Nonlin. Processes Geophys., 19, 569–575, 2012
www.nonlin-processes-geophys.net/19/569/2012/
doi:10.5194/npg-19-569-2012

Nonlinear Processes in Geophysics

Forecast improvement in Lorenz 96 system

L. Basnarkov[1,2] **and L. Kocarev**[1,2,3]

[1]SS. Cyril and Methodius University, Faculty of Computer Science and Engineering, P.O. Box 393, Skopje, Macedonia
[2]Macedonian Academy of Sciences and Arts, Skopje, Macedonia
[3]BioCircuits Institute, University of California San Diego, USA

Correspondence to: L. Basnarkov (lasko.basnarkov@finki.ukim.mk)

Received: 5 July 2012 – Accepted: 9 September 2012 – Published: 2 October 2012

Abstract. Contemporary numerical weather prediction schemes are based on ensemble forecasting. Ensemble members are obtained by taking different (perturbed) models started with different initial conditions. We introduce one type of improved model that represents interactive ensemble of individual models. The improved model's performance is tested with the Lorenz 96 toy model. One complex model is considered as reality, while its imperfect models are taken to be structurally simpler and with lower resolution. The improved model is defined as one with tendency that is weighted average of the tendencies of individual models. The weights are calculated from past observations by minimizing the average difference between the improved model's tendency and that of the reality. It is numerically verified that the improved model has better ability for short-term prediction than any of the individual models.

1 Introduction

We are witnessing steady increase of accuracy of numerical weather prediction. It is based on improvements in understanding the processes that govern the atmosphere and the ocean and availability of computational power that directly affects the grid resolution. Although modeling of the dynamics of the fluids is build upon the same physical laws, representation of sub-grid processes is considered differently at different meteorological centers. Thus we are facing now with dozens of operational atmospheric models that generally differ in parameterisation of the unresolved physical processes. From the other side increasing of the computational resources enabled ensemble forecasting – running multiple simulations of the same model with different initial conditions, or even using grand ensembles – multiple runs with different initializations for each of the multitude of models obtained with perturbations of the operational model (Stainforth et al., 2005; Lewis, 2005). Even an ensemble of different models is under consideration (Bougeault et al., 2010). Already ideas about dynamically connecting the models are emerging, at least coupling of ocean and atmospheric models (Kirtman and Shukla, 2002). Further improvement of weather forecasting and climate projection is expected if one applies ensemble of dynamically interacting models – perturbed variants of single models or different models. The problem that has to be solved is then how to combine the individual models: coupling the dynamical variables or exchanging some fields or fluxes as is the case of coupling the atmospheric and oceanic models. In this work we show how, in a medium-dimensionality dynamical system – Lorenz 96 model, short-term prediction can be improved by using weighted combination of different models.

A century ago Bjerknes realized that, in weather forecasting, one faces two types of errors: inability to copy the atmospheric dynamics – model error, and limitation of the knowing its state – initial condition or analysis error (Bjerknes, 1911). The second issue has gained importance after Lorenz's discovery that, for nonlinear dynamical systems, the separation of trajectories starting with close initial conditions is exponential on average (Smith et al., 1999; Lorenz, 1963). To ameliorate the effect of this type of divergence, scientists have developed different data assimilation techniques. The problem with the inability to build a perfect model has received much interest in recent years (Orrell et al., 2001; Judd and Smith, 2004; Judd et al., 2008). Comparison of the higher-dimensional truth (atmosphere in this case) and its model is done in the state space of the model – the state of the truth is projected onto that space. The evolution of states of the atmosphere and its model are governed with

Published by Copernicus Publications on behalf of the European Geosciences Union & the American Geophysical Union.

their respective tendencies. So, the atmosphere tendency is projected onto the model space as well. Then it is almost certain that the tendency vectors are different and their mismatch is the so-called tendency error (Orrell et al., 2001). We propose a particular approach for decreasing tendency error and thus an improvement of the short-term forecast. It is based on using a model with tendency that is weighted average of tendencies of individual models. The weights are obtained with statistical techniques based on past observations. Our work was based on encouraging results with the Lorenz 63 model (van den Berge et al., 2011; Wiegerinck et al., 2011; Mirchev et al., 2012). We shortly introduce the Lorenz 96 model in Sect. 2. The interacting ensemble is introduced in Sect. 3, and the results are presented in Sect. 4. We finish with the Conclusion.

