GEOPHYSICAL RESEARCH LETTERS, VOL. ???, XXXX, DOI:10.1029/,

A stochastic method for improving seasonal predictions

L. Batté,[1] M. Déqué,[1]

L. Batté, CNRM-GAME, Météo France, 42 avenue G. Coriolis, 31057 Toulouse, FRANCE (lauriane.batte@meteo.fr)

M. Déqué, CNRM-GAME, Météo France, 42 avenue G. Coriolis, 31057 Toulouse, FRANCE (michel.deque@meteo.fr)

[1]CNRM-GAME, Toulouse, France.

Ensemble seasonal forecasts during boreal winter suffer from insufficient spread and systematic errors. In this study we suggest a new stochastic dynamics method to address both issues at a time. Our technique relies on random additive corrections of initial tendency error estimates of the atmospheric component of the CNRM-CM5.1 global climate model, using ERA-Interim as a reference over a 1979-2010 hindcast period. The random method improves deterministic scores for 500-hPa geopotential height forecasts over the Northern Hemisphere extratropics (NH Z500), and increases the ensemble spread. An optimal method consisting in drawing the error corrections within the current month of the hindcast period illustrates the high potential of future improvements, with NH Z500 anomaly correlation reaching 0.65 and North Atlantic Oscillation index correlation 0.71 with ERA-Interim. These substantial improvements using current year corrections pave the way for future forecasting methods using classification criteria on the correction population.

1. Introduction

Seasonal prediction using coupled general circulation models (GCMs) has been an active field of research over the last two decades. International research efforts such as the European Commission-funded DEMETER [*Palmer et al.*, 2004] and ENSEMBLES [*Weisheimer et al.*, 2009; *Doblas-Reyes et al.*, 2009] projects as well as the APEC Climate Center-sponsored CliPAS project [*Wang et al.*, 2009] illustrated the potential of state-of-the-art numerical climate models in forecasting temperature and geopotential, and to a lesser extent precipitation, at a seasonal timescale. Predictability is generally higher over the Tropics, but models show positive skill with respect to climatology over some midlatitudinal regions. Most model ensembles suffer from systematic errors and lack of spread. Multi-model techniques pooling together predictions from several models address both issues : some systematic errors are cancelled out provided that individual model errors are different, and reliability is improved [*Hagedorn et al.*, 2005]. However, the success of a multi-model ensemble technique relies mainly on the quality of the individual models used. In addition, if a model has insufficient spread and a large prediction error over a given region, it will lead the multi-model towards a wrong prediction.

In recent years a variety of stochastic perturbation methods has been implemented in atmospheric models to account for model error, both for short-term ensemble predictions and monthly-to-seasonal forecasts using these models as the atmospheric component of an earth-system model. *Buizza et al.* [1999] introduced random perturbations of model physical tendencies into the ECMWF ensemble prediction system. An additional scheme called Stochastic Kinetic Energy Backscatter (SKEB) algorithm is used by ECMWF to scatter kinetic energy dissipated by the model at the sub-grid scale back to larger scales [*Shutts*, 2005], and *Berner et al.* [2008] highlights the reduction of systematic error and improvements of most deterministic and probabilistic skill scores over differents regions at a seasonal time scale due to this algorithm. SKEB is used alongside a perturbed parameters scheme described in *Bowler et al.* [2008] in the Met Office's GloSea4 seasonal forecast model [*Arribas et al.*, 2011]. Similar stochastic physics schemes are also used for medium-range forecasts in the Canadian ensemble prediction system [*Charron et al.*, 2010].

In the present study, an alternative stochastic perturbation technique is applied to the CNRM-CM5.1 GCM [*Voldoire et al.*, 2012] for seasonal forecasting. Predictions are stochastically corrected by adding randomly drawn initial tendency residuals to the temperature, specific humidity and vorticity fields in the prognostic equations of the ARPEGE-Climat v5.2 atmospheric model component. The initial tendency residuals are estimated using a nudging technique as described in *Kaas et al.* [1999] and *Guldberg et al.* [2005]. Several past studies such as *Yang and Anderson* [1999], *Barreiro and Chang* [2004] and *Guldberg et al.* [2005] have suggested that correcting systematic errors in atmospheric or coupled ocean-atmosphere GCMs reduce model bias with some impact on seasonal prediction skill. However, *Guldberg et al.* [2005] found that systematic error correction in a previous version of ARPEGE-Climat showed no improvement over the Tropics and the Northern Hemisphere. The originality of the method presented here relies on the stochasticity of the error corrections. A more detailed description of the stochastic dynamics technique is given in section 2, and results are shown in section 3. They illustrate the significant gain in seasonal forecasting skill during Northern Hemisphere winter. An upper limit for possible future improvements using this method is also shown.

