Geoscientific
Model Development

Optimization of NWP model closure parameters using total energy norm of forecast error as a target

P. Ollinaho[1,2], H. Järvinen[2], P. Bauer[3], M. Laine[1], P. Bechtold[3], J. Susiluoto[1,2], and H. Haario[4]

[1]Finnish Meteorological Institute, Helsinki, Finland
[2]University of Helsinki, Department of Physics, Helsinki, Finland
[3]European Centre for Medium-Range Weather Forecasts, Reading, UK
[4]Lappeenranta University of Technology, Lappeenranta, Finland

Correspondence to: P. Ollinaho (pirkka.ollinaho@fmi.fi)

Abstract. We explore the use of dry total energy norm in improving numerical weather prediction (NWP) model forecast skill. The Ensemble Prediction and Parameter Estimation System (EPPES) is utilized to estimate ECHAM5 atmospheric GCM (global circulation models) closure parameters related to clouds and precipitation. The target criterion in the optimization is the dry total energy norm of 3-day forecast error with respect to the ECMWF (European Centre for Medium-Range Weather Forecasts) operational analyses. The results are summarized as follows: (i) forecast error growth in terms of energy norm is slower in the optimized than in the default model up to day 10 forecasts (and beyond), (ii) headline forecast skill scores are improved in the training sample as well as in independent samples, (iii) the decrease of the forecast error energy norm at day three is mainly because of smaller kinetic energy error in the tropics, and (iv) this impact is spread into midlatitudes at longer ranges and appears as a smaller forecast error of potential energy. The interpretation of these results is that the parameter optimization has reduced the model error so that the forecasts remain longer in the vicinity of the analyzed state.

1 Introduction

Tuning of closure parameters in atmospheric modeling is a recurring topic. In research, the aim is to improve physical realism of subgrid-scale physical processes and to maintain or improve the general model behavior, such as reproduction of observed variability. In operational applications, such as numerical weather prediction (NWP), the aim is also to increase the predictive skill. Tuning procedures in modeling are predominantly manual and there are no generally applicable or accepted algorithmic tools in everyday use. One reason is that in multiscale and multiphase systems the model response to closure parameter variations is very nonlinear and general nonstationary inverse problem tools can fail. Therefore results may be promising in idealized cases but this does not seem to carry on to more demanding real-world estimation cases. This difficulty is nicely illustrated in Schirber et al. (2013), where the inverse problem realism is gradually increased from a synthetic to fully realistic estimation in case of an atmospheric general circulation model. The parameter-augmented state filter works well in an idealized setup but is less successful in realistic estimation cases.

The aim of this paper is by no means to declare that a final solution has been found to this generic problem. Some success has nevertheless been obtained by applying the so-called Ensemble Prediction and Parameter Estimation System (EPPES; Järvinen et al., 2012; Laine et al., 2012). We have reported earlier (Ollinaho et al., 2013a) that the EPPES algorithm is able to recognize models with superior performance with respect to a given target criterion, even in case of a highly tuned system of full complexity, such as the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF). EPPES is thus clearly a good candidate for a general-purpose tuning algorithm. The remaining key question is the definition of a proper target, the optimization of which can lead to a univocal improvement of the model performance. Targeting

Table 1. ECHAM5 closure parameter subset used in model optimization.

Parameter	Description
CAULOC	A parameter influencing the accretion of cloud droplets by precipitation (rain formation in stratiform clouds).
CMFCTOP	Relative cloud mass flux at the level above nonbuoyancy (in cumulus mass flux scheme).
CPRCON	A coefficient for determining conversion from cloud water to rain (in convective clouds).
ENTRSCV	Entrainment rate for shallow convection.

improvements in all model fields would assure a model-wide improvement, but the construction of correct weights for the all the variables would be impractical. However, a too simple target is not likely to lead to a univocally improved model. This paper presents atmospheric dry total energy norm as a target for model optimization. In recent years, various energy norms have appeared in NWP literature mainly in the context of seeking the fastest growing structures to be used as initial state perturbations in ensemble prediction systems (e.g., Farrel, 1988; Palmer et al., 1994; Errico, 2000), as well as in forecast sensitivity studies (e.g., Gelaro et al., 1998; Orrell et al., 2001; Mitchell et al., 2002). Here we apply the dry total energy norm in the opposite sense of the former: we seek a model which tends to have the slowest possible forecast error growth in terms of dry total energy norm. As the energy norm is computed as an integral over the entire model atmosphere, it is not selective to any particular model variable, level, or geographical region. It is thus a potentially powerful target.