2 The Lorenz 96 model

In 1996 Lorenz introduced a medium-dimensionality model for modeling the evolution of one scalar atmospheric variable defined over single latitude circle (Lorenz, 1996, 2004). Although artificial, the model shares some basic properties that any atmospheric model possesses: damping, advection and forcing. Since its introduction it has been used widely as a testbed for different ideas (e.g., Orrell, 2003; Lorenz, 2004; Ott et al., 2004). One can see the model as discretized version of partial differential equation describing the evolution of one-dimensional quantity. There are three versions of the model with increasing complexity. The basic one – version I – simply captures the chaotic nature of the atmosphere and gives solution profile with irregular traveling waves. We will not use it and so its definition is skipped. The version II is generally similar to the simpler one but with its increased complexity the solution is smoother. The equation of motion of the scalar field X at point n (there are N points around the circle) reads

$$\dot{X}_n = [X,X]_{K,n} - X_n + F_n, \tag{1}$$

where the linear term $-X_n$ corresponds to the damping, F_n is spatially dependent forcing and the term in the brackets models the advection. It is short hand notation of the sum

$$[X,Y]_{K,n} = \sum_{j=-J}^{J}{}' \sum_{i=-J}^{J}{}' (-X_{n-2K-i}Y_{n-K-j} + X_{n-K+j-i}Y_{n+K+j})/K^2, \tag{2}$$

where the number K determines the extension of the influence, and $J = K/2$ if K is even and $J = (K-1)/2$ if K is odd. The sign prime at the sum means that, in the case when K is even, the first and last terms are divided by 2 and when K is odd one has ordinary sum. The most complex version (model III) has small-scale activity added to the large-scale one which is given by the model II. The dynamical variable of the model III is Z_n, and it evolves according to

$$\dot{Z}_n = [X,X]_{K,n} + b^2[Y,Y]_{1,n} + c[Y,X]_{1,n} - X_n - bY_n + F_n, \tag{3}$$

where b and c are parameters and the large- and small-scale variables are given by

$$X_n = \sum_{i=-I}^{I}{}' (\alpha - \beta|i|) Z_{n+i}$$
$$Y_n = Z_n - X_n. \tag{4}$$

The integer I and parameters α and β in the last equation are chosen so as the large-scale variable X_n is smoothed version of Z_n (the sum acts as a low-pass filter), while Y_n represents the fast processes. The brackets and the sign prime in the sum have the same meaning as for the model II. Lorenz suggested following constraint for the parameters

$$\alpha = (3I^2+3)/(2I^3+4I),$$
$$\beta = (2I^2+1)/(I^4+2I^2). \tag{5}$$

To achieve a model that has desired properties (chaotical behavior, large and small-scale dynamics, traveling wave solution), Lorenz took the following parameter values: $b = 10$, $c = 2.5$. Within numerical experiments the number of grid-points is standardly chosen to be $N = 960$ and also $K = 32$ and $I = 12$. For integration of the equations of motion, a Runge-Kutta of fourth order is used with time step 0.001. In original formulation the external forcing was taken constant $F = 15$ which induced chaotic behavior of the solution of the models.

While trying to estimate the performance of a model in explaining some physical reality, one should bear in mind that the reality is much more complex than any of its models. From this observation one can conclude that the reality has more degrees of freedom than the model, or in mathematical description the reality has more equations and variables. In majority of the studies, scientists generally assume that the models have the same dimensionality as the truth and that they differ only in the values of the parameters in the equations. We abandon that simplification and consider that the reality has to be more complex than the model. In this toy example as a truth, we take model III from the Lorenz hierarchy and use model II for explanation of the reality. To have a more real setting, we assume spatially dependent forcing F_n. In order to have a smoothly varying forcing, we took perturbation of the constant $f_0 = 15$ that has randomly chosen Fourier components up to the order 10, while the higher were taken to be zero. More precisely the forcing is given by the sum

$$F_n = f_0\left[1 + \sum_{m=1}^{10} f_m^c \cos\left(\frac{2\pi mn}{N}\right) + f_m^s \sin\left(\frac{2\pi mn}{N}\right)\right], \tag{6}$$

where the spectral components f_m^c and f_m^s have random values from the interval $[-0.5, 0.5]$. Also we assume that the models have different values of the forcing from the truth and between themselves as well. That can be represented if to the forcing of the truth is added another sum with random coefficients for each model. This assumption should mimic different parameterisations of the unresolved physical processes obtained by different scientific groups. So, there are three different models describing one truth. Another important issue is the fact that all atmospheric models have finite resolution, and this can be modeled conveniently by taking that the models have less gridpoints than the truth – M in total.