2. Stochastic Dynamics Method

The stochastic dynamics method implemented in the ARPEGE-Climat v5.2 atmospheric model for seasonal forecasts is an additive stochastic perturbation of three prognostic ARPEGE variables $\mathbf{X}$: temperature, specific humidity and vorticity, following equation 1. $\mathbf{M}(\mathbf{X}(t), t)$ represents the evolution of variable $\mathbf{X}$ due to the initial ARPEGE-Climat model formulation, and $\delta\mathbf{X}_t$ is the stochastic perturbation.

$$\mathbf{X}(t + \Delta t) = \mathbf{X}(t) + \mathbf{M}(\mathbf{X}(t), t) + \delta\mathbf{X}_t \tag{1}$$

Our method derives from *Guldberg et al.* [2005] and consists in using the nudging technique to estimate initial tendency errors of ARPEGE-Climat v5.2 and then perturbing a seasonal forecast with random initial tendency error corrections drawn within these estimates. The stochastic dynamics method follows three steps. The first step is to run the CNRM-CM5.1 model during 32 years (1979-2010), nudging it towards the ECMWF ERA-Interim reanalysis data [*Dee et al.*, 2011]. ERA-Interim data is re-interpolated on the ARPEGE-Climat reduced gaussian grid. Prognostic variables temperature, specific humidity and vorticity are relaxed towards the ERA-Interim fields with relaxation times of a day for temperature and specific humidity and 6 hours for vorticity. This run provides initial conditions on November 1st 1979 to 2010 (for boreal winter forecasts) for each component of CNRM-CM5.1.

In a second step, a four-member ensemble is implemented for each November-December-January-February season (NDJF) of the 1979-2010 period. This second run is relaxed more weakly towards ERA-Interim and started with initial conditions from the first run, thus reducing spin-up effects due to differences between ERA-Interim and model climatology. Relaxation times are selected close to one month for temperature and specific humidity, and ten days for vorticity. A vertical profile for relaxation coefficients is introduced in the five lowest levels of the model so as to tune relaxation down to zero and avoid inconsistencies at the surface. Differences between ERA-Interim fields and each member for the three relaxed variables are stored daily. The opposite of these fields, thus corresponding to model corrections towards ERA-Interim, make up the $\{\delta\mathbf{X}\}$ population from which the perturbations are drawn in forecast mode.

The third step consists in the actual retrospective forecast, started with initial conditions each November 1st from the first run and with perturbations drawn from the $\{\delta\mathbf{X}\}$ population designed in the second step of the method. In this study perturbations were drawn within the corresponding calendar month, meaning that $\{\delta\mathbf{X}\}$ was in fact separated in four bins for NDJF coherent with the forecast lead-time. A different $\delta\mathbf{X}$ was drawn for each ensemble member every six hours of the forecast. Perturbations for temperature, specific humidity and vorticity are drawn together, and correspond to an error correction for a given day of the second step re-forecast. This ensures that perturbations are coherent between the three corrected fields, and avoids partially cancelling out the effects of one correction with that of another field.

3. Experiments and Results

Three sets of seasonal re-forecasts of December to February (DJF) 1979-80 to 2010-11 were run with 15 ensemble members:

1. The reference seasonal forecast ensemble (REF) was perturbed with random $\delta\mathbf{X}$ drawn from the initial tendency error correction population only at the initial time step.

2. A random stochastic dynamics ensemble (SD_RAND) was perturbed with $\delta\mathbf{X}_t$ at each time step.

3. An optimal stochastic dynamics ensemble (SD_OPT) was perturbed with $\delta\mathbf{X}_t$ at each time step drawn in the same month and year as the actual forecast.