2 Experiment configuration

The ECHAM5.4 atmospheric general circulation model (Roeckner et al., 2003) is used here with a coarse horizontal resolution of T42 and 31 vertical layers, the model top being at 10 hPa. We consider the same four closure parameters (Table 1) that were estimated in Ollinaho et al. (2013b), and studied in Järvinen et al. (2010). These influence parametrized clouds and precipitation, and, even though considered here only from the NWP viewpoint, they are also of great interest when considering the model climatology.

A more complete description of the ensemble prediction system (EPS) emulator is given in detail in Ollinaho et al. (2013b). A concise overview is provided in the following: the operational ensemble of initial states produced by ECMWF EPS (ENS) has been used to generate initial uncertainties. A total of 50 perturbed initial states, as well as the control state, are used for twice-daily (00:00 and 12:00 UTC – universal time coordinated) forecasts over a period of 3 months (January–March 2011). The initial-time parameter

variations, sampled via the EPPES algorithm, represent the model error.

The EPPES algorithm was introduced in Laine et al. (2012), who also demonstrated the algorithm use with a stochastic version of the Lorenz-95 model (Lorenz, 1996; Wilks, 2005). The EPPES algorithm approaches the problem of estimating model parameters $\boldsymbol{\theta}$ by assuming it to be a realization from a background parameter uncertainty distribution that is approximated by a multivariate Gaussian distribution, with a mean vector $\boldsymbol{\mu}$ (of dimension p) and a $p \times p$ covariance matrix $\boldsymbol{\Sigma}$. For each time window i, the optimal parameters, $\boldsymbol{\theta}_i$, are a sample from this distribution as

$$\boldsymbol{\theta}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), i = 1, 2, \ldots \tag{1}$$

The estimation problem is thus shifted to estimating these unknown, but static in time, distribution parameters (or hyperparameters). The mean of the distribution $\boldsymbol{\mu}$ corresponds to parameter values that perform best on average considering all weather types, seasons, etc., and $\boldsymbol{\Sigma}$ indicates how much these values vary between time windows due to inaccurate parametrization schemes and other modeling errors. Thus, $\boldsymbol{\Sigma}$ provides objective information about uncertainties related to the estimated parameters.

The algorithm first draws a sample from an initial distribution, and these parameter values are used in an ensemble of forecasts. The likelihood of each forecast is then evaluated with respect to given criteria, and each parameter vector is weighted by the likelihood. A resample is drawn from the weighted parameter sample, favoring parameter values associated with high likelihood (known as importance sampling). Finally, the hyperparameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are updated with the weighted sample. A new sample is then drawn for the next time window from the updated distribution. The algorithm steps can be written are as follows:

1. Initialize the hyperparameters $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$. The distribution $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ acts as the prior for $\boldsymbol{\theta}$ for the first time window and as the proposal distribution for the first sample.

2. For each time instance i, draw a sample of proposed values for the parameters $\boldsymbol{\theta}_i$ – call them $\tilde{\boldsymbol{\theta}}_i^j$ – from the multivariate Gaussian distribution, $\tilde{\boldsymbol{\theta}}_i^j \sim N(\boldsymbol{\mu}_{i-1}, \boldsymbol{\Sigma}_{i-1})$; $j = 1, \ldots, n_{\text{ens}}$, where n_{ens} is the ensemble size.

3. Using the parameters $\tilde{\boldsymbol{\theta}}_i^j$, generate an ensemble of predictions.

4. Evaluate the fit of each ensemble member with the cost function $J(\tilde{\boldsymbol{\theta}}_i^j)$ and calculate the importance weights $w_i^j \propto \exp(-\frac{1}{2} J(\tilde{\boldsymbol{\theta}}_i^j))$, such that $\sum_{j=1}^{n_{\text{ens}}} w_i^j = 1$.

