors (stars) are smaller on average than errors from the emulators, and indeed are smaller than the errors from the fourth-order polynomial parameterization in Wilks (2005) through lead times of 2.5 units, and essentially the same thereafter, even though that polynomial parameterization was trained on a large sample of realizations of the "true" system. However, remarkably the superparameterization results are only slightly better than those for the autoregressive regression emulator (that has been derived from a very small sample of superparameterization realizations), which itself shows RMSE comparable to or smaller than the earlier polynomial parameterization results for the first three lead times. Deterministic superparameterization errors are slightly smaller on average for lead times of 0.5 and 1 time unit, and stochastic superparameterization errors are slightly smaller for lead times of 2 and longer. Among the emulators, best results are generally exhibited by the autoregressive regression emulator (open circles), although the independent (filled triangles) and autoregressive (open triangles) Bayesian emulators yield generally similar error magnitudes. Errors for the independent regression emulator (filled circles) are generally comparable to the others until the later lead times, where they are clearly worse.

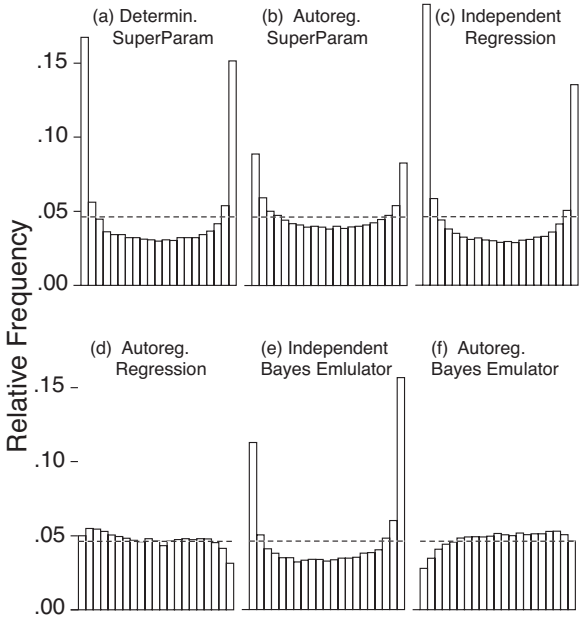

Figure 3. Verification rank histograms at 1 time unit lead, for (a) deterministic superparameterization, (b) autoregressive superparameterization, (c), independent regression emulator, (d), autoregressive regression emulator, (e), independent Bayesian emulator, and (f), autoregressive Bayesian emulator. Dashed lines show level of perfect rank uniformity.

Figure 3 shows verification rank histograms at lead times of 1 unit. The rank histograms for the deterministic superparameterization (a), the independent regression emulator (c), and the independent Bayesian emulator (e) all indicate notable underdispersion of the ensembles. Adding autoregressive components to the superparameterization (b), regression emulator (d) and Bayesian emulator (f) greatly improves the underdispersion, and in the case of the emulators nearly eliminates it. Much of the lack of rank uniformity in panels (d) and (f) result from the positive bias (+0.05) of the autoregressive regression emulator (d) and the negative bias (–0.16) for the

autoregressive Bayesian emulator (f), which are large enough to be detected by the rank histograms but too small to have an appreciable impact on RMSE. Comparable biases are also evident in the rank histograms for the respective independent emulators (c) and (e).

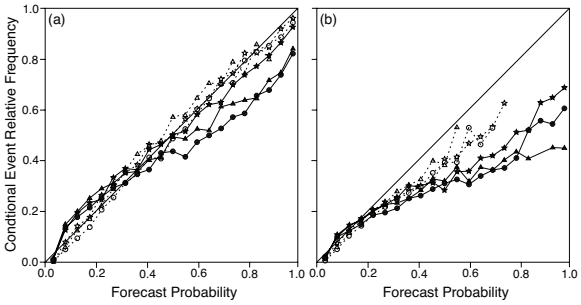

Figure 4. Calibration functions from reliability diagrams for forecasts above the climatological 90th percentile ($X \geq 10.9$), at lead times of (a) 1 unit and (b) 2 time units. Plotting symbols are as defined in Figure 2. Points are plotted only for subsamples of size 50 or larger. Total sample size is 10,000.