The SD_OPT experiment cannot be implemented for operational forecasts, since initial tendency errors can only be estimated for a set of hindcasts. Perturbations are consistent with the errors the model makes in a given month. Therefore, results for SD_OPT determine the upper limit for scores using this stochastic perturbation technique, provided that corrections are relevant to the model initial tendency errors at a given time.

The impact of the stochastic dynamics method on DJF 500 hPa geopotential height (Z500) bias over the Northern Hemisphere is shown in figure 1. The negative bias over the polar region is reduced in SD_RAND, and Z500 bias gradients over the northern Pacific and northern Atlantic are less pronounced. SD_OPT biases are very similar to SD_RAND (not shown). Figure 2 shows anomaly correlation coefficients (ACC) for DJF Z500 over the Northern Hemisphere extra-tropics (30 to 75 degrees North) for each forecast ensemble. The random stochastic dynamics method improves anomaly correlation for 22 out of 32 seasons. The associated binomial test shows that this improvement is statistically significant ($p = 0.025$). While the REF ensemble yields correlation values lower than 0.2 for 15 seasons, correlation remains lower than this threshold for only 8 seasons with the SD_RAND ensemble. SD_OPT anomaly correlation scores reach over 0.6 for 19 seasons and are lower than 0.4 for only 4 seasons. This suggests that an appropriate set of perturbations in a given season could lead to significant improvements in forecasting skill.

Mean ACC values for different variables and regions were calculated for the three ensembles and are listed in table 1. Mean ACC is considerably improved with stochastic dynamics for Z500 over the Northern Hemisphere extra-tropics, in coherence with results shown earlier. Results over the Tropics for 2-meter temperature (T2m) and precipitation and the Niño 3.4 region for T2m exhibit no significant impact of the stochastic dynamics method on mean ACC scores for SD_RAND, whereas SD_OPT improves precipitation and T2m scores over the Tropics.

Improvement over the Northern Hemisphere extra-tropics is also found when looking at monthly root mean square error (RMSE) of the forecasts over the 1979-2010 time period. Figure 3 illustrates the improvement of the spread-to-skill ratio of the forecast ensemble for NH Z500. While RMSE is reduced by over 15 meters in months 3 and 4 of the forecast, the SD_RAND ensemble also has a higher spread during the first two months, and similar spread in the following two months. The stochastic dynamics method therefore improves model error and dispersion, as intended. SD_OPT has the same spread as SD_RAND, with an ensemble spread larger than the RMSE after a 2-month lead.

Skill was further assessed over the Euro-Atlantic region by investigating model performance in forecasting the North Atlantic Oscillation (NAO). Following a method similar to *Doblas-Reyes et al.* [2003], the NAO is defined as the leading empirical orthogonal function (EOF) of December to February monthly Z500 ERA-Interim data from 1979 to 2010 over the region 20°N-80°N and 80°W-40°E. Model NAO indexes are calculated by projecting monthly grid point anomalies for each member onto this EOF. Forecasts and ERA-Interim verification series are standardized in cross-validation mode. The introduction of stochastic dynamics has little impact on the ensemble spread of the forecasts at a seasonal time scale. The SD_RAND ensemble has slightly higher skill than REF in forecasting the NAO index, with a correlation of 0.36 versus 0.32 between the ensemble mean index and the reference ERA-Interim index. The SD_OPT ensemble exhibits significant improvement with a correlation of 0.71 with ERA-Interim.

Probabilistic skill was evaluated with a ranked probability score (RPS) for tercile prediction defined following *Toth et al.* [2003] as the average of Brier Scores for a given variable remaining below the climatological terciles. The RPS ranges between 0 (perfect forecast) and 1 and consists in a sum over the 32 seasons of quadratic distances in probabilistic space between forecasts and observations (worth 0 or 1 whether the event occurs or not a given season). Reliability, resolution [*Murphy*, 1973] and RPS scores are calculated as in *Batté and Déqué* [2011] for each grid point over land and averaged over the region of interest. Results for T2m terciles over NH land grid points and NH Z500 are shown in table 2. A ranked probability skill score is defined as $\mathrm{RPSS} = 1 - \mathrm{RPS}/\mathrm{RPS}_c$ where RPS_c is the climatology RPS. Similar scores are found for ensembles REF and SD_RAND, which outperform climatological forecasts over the region, yielding positive RPSS values. The improvement in scores noted for SD_OPT is mainly due to an increase in resolution, which evaluates the ability of the model to separate events that have different probabilities of occurence.