5. Make a weighted resample of $\tilde{\boldsymbol{\theta}}_i^j$ using the weights $w(\tilde{\boldsymbol{\theta}}_i^j)$ as $\boldsymbol{\theta}_i^j$, $j = 1, \ldots, n_{\text{ens}}$, and use the sample to

update the hyperparameters ($\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$) by the EPPES update formulae (see Laine et al., 2012)

6. For the next time window $i+1$, specify the proposal distribution for parameter $\boldsymbol{\theta}_{i+1}$ as $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ and go back to step 2.

The initial distribution is defined according to expert knowledge ("Prior" in Table 2). Default model parameter values provide practical values for $\boldsymbol{\mu}_0$. The initial time parameter uncertainties $\boldsymbol{\Sigma}_0$ can be set rather freely, though too small or too large uncertainties will slow down the estimation process. If no prior information about parameter correlations is available, a diagonal matrix can be used. The estimation process will reveal potential parameter covariances. Parameter bounds are also set to prevent the selection of unrealistic parameter values (Table 2).

3 Total energy norm

3.1 Target criterion

The dry total energy norm in discretized form can be written as

$$\Delta E = \frac{1}{2} \sum_{p_0}^{p_1} \sum_A \left((\Delta u)^2 + (\Delta v)^2 + \frac{c_p}{T_r} (\Delta T)^2 \right) \mathrm{d}A \mathrm{d}p$$
$$+ \frac{1}{2} R_d T_r p_r \sum_A (\Delta \ln p_{\mathrm{sfc}})^2 \mathrm{d}A. \qquad (2)$$

Here, u and v denote the zonal and meridional wind components, T the temperature, and $\ln p_{\mathrm{sfc}}$ the logarithmic surface pressure. ΔS indicates difference between two atmospheric states; i.e., $\Delta S = S_{\mathrm{an}} - S_{\mathrm{fc}}$, where subscripts denote analysis (an) and model forecast (fc). c_p is the specific heat at constant pressure, R_d the gas constant of (dry) air, T_r a reference temperature (280 K), p_r a reference surface pressure (1000 hPa) and $\mathrm{d}A$ the areal element of the model grid. $\mathrm{d}p$ is the pressure difference between two pressure levels, we use $\mathrm{d}p = 1$ throughout the atmosphere. Thus every model layer has the same weight in the summation. This treatment emphasizes the surface pressure term since the correct $\mathrm{d}p$ values in ECHAM5 with 31 vertical model levels vary between 10 and 50 hPa.

The first two terms in the right-hand side of Eq. (2) (u and v) are identifiable as kinetic energy, and the third (T) and fourth ($\ln p_{\mathrm{sfc}}$) terms as available potential energy (Lorenz, 1955, 1960). Equation (2) can also be extended to include a term related to the latent energy. We have restricted this study to the dry total energy norm. Optimal inclusion of the latent energy term requires defining a vertically changing weighting term (see Barkmeijer et al., 2001).

The ECMWF operational analyses are used in computation of Eq. (2). The target criterion, or cost function, for the EPPES estimation is then the forecast error from the analysis, the norm being the dry total energy norm.

$$J(\boldsymbol{\theta}) = w \Delta E^{72}(\boldsymbol{\theta}), \qquad (3)$$

where ΔE^{72} denotes the energy state difference between the analysis and a 72 h forecast, and w is an ad hoc scaling term (a value of $1/6 \, (\mathrm{J/kg} \, \mathrm{m}^2 \, \mathrm{Pa})^{-1}$ is used here). The scaling term widens or narrows down the probability density function (pdf) of the analysis field errors. It acts to prevent (i) that the ensemble member with the best fit to the analysis would solely affect the distribution update, and (ii) that all ensemble members would appear as likely. The 72 h forecast range is selected because it is beyond the tangent-linear regime of the system and not seriously affected by the spin-up/down of the model hydrology, and not yet affected by the nonlinear forecast error saturation.