We run simultaneously the truth and the models and estimate the quality of the models by comparison of the output of the models with that of the truth which is considered as obtained by "measurements". For short-term prediction purposes, one can take the values from the truth Z_n as initial conditions for the models, with small noise added to incorporate the measurement error. One can verify that in this toy example the models are good for short-term prediction by visual comparison of the solutions of the models and the truth (see Fig. 1).

For chaotical dynamical systems, an important property related to the predictability is the divergence between the phase points originating from two nearby initial conditions. It is well established that the trajectories of nonlinear dynamical systems starting from very close initial conditions typically separate exponentially fast. But that happens if both trajectories are obtained from the same dynamical system. However, in reality the truth and the system belong to different classes of functions (Smith, 1997) and the divergence between them must not be exponential in the beginning. Comparison of the solutions of a model and the truth (and their mismatch) is the proper measure of the predictability power of the model. In Fig. 2 are shown root-mean-square difference between the truth and three different models of it – model with the same complexity M3, and two models of class II with different number of gridpoints. As can be seen, only the difference is exponentially increasing only when the model is the same class of function as the truth. In the lower panel of the same figure are shown the same differences but in linear plot, where it is clear that the growth of the error is linear (Orrell et al., 2001).

This result can be easily explained by comparison of the tendencies of the truth and the model. In this toy case, as well as in the reality, the state of the truth and its tendency can be represented (projected) in the space of the model. The tendencies are velocity vectors defined in the space of the model. It is highly unlikely to expect that the (projected) vector of the truth will be almost all the time equal to the vector of any model – there is at least some small angle between them and small intensity difference – there is always tendency error. This means that the states starting even from the same initial condition will evolve along those vectors, and for short time one can assume that the tendency vectors are constant. With this reasoning the divergence of the states of the truth and the model can be represented with separation of two particles moving along two intersecting lines with constant velocities and starting from the intersection point. Then one can easily show that the distance between those points will be linearly growing function of time. Linear growth is obtained as well when the vectors are parallel but have different intensities, which is also more reasonable to assume than the case of equal intensities.

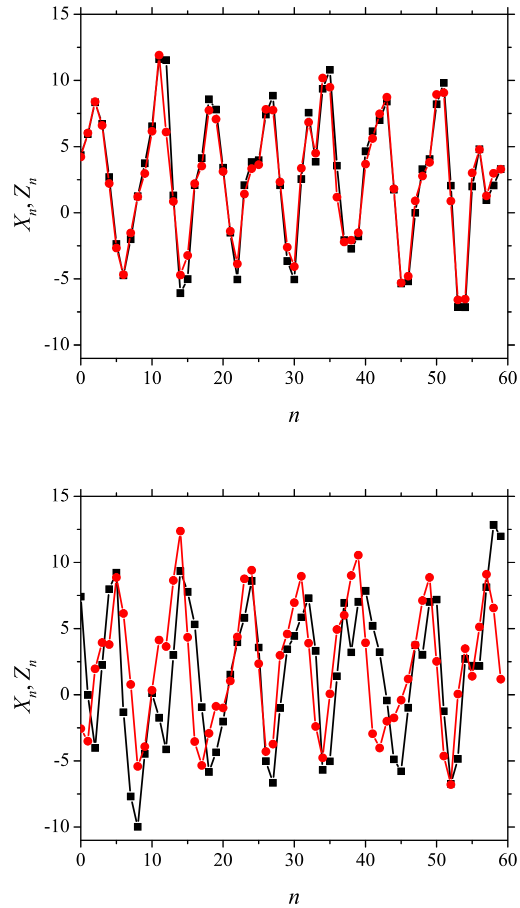

Fig. 1. Solution profiles of the truth – Lorenz 96 model III (in black) and its simpler representation – Lorenz 96 model II (in red). The upper panel shows the solutions at moment 0.2 after initiation at close initial solution, while at the bottom one the moment when the solutions are taken is $t = 1$.