4. Conclusion and Discussion

This study presents an original technique for stochastic perturbations combining the assets of random perturbation and systematic error correction in coupled models used for seasonal forecasts. Re-forecasts of DJF 1979-2010 using this method with the CNRM-CM5.1 GCM show enhanced performance over the Northern Hemisphere for 500hPa geopotential height, with similar skill over the Tropics. RMSE and anomaly correlation coefficients for Z500 show that random stochastic perturbations as designed in our study can enhance scores and improve the model spread-to-skill ratio. These improvements are triggered by a reduced seasonal bias consistent with previous studies that corrected average errors, and an enhanced ensemble spread consistent with other stochastic techniques.

Results with an ensemble using optimal corrections drawn from the current forecast month suggest room for improvement in seasonal forecasting skill, provided that corrections are drawn from a population that is representative of the common initial tendency errors of the current season. Correlation coefficients for the NAO index with the optimal ensemble reach 0.7 and therefore illustrate the potential of such a technique, as long as an appropriate classification of the correction population is found. Further work should therefore focus on exploring classification criteria for the perturbation population based on the state of the ocean or the atmosphere, using analogues to classify perturbations according to tropical sea surface temperature or weather regimes as in *D'Andrea and Vautard* [2000]. It is worth mentioning that although RMSE was further reduced with optimal perturbations, ensemble spread remained very close to the random perturbation ensemble. A concise study of probabilistic skill showed that ranked probability score improvements with the optimal ensemble relied mainly on increased resolution. Lack of improvement in reliability could be corrected by multi-model forecasting. Given the current impact of our method on model spread, other stochastic perturbations with a longer time scale could be included in the model. Future experiments will study the impact of the perturbation frequencies and drawing several successive chronological corrections on model spread and skill.

Acknowledgments.

ERA-Interim data used in this study were provided by ECMWF. We are grateful to two anonymous reviewers who helped us improve this manuscript.

References

Arribas, A., M. Glover, A. Maidens, K. Peterson, M. Gordon, C. MacLachlan, R. Graham, D. Fereday, J. Camp, A. A. Scaife, P. Xavier, P. McLean, A. Colman, and S. Cusack, The GloSea4 ensemble prediction system for seasonal forecasting, *Mon. Wea. Rev.*, *139*, 1891–1910, doi:10.1175/2010MWR3615.1, 2011.

Barreiro, M., and P. Chang, A linear tendency correction technique for improving seasonal prediction of SST, *Geophys. Res. Lett.*, *31*, L23,209, doi:10.1029/2004GL021148, 2004.

Batté, L., and M. Déqué, Seasonal predictions of precipitation over Africa using coupled ocean-atmosphere general circulation models: skill of the ENSEMBLES project multi-model ensemble forecasts, *Tellus*, *63A*, 283–299, doi:10.1111/j.1600-0870.2010.00493.x, 2011.

Berner, J., F. Doblas-Reyes, T. Palmer, G. Shutts, and A. Weisheimer, Impact of a quasi-stochastic cellular automaton backscatter scheme on the systematic error and seasonal prediction skill of a global climate model, *Philos. Trans. R. Soc. London, Ser. A*, *366*(1875), 2559–2577, doi:10.1098/rsta.2008.0033, 2008.

Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, The MOGREPS short-range ensemble prediction system, *Q. J. R. Meteorolog. Soc.*, *134*, 703–722, 2008.

Buizza, R., M. Miller, and T. Palmer, Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System, *Q. J. R. Meteorolog. Soc.*, *125*, 2887–2908, 1999.

Charron, M., G. Pellerin, L. Spacek, P. Houtekamer, N. Gagnon, H. Mitchell, and L. Michelin, Toward random sampling of model error in the Canadian Ensemble Prediction System, *Mon. Wea. Rev.*, *138*, 1877–1901, 2010.

D'Andrea, F., and R. Vautard, Reducing systematic errors by empirically correcting model errors, *Tellus*, *52A*, 21–41, 2000.

Dee, D., S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. Beljaars, L. van de Berg, L. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. Geer, L. Haimberger, S. Healy, H. Hersbach, E. Hólm, L. Isaksen, P. Kallberg, M. Köhler, M. Matricardi, A. McNally, B. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart, The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorolog. Soc.*, *137*, 553–597, doi:10.1002/qj.828, 2011.