Figure 4 shows calibration functions from reliability diagrams for forecasts of X at or above its 90th percentile, for (a) 1 and (b) 2 time unit leads. As indicated also by the rank histograms in Figure 3, the deterministic superparameterization (solid stars) and especially the independent regression (solid circles) and Bayesian (solid triangles) emulators exhibit overconfidence, with overforecasting of the larger probabilities at both lead times, and underforecasting of the smaller probabilities in panel (a). In contrast, the stochastic superparameterization and the autoregressive emulators (open symbols) all show excellent calibration at the 1-unit lead time, and more modest overconfidence at the 2-unit lead time.

Table 1 shows the reliability (REL) and resolution (RES) components of the Murphy (1973) decomposition of the Brier Score (BS) for the six ensemble formulations forecasting the probability of X above its 90th percentile, as functions of lead time. Table 1a shows that the autoregressive Bayesian emulator exhibits the best (smallest) REL (which corresponds to subsample-size-weighted probability calibration) at all except the first lead time, and is nearly best for lead time of 1 also. In terms of reliability, the ensembles including autocorrelated noise are in each case better calibrated than their counterparts with no or independent random forcing, at all lead times.

The resolution results in Table 1b, corresponding in a sense to intrinsic information content without regard to correct labeling of the forecasts, is not surprisingly best for the superparameterization, of which the emulators are statistical abstractions. Among the emulators, the independent regression emulator shows better resolution at early leads, with the independent Bayesian emulator being better at later leads. Here the deterministic superparameterization and independent emulators show better resolution than their autoregressive counterparts, but only slightly so in most cases.

Table 1c shows best (smallest) overall Brier scores for the superparameterizations, with the autoregressive regression best among the emulators at early leads, and the independent Bayesian emulator best at the later leads. Here the Brier score for climatological forecasts is (0.9)(1–0.9) = 0.09, so skill scores corresponding to the results in Table 1c range from 46.4% for the deterministic superparameterization at lead 1 to slightly negative skills for all formulations at lead 5.

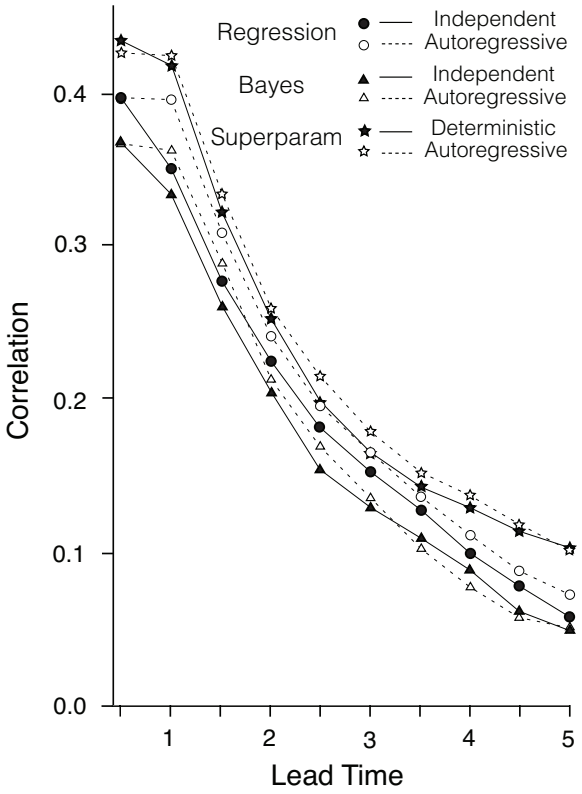

Figure 5. Spread-skill correlations between ensemble standard deviation and mean-absolute error of the ensemble mean forecasts, for the six ensemble formulations as a function of lead time.

Finally, Figure 5 shows spread-skill correlations (between ensemble standard deviation and mean-absolute error of the ensemble mean forecasts), for forecast leads through 5 time units. These correlations are generally high for the early lead times but are still positive at the later lead times. With only a few exceptions, the superparameterization correlations are higher than the regression emulator correlations, which in turn are higher than the Bayesian emulator correlations. In most cases the ensembles forced with autoregressive noise exhibit higher spread-skill correlations than their deterministic or independent-noise counterparts.