3 Weighted averaging

The existence of tendency error suggests that possibly a linear combination of the tendencies of the models can give a tendency closer to the one of the truth. This is the idea behind the improvement of models. For every grid point of the models, the improved model is a model with a tendency that is weighted combination of the tendencies of the individual models. More formally, if the tendency at grid point n of the

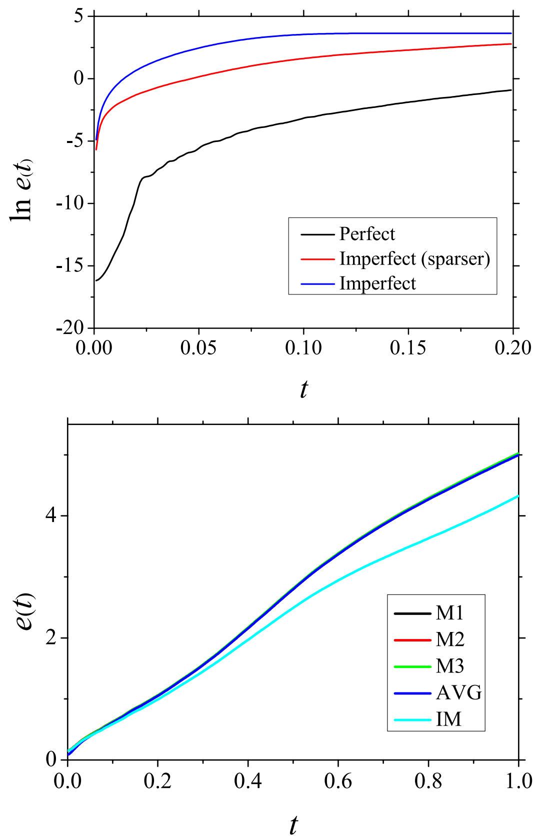

Fig. 2. Growth of the error between the model and the truth. In the upper figure the ordinate axis has logarithmic scale and verifies the exponential divergence between the trajectories when the model is perfect (black curve). From the lower figure, it is clear that the error grows linearly – the improved model's curve is the lowest one.

model μ is

$$\dot{X}_n^{\mu} = T_n^{\mu} = [X^{\mu}, X^{\mu}]_{K,n} - X_n^{\mu} + F_n^{\mu}, \tag{7}$$

then the improved model at grid point n has tendency

$$\dot{X}_n^{s} = T_n^{s} = \sum_{\mu} w_n^{\mu} T_n^{\mu}. \tag{8}$$

Then to obtain an improved model (assuming that the models are made as good as possible), the designer is left to find optimal values of the weights w_n^{μ}. Because of non-linear nature of the systems, the tendencies of the truth and the models are not strongly correlated, and one approach for determination of optimal weights is the statistical one. Assuming that the main source of limitation of the prediction is the error in determination of tendency of the truth, one should use the tendency error as a measure of quality of a improved model. The average tendency error is given by

$$D = \langle \sum_{n=0}^{M-1} |T_n^t - T_n^s|^2 \rangle = \sum_{n=0}^{M-1} \langle \left| T_n^t - \sum_{\mu} w_n^{\mu} T_n^{\mu} \right|^2 \rangle, \tag{9}$$

where the tendency of the truth T_n^t is given by the RHS of Eq. (3) and angle brackets denote time average. Optimal weights (according to the training set of data) are obtained by differentiating the last expression with respect to the weights

$$\frac{\partial D}{\partial w_n^{\mu}} = \frac{\partial \langle \left| T_n^t - \sum_{\nu} w_n^{\nu} T_n^{\nu} \right|^2 \rangle}{\partial w_n^{\mu}} = 2 \langle T_n^{\mu} \left(T_n^t - \sum_{\nu} w_n^{\nu} T_n^{\nu} \right) \rangle = 0. \tag{10}$$

To simplify the notations, one could introduce the covariances between the tendencies

$$C_n^{\mu,\nu} = \langle T_n^{\nu} T_n^{\mu} \rangle,$$
$$C_n^{\mu,t} = \langle T_n^{t} T_n^{\mu} \rangle. \tag{11}$$