Doblas-Reyes, F., V. Pavan, and D. Stephenson, The skill of multi-model seasonal forecasts of the wintertime North Atlantic Oscillation, *Clim. Dyn.*, *21*, 501–514, doi: 10.1007/s00382-003-0350-4, 2003.

Doblas-Reyes, F., A. Weisheimer, M. Déqué, N. Keenlyside, M. MacVean, J. Murphy, P. Rogel, D. Smith, and T. Palmer, Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts, *Q. J. R. Meteorolog. Soc.*, *135*, 1538–1559, doi: 10.1002/qj.464, 2009.

Guldberg, A., E. Kaas, M. Déqué, S. Yang, and S. Vester Thorsen, Reduction of systematic errors by empirical model correction: impact on seasonal prediction skill, *Tellus*, *57A*, 575–588, 2005.

Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer, The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept, *Tellus*, *57A*, 219–233, 2005.

Kaas, E., A. Guldberg, W. May, and M. Déqué, Using tendency errors to tune the parameterisation of unresolved dynamical scale interactions in atmospheric general circulation models, *Tellus*, *51A*, 612–629, 1999.

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

Palmer, T., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Délécluse, M. Déqué, E. Díez, F. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J.-F. Guérémy,

R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonnave, V. Marletto, A. Morse, B. Orfila, P. Rogel, J.-M. Terres, and M. Thomson, Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER), *Bull. Am. Meteorol. Soc.*, *85*, 853–872, 2004.

Shutts, G., A kinetic energy backscatter algorithm for use in ensemble prediction systems, *Q. J. R. Meteorolog. Soc.*, *131*, 3079–3102, doi:10.1256/qj.04.106, 2005.

Toth, Z., O. Talagrand, G. Candille, and Y. Zhu, Probability and ensemble forecasts, in *Forecast Verification, A Practitioner's Guide in Atmospheric Science*, edited by I. Joliffe and D. Stephenson, pp. 137–163, John Wiley & Sons Ltd, 2003.

Voldoire, A., E. Sanchez-Gomez, D. Salas y Mélia, B. Decharme, C. Cassou, S. Sénési, S. Valcke, I. Beau, A. Alias, M. Chevallier, M. Déqué, J. Deshayes, H. Douville, E. Fernandez, G. Madec, E. Maisonnave, M.-P. Moine, S. Planton, D. Saint-Martin, S. Szopa, S. Tyteca, R. Alkama, S. Belamari, A. Braun, L. Coquart, and F. Chauvin, The CNRM-CM5.1 global climate model: Description and basic evaluation, *Clim. Dyn.*, doi:10.1007/s00382-011-1259-y, 2012.

Wang, B., J.-Y. Lee, I.-S. Kang, J. Shukla, C.-K. Park, A. Kumar, J. Schemm, S. Cocke, J.-S. Kug, J.-J. Luo, L. Zhou, B. Wang, X. Fu, W.-T. Yun, O. Alves, E. K. Jin, J. Kinter, B. Kirtman, T. Krishnamurti, N. C. Lau, W. Lau, P. Liu, P. Pegion, T. Rosati, S. Schubert, W. Stern, M. Suarez, and T. Yamagata, Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004), *Clim. Dyn.*, *33*, 93–117, doi:10.1007/s00382-008-0460-0, 2009.

Weisheimer, A., F. Doblas-Reyes, T. Palmer, A. Alessandri, A. Arribas, M. Déqué, N. Keenlyside, M. MacVean, A. Navarra, and P. Rogel, ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions — skill and progress beyond DEMETER in forecasting tropical Pacific SSTs, *Geophys. Res. Lett.*, *36*, doi: 10.1029/2009GL040896, 2009.

Yang, X.-Q., and J. Anderson, Correction of systematic errors in coupled GCM forecasts, *J. Climate*, *13*, 2072–2085, 1999.