6. SUMMARY AND CONCLUSIONS

An idealized "superparameterization", which is an abbreviated but dynamically explicit representation of small-scale influences on the conventionally resolved larger scales, has been constructed within the Lorenz '96 system. This superparameterization has been used to investigate the feasibility of abstracting the greater portion of its information content through use of computationally faster statistical summaries, or emulators, based on an extremely limited training sample size. Both a simple regression emulator and a mathematically more sophisticated Bayesian emulator were constructed, and a large number of ensemble forecasts based on these with both temporally independent and autocorrelated stochastic components were integrated. Results were compared to ensembles based on the superparameterization, formulated both with and without stochastic forcing.

Ensembles based on the superparameterizations yielded better ensemble-mean RMSE; but considering the very small amount of training data the best of the statistical emulators, the regression emulator with autoregressive forcing, performed remarkably well. Consistent with previous studies, the nonstochastic superparameterization and temporally independent emulators yielded strongly underdispersed ensembles, whereas autoregressive random forcing ameliorated the underdispersion to a very substantial degree in all three cases.

The Bayesian emulator with autocorrelated random forcing yielded the best-calibrated probability forecasts, as measured by the reliability term of the Murphy (1973) decomposition of the Brier score, and for lead times beyond 1 unit yielded better reliability than either of the superparameterization ensembles. The superparameterization ensembles exhibited best resolution and overall Brier score, with regression emulators better than Bayesian emulators at the earlier lead times, and the reverse at later lead times, with respect to both of these attributes.

The superparameterization and its emulators used here were constructed in a deliberately imperfect way in order to mimic inaccuracies in the structure of real atmospheric models. Future improvements might include stochastic models more sophisticated than Eq. (4) for the time dependence of the random innovations to the superparameterization and its regression emulator, the details of which could be informed by the statistical characteristics of the resolved model errors. Similarly explicit time dependence could also be built into the structure of the Bayesian emulator, of course at the cost of greater complexity.

Overall the present results are encouraging with respect to the prospects for capturing much of the information content of a computationally expensive superparameterization using well-designed and computationally fast statistical emulators. Even though the present training data were extremely limited in number, the autoregressive regression emulator yielded results comparable to those for a similar formulation trained on a very large sample from the "true" system (Wilks 2005). Of course the Lorenz '96 setting is a highly idealized abstraction of the real atmosphere, but the results here suggest that it should be worthwhile to extend the present experimental protocol to forecast ensembles within more realistic dynamical models. In that case the implementation will be substantially more complicated, probably requiring emulators with multiple inputs and outputs.

Interestingly, forecast performance for superparameterization ensembles was improved by addition of autocorrelated random forcing to the superparameterization output. In realistic models of the atmosphere this procedure would correspond to accounting for deficiencies in (conventional) parameterizations of such process as cloud microphysics and radiative transfer whose scales are too small to be resolved by the superparameterization. Even in settings where the computational resources may be adequate for use of superparameterizations rather than their more economical statistical emulators, the results here suggest that better forecasts may be achievable by also including explicitly stochastic elements.

7. REFERENCES

Berner J, Shutts GJ, Leutbecher M, Palmer TN. 2009. A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system, *J. Atmos. Sci*., **66**, 603–626.

Buizza R, Miller M, Palmer TN. 1999. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, **125**, 2887-2908.

Crommelin D, Vanden-Eijnden E. 2008. Subgrid-scale parameterization with conditional Markov chains. *J. Atmos. Sci.*, **65**, 2661-2675.

Grabowski WW. 2001. Coupling cloud processes with the large-scale dynamics using the cloud-resolving convection parameterization (CRCP). *J. Atmos. Sci*., **58**, 978-997.

Grabowski WW. 2004. An improved framework for superparameterization. *J. Atmos. Sci*., **61**, 1940-1952.

Grabowski WW, Smolarkiewicz PK. 1999. CRCP: a cloud resolving convection parameterization for modeling the tropical convective atmosphere. *Physica D*, **133**, 171-178.

Horenko I. 2010. On the identification of nonstationary factor models and their application to atmospheric data analysis. *J. Atmos. Sci.*, **67**, 1559-1574.

Kennedy MC, O'Hagan A. 2001. Bayesian calibration of computer models. *J. Roy. Stat. Soc. B*, **63**, 425-464.