Then the equations for optimal weights at every grid point n become linear:

$$\sum_{\nu} C_n^{\mu,\nu} w_n^{\nu} - C_n^{\mu,t} = 0, \tag{12}$$

where the factor 2 was removed with cancelation. The system of Eq. (12) can be written more succinctly by using matrix of covariances between the models $\mathbf{C}_n$, vector of covariances with the truth $\boldsymbol{c}_n$ and vector of weights $\boldsymbol{w}_n$ at every grid point n:

$$\boldsymbol{C}_n \boldsymbol{w}_n = \boldsymbol{c}_n. \tag{13}$$

The linear regression technique suggests adding regularization term to avoid over-fitting of the parameters – weights in our case (Bishop, 2006). Then instead of minimizing only the average tendency error (Eq. 9), the function to be minimized has the form

$$D + \lambda \sum_{n,\nu} (w_n^{\nu})^2, \tag{14}$$

where λ is the regularization coefficient. The minimization is obtained again by taking partial derivatives with respect to the weights. Then the system of equations for the weights (Eq. 12) will have a slightly modified form:

$$\sum_{\nu} \left(C_n^{\mu,\nu} - \lambda \right) w_n^{\nu} - C_n^{\mu,t} = 0. \tag{15}$$

Using matrix notation, one concludes that for every grid point n the following matrix equation should be solved:

$$(\mathbf{C}_n - \lambda \mathbf{I})\, \boldsymbol{w}_n = \boldsymbol{c}_n, \tag{16}$$

which has the solution

$$\boldsymbol{w}_n = (\mathbf{C}_n - \lambda \mathbf{I})^{-1} \boldsymbol{c}_n. \tag{17}$$

4 Numerical experiments

To perform numerical experiments on a PC, we have considered as truth the Lorenz model III with $N = 960$ gridpoints. As said above the forcing term was randomly perturbed to account for spatial inhomogeneity of the atmospheric forcing (see Eq. 6). Because any model of the atmosphere is its coarse representation, we have taken $M = 60$ gridpoints for the models. Individual models of the truth differ in forcing terms, which were obtained with random perturbation of those of the truth at the corresponding points. This difference should represent the differences between the models designed at different meteorological centers. For comparison of the models and the truth, it was considered that the measurements are performed only at the gridpoints of the models. In calculations of the covariances, we have assumed that the tendency of the truth is known. In reality the tendency of the atmosphere can be estimated with interpolation and the estimation will be different from the true value. To incorporate this fact, we have added noise to the tendency of the truth. For short-term forecasting purposes, the models should be initiated from the state of the truth, and again with some perturbation that models the observation noise.

First verification of the models can be done with visual comparison of the solution profiles at some moment. In Fig. 1 are shown the fields of the truth and one model at the moment $t = 0.2$, which corresponds approximately to one day according to Lorenz (we remind the reader that the truth and the model were started with close initial conditions). As can be seen for short times, the model's profile follows the pattern of the truth. Later, the states of the solutions are much more different.

The distance between the solutions of the models and the truth is a measure of the predictability of the state of the reality. Starting with the same initial condition, the mismatch, or prediction error, between the model μ and the truth at moment t is

$$e^{\mu}(t) = \sqrt{\sum_{m=0}^{M-1} |Z_m(t) - X_m^{\mu}(t)|^2}. \quad (18)$$

In the bottom panel of Fig. 2 are shown the prediction errors, for the individual models, their average[1], and of the improved model. It is clear that the improved model outperforms all of them.

Within meteorological scientific community, a measure for estimation of the predictability range of a model is the anomaly correlation – AC (Allgaier et al., 2012). AC for two time series is simply defined as a correlation between the two variables at the same moment. It measures how much, on average, the deviation from their respective means at the same moment is at same direction and with similar magnitude. The AC between the truth and any model μ is given by

[1]The average field of the individual models has values that are simple average of the fields of the individual models $X_m^{\text{av}} = (\sum_{\mu} X_m^{\mu})/3$.