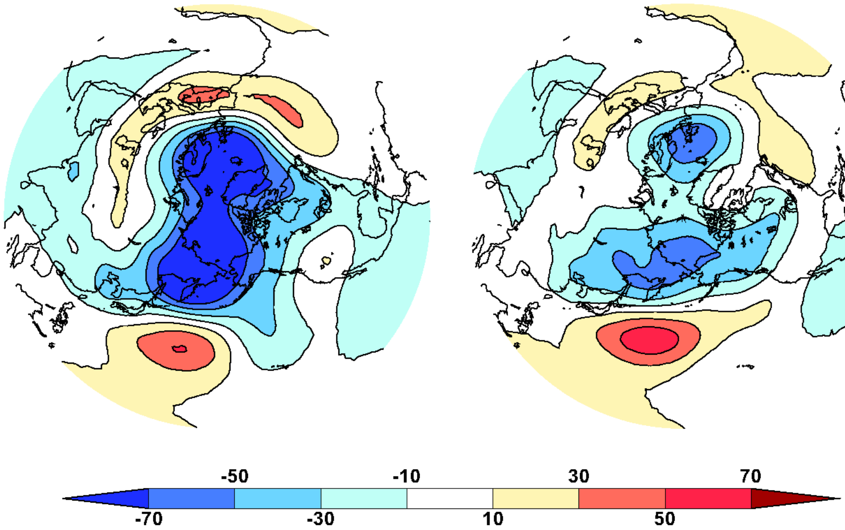

Figure 1. DJF NH Z500 mean bias (in meters) for ensembles REF (left) and SD_RAND (right).

Table 1. Mean ACC values for REF, SD_RAND and SD_OPT. Statistical significance of differences between the SD ensembles and REF are tested using a binomial test for season ACC scores. Bold scores are significantly better than REF at a 95% level.

Region	Variable	REF	SD_RAND	SD_OPT
NH [a]	Z500	0.25	**0.37**	**0.65**
Tropics [b]	Precipitation	0.45	0.45	**0.52**
Tropics	T2m	0.47	0.47	**0.51**
Niño 3.4 [c]	T2m	0.83	0.81	0.82

[a] 30°N-75°N

[b] 23°N-23°S

[c] 170°W-120°W and 5°N-5°S

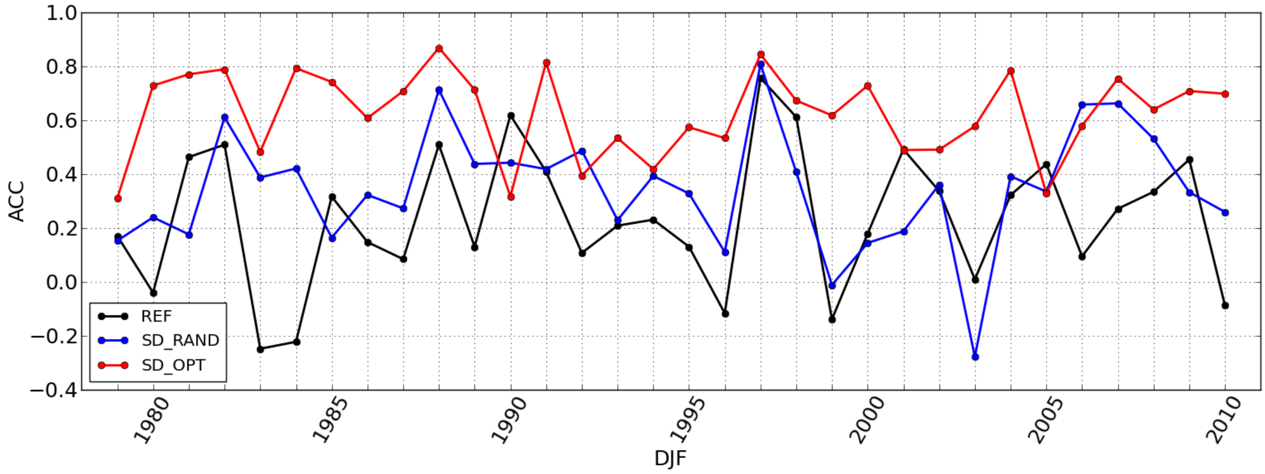

Figure 2. DJF NH Z500 anomaly correlation coefficient for ensembles REF, SD_RAND and SD_OPT.

Table 2. Reliability, resolution, RPS and RPSS values for ERA-Interim climatology, REF, SD_RAND and SD_OPT for NH T2m (land grid points only) and Z500. Bold RPS values indicate scores significantly better than REF at a 95% level using a binomial test for season RPS scores.