Khairoutdinov MF, Randall DA. 2001. A cloud resolving model as a cloud parameterization in the NCAR Community Climate System Model: preliminary results. *Geophys. Res. Lett.*, **28**, 3716-3620.

Khairoutdinov M, Randall D, DeMott C. 2005. Simulations of the atmospheric general circulation using a cloud-resolving model as a superparameterization of physical processes. *J. Atmos Sci.*, **62**, 2136-2154.

Lin J, Neelin JD. 2000. Influence of a stochastic moist convective parameterization on tropical climate variability. *Geophys. Res. Lett.*, **27**, 3691-3694.

Lin J, Neelin JD. 2003. Toward stochastic deep convective parameterization in general circulation models. *Geophys. Res. Lett.*, **30**, 1162-1165.

Lorenz EN. 1996. 'Predictability — A problem partly solved'. Pp. 1–18 in Proceedings of seminar on predictability: Volume 1. ECMWF, Reading, UK.

Lorenz EN. 2006. Predictability — A problem partly solved. In: Palmer T and Hagedorn R, Eds., *Predictability of Weather and Climate*. Cambridge University Press, 40-58.

Murphy AH. 1973. A new vector partition of the probability score. *J. Appl. Meteorol.*, **12**, 595-600.

O'Hagan A. 1978. Curve fitting and optimal design for prediction. *J. R. Stat. Soc. B*, **40**, 1-42.

O'Hagan A. 1992. 'Some Bayesian numerical analysis'. In: Bernardo JM, Berger JO, Dawid AP and Smith AFM (Eds.), *Bayesian Statistics 4*. Oxford University Press, 345-363.

O'Hagan A. 2006. Bayesian analysis of computer code outputs: a tutorial. *Rel. Eng. System Safety*, **91**, 1290-1300.

Randall D, Khairoutdinov M, Arakawa A, Grabowski W. 2003. Breaking the cloud parameterization deadlock. *Bull. Amer. Meteor. Soc.*, **84**, 1547-1564.

Wilks DS. 2005. Effects of stochastic parametrizations in the Lorenz '96 system. *Q. J. R. Meteorol. Soc.*, **131**, 389-407.

Wilks DS. 2011. *Statistical Methods in the Atmospheric Sciences*, 3rd Edition. Academic Press, 676 pp.

Wilks, DS. 2012. "Superparameterization" and statistical emulation in the Lorenz '96 system. *Q. J. R. Meteorol. Soc.*, in press.

Table 1. Reliability (a) and Resolution (b) components of the Murphy (1973) decomposition of the Brier Score (c), for probability forecasts of *X* larger than its 90th percentile (> 10.9). Best results at each lead time are indicated in boldface.

		Superparameterizations		Regression Emulators		Bayesian Emulators	
	Lead	Determ.	Autoreg.	Indep.	Autoreg.	Indep.	Autoreg.
(a) REL							
	1	0.0008	**0.0006**	0.0027	0.0007	0.0019	0.0008
	2	0.0029	0.0010	0.0050	0.0010	0.0026	**0.0002**
	3	0.0031	0.0015	0.0049	0.0017	0.0019	**0.0006**
	4	0.0034	0.0021	0.0052	0.0025	0.0024	**0.0012**
	5	0.0032	0.0024	0.0056	0.0030	0.0025	**0.0019**
(b) RES							
	1	**0.0429**	0.0410	0.0364	0.0357	0.0333	0.0308
	2	**0.0128**	0.0123	0.0103	0.0102	0.0088	0.0075
	3	**0.0063**	0.0061	0.0046	0.0045	0.0048	0.0031
	4	**0.0034**	**0.0034**	0.0020	0.0019	0.0023	0.0011
	5	**0.0023**	0.0021	0.0008	0.0008	0.0012	0.0004
(c) BS							
	1	**0.0482**	0.0498	0.0566	0.0552	0.0588	0.0603
	2	0.0800	**0.0787**	0.0846	0.0808	0.0838	0.0827
	3	0.0878	**0.0864**	0.0913	0.0882	0.0882	0.0885
	4	0.0900	**0.0888**	0.0933	0.0907	0.0902	0.0901
	5	0.0920	**0.0914**	0.0959	0.0932	0.0924	0.0926