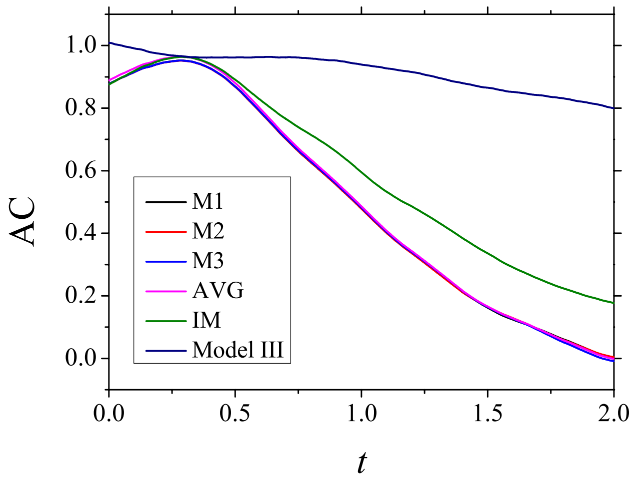

Fig. 3. Anomaly correlation between the truth and the models. Top curve (in blue) is for model that has same complexity as the truth – model III. The middle curve (in green) corresponds to the improved model, and lower curves (almost indistinguishable) are for the individual models and average output of them.

$$\text{AC}^{\mu} = \frac{\sum_{m=0}^{M-1} \langle (Z_m - \langle Z_m \rangle)(X_m^{\mu} - \langle X_m^{\mu} \rangle) \rangle}{\sqrt{\sum_{m=0}^{M-1} (Z_m - \langle Z_m \rangle)^2} \sqrt{\sum_{m=0}^{M-1} (X_m^{\mu} - \langle X_m^{\mu} \rangle)^2}}. \quad (19)$$

The angular brackets in the last equation again denote time averaging – in this case averaging is performed in the examination period. The predictability range extends to the moment when AC falls below value 0.6. In Fig. 3 we show the AC for the individual models, their average (calculated in the same way as for the prediction error) and the improved model. By using the threshold $\text{AC} = 0.6$ as a criterion for the predictability, it is obtained that the improved model extends the predictability window for 17 %. In the same figure is shown also the AC between the truth and a model that is the same as truth, started with close initial condition. That curve has typical behavior because it is decreasing. We think that the other AC curves first increase and then decrease due to the structural difference between the truth and the models.

We have also tried to estimate whether the improved model will give "climatology" closer to that of the truth, as climatology can be considered time averages of the fields at certain gridpoints $\langle Z_n \rangle$ and $\langle X_n^{\mu} \rangle$. After calculating average mismatch $(\langle Z_n \rangle - \langle X_n^{\mu} \rangle)^2$ for whole space, we found that it is not smaller for the improved model than the individual models. In that case, the average output of the models X_n^{av} has best performance. This can be understood because the improved model "is trained" for short-term forecasts – it is based on optimization of the tendency, which is a short-term property.

5 Conclusions

In this work we have shown a proof-of-concept that a combination of different models of one-dimensional scalar atmospheric quantity can be better short-term forecaster than any of them. The idea was tested on the Lorenz 96 toy model because it is simple enough, but also nontrivial and shares some basic properties with the real atmosphere. The atmosphere has (and always will) higher dimensionality than any model of it, and so in this example we have taken that the models have less degrees of freedom than the truth. That causes the tendencies of the models and the truth to become different and possibly the largest factor that contributes to divergence of the prediction and the realization of the truth. Then a weighted combination of the tendencies of the individual models with weights learned by using the past observations can be used to construct an improved model for the short-term prediction. As observations in the learning were used the tendencies, which are not available for any reality, and the atmosphere in particular. It should be estimated with appropriate techniques. To incorporate that limitation in this toy model, we have added a noise. However, the noise that will inevitably emerge by using estimation of the tendency of the atmosphere can be even larger, and thus limit (or even eliminate) the improvement of short-term forecast in this way.

To our opinion there are two main lines of future research related to this work. The first one is towards the search for different techniques for combination of the individual models. One possible option is the coupling of the variables of state, or subset of them. For more complex atmospheric models, exchange of fluxes is already in use – coupling of atmospheric and ocean models. However, coupling of different atmospheric models, to our knowledge, is not applied yet. The interaction structure between the models will influence the strategies of searching for the best coupling parameters. Besides different techniques from machine learning, expert knowledge is welcomed also. The weights should not be constant in time, but time (e.g., seasonally) dependent instead. Or they can be adjusted and improved all the time because the measurement data are accumulating.

The second direction for further research is attempt to apply these results in more real atmospheric models, or even for those that are used for numerical weather prediction. The main obstacle can be estimation of the tendency of the atmosphere. We think this kind of combination of state-of-the-art models is worth testing because of importance of the weather prediction. In the worst case the weights can have unit values for the best member of the ensemble, and thus there is at least one combination of weights that is as good as the best individual model. We expect that mixing the tendencies can lead to improvement of the numerical weather prediction. Another issue is improvement of the projections of the future climate. However, as our results have shown, maybe other techniques for optimization of the connection parameters should be applied to construct an interactive ensemble that will outperform the individual members.