Ensemble	Rel	Res	RPS	RPSS
NH T2m (over land)				
Climatology	0.	0.	0.222	-
REF	0.095	0.099	0.218	0.019
SD_RAND	0.094	0.100	0.217	0.026
SD_OPT	0.094	0.112	**0.204**	**0.080**
NH Z500				
Climatology	0.	0.	0.222	-
REF	0.090	0.095	0.217	0.022
SD_RAND	0.088	0.097	0.213	0.042
SD_OPT	0.091	0.120	**0.193**	**0.131**

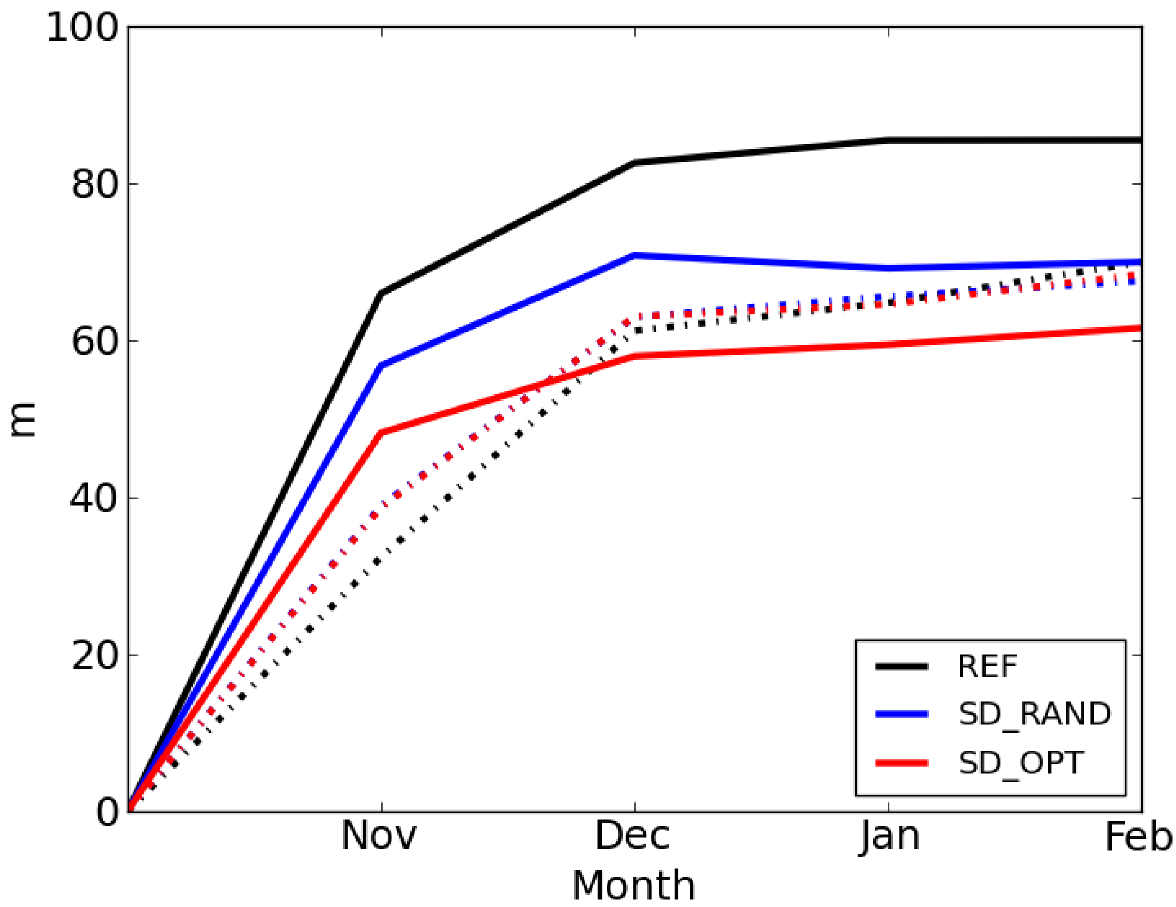

Figure 3. Evolution of monthly root mean square error (full lines) and ensemble spread (dashed lines) for NH Z500 with forecasts REF, SD_RAND and SD_OPT.