Acknowledgements. This work was partially supported by project ERC Grant # 266722 (SUMO project).

Edited by: J. Kurths
Reviewed by: two anonymous referees

References

Allgaier, N. A., Harris, K. D., and Danforth, C. M.: Empirical correction of a toy climate model, Phys. Rev. E, 85, 026201, doi:10.1103/PhysRevE.85.026201, 2012.

Bishop, C. M.: Pattern Recognition and Machine Learning, Springer, 2006.

Bjerknes, V.: Dynamic Meteorology and hydrography, Part II. Kinematics, Gibson Bros., Carnegie Institute, New York, 1911.

Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D. Chen, D. H., Ebert, B., Fuentes, M., Hamill, T. M., Mylne, K., Nicolau, J., Paccagnella, T., Park, Y. Y., Parsons, D., Raoult, B., Schuster, D., Dias, P. S., Swinbank, R., Takeuchi, Y., Tennant, W., Wilson, L., and Worley, S.: The THORPEX Interactive Grand Global Ensemble, Bull. Am. Meteorol. Soc., 91, 1059–1072, 2010.

Judd, K. and Smith, L. A.: Indistinguishable states II: The imperfect model scenario, Physica D, 196, 224–242, 2004.

Judd, K., Reynolds, C. A., Rosmond, T. E., and Smith, L. E.: The Geometry of Model Error, J. Atmos. Sci., 65, 1749–1772, 2008.

Kirtman, B. P. and Shukla, J.: Interactive coupled ensemble: A new coupling strategy for CGCMs, Geophys. Res. Lett., 29, 1367, doi:10.1029/2002GL014834, 2002.

Lewis, J. M.: Roots of Ensemble Forecasting, Mon. Weather Rev., 133, 1865–1885, 2005.

Lorenz, E. N.: Deterministic nonperiodic flow, J. Atmos. Sci. 20, 130–141, 1963.

Lorenz, E. N.: Predictability – A problem partly solved, Proc. Seminar on Predictability, Reading, Berkshire, UK, ECMWF, 1–18, 1996.

Lorenz, E. N.: Designing chaotic models, J. Atmos. Sci., 62, 1574–1587, 2004.

Mirchev M., Duane G., Tang W. K. S., and Kocarev L.: Improved modeling by coupling imperfect models, Commun. Nonlinear Sci. Numer. Simul., 17, 2741–2751, doi:10.1016/j.cnsns.2011.11.003, 2012.

Orrell, D.: Model Error and Predictability over Different Timescales in the Lorenz '96 Systems, J. Atmos. Sci., 60, 2219–2228, 2003.

Orrell, D., Smith, L., Barkmeijer, J., and Palmer, T. N.: Model error in weather forecasting, Nonlin. Processes Geophys., 8, 357–371, doi:10.5194/npg-8-357-2001, 2001.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corrazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A.: A local ensemble Kalman filter for atmospheric data assimilation, Tellus A, 56, 415–428, 2004.

Smith, L. A.: The maintenance of uncertainty, in: Past and Present Variability in the Solar-Terrestial System: Measurement, Data Analysis and Theoretical Models, edited by: Castagnoli, G. C. and Provenzale, A., vol. CXXXIII of International School of

Physics “Enrico Fermi”, Il Nuovo Cimento, Bologna, 177–246, 1997.

Smith, L. A., Ziehmann, C., and Fraedrich, K.: Uncertainty dynamics and predictability in chaotic systems, Q. J. R. Meteorol. Soc., 125, 2855–2886, 1999.

Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., Knight, S., Martin, A., Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe, A. J., and Allen, M. R.: Uncertainty in predictions of the climate response to rising levels of greenhouse gasses, Nature, 433, 403–406, 2005.

van den Berge, L. A., Selten, F. M., Wiegerinck, W., and Duane, G. S.: A multi-model ensemble method that combines imperfect models through learning, Earth Syst. Dynam., 2, 161–177, doi:10.5194/esd-2-161-2011, 2011.

Wiegerinck, W., Burgers, W., and Selten, F. M.: SUMO – Supermodeling by imperfect models, Technical Report, Workpackage 2, 2011.