3.2 Model sensitivity

We first study (i) how the model performs in terms of energy norm, and (ii) how much impact the initial state and parameter perturbations have on forecasts with respect to the energy norm. Figure 1 illustrates the ensemble spread of the zonal mean energy norm at a 72 h forecast range, averaged over 15 days (1–15 January 2011). We divide the dry total energy norm (dark blue) into surface pressure (light blue), temperature (dark green) and kinetic energy (light green) terms in order to better understand the respective contributions to dry total energy norm variability. The width of the colored area represents ± two standard deviations (SD) from the mean, thus indicating the impact of initial state and parameter perturbations on the system. Moreover, the mean (continuous black lines) indicates how far the forecast is from the analyses in general.

The largest mean forecast error of the dry total energy is in the midlatitudes, especially so in the Northern Hemisphere (30 to 60° N), where all three energy norm terms also reach their individual maximum values. There is also an increased ensemble spread associated with both of the hemispheric maxima as well as in the tropics (shifted slightly towards the summer hemisphere). The impact of initial state and parameter perturbations separately to the spread of dry total energy norm was also tested by running the model with only one perturbation type active at a time. Figure 2 shows the ensemble spread caused by the combined effect of parameter and initial-state perturbations (thick continuous lines), as well as the independent contributions of parameter perturbations (thin continuous lines) and initial-state perturbations (dashed lines). The spread of the dry total energy (total), and the individual spread of surface pressure (surf pres), temperature and kinetic energy (kinetic) terms are shown. The separate contributions to the dry total energy norm are as follows: parameter variations dominate in the tropics, initial state perturbations

Table 2. Parameter values for ECHAM5 (T42L31) in EPPES tests.

Parameter	Prior		Bounds	Posterior	
	mean	SD		mean	SD
CAULOC	2.0	7.0	0–30	10.79	4.29
CMFCTOP	0.3	0.2	0–1.0	0.42	0.12
CPRCON	1.5×10^{-4}	4.0×10^{-3}	0–1.5×10^{-2}	3.63×10^{-3}	1.43×10^{-3}
ENTRSCV	3.0×10^{-4}	1.0×10^{-3}	0–5.0×10^{-3}	2.12×10^{-4}	0.91×10^{-4}

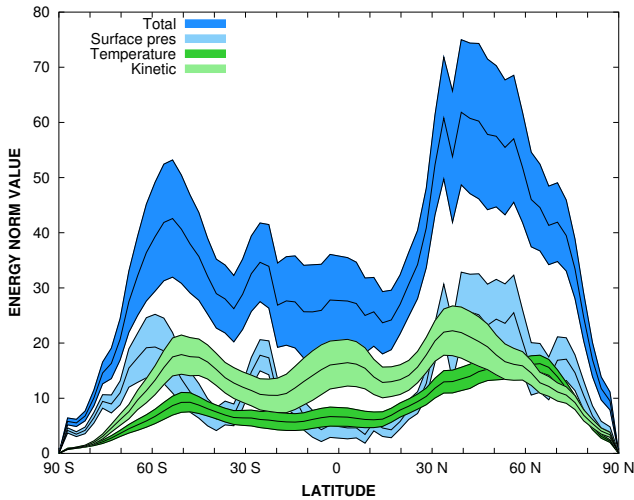**Figure 1.** Mean error and ensemble spread of zonally averaged and areal-weighted energy norm (unit J/kg m^2 Pa) for 15 days (1–15 January 2011) from a +72 h forecast. Dry total energy norm (dark blue), and individual terms: surface pressure (light blue), temperature (dark green) and kinetic energy (light green). Continuous black line indicates the mean model error. Width of the colored area represents ± two standard deviations from the mean.

dominate in the Southern Hemisphere, and both sources are approximately equal in the Northern Hemisphere.

The surface-pressure term has three mean error maxima: two in the Southern Hemisphere (22 and 60° S) and a broader one in the Northern Hemisphere (35–57° N). The peaks at 22° S and 35° N, namely the Andes and the Himalayas regions, are caused by orographical differences between ECHAM5 and the originally higher-resolution analysis data. Ensemble spread is the largest within the peak areas of 60° S and 40–57° N. The southern hemispheric maximum is dominated by initial state perturbations, whereas in the Northern Hemisphere both perturbations have an equal effect.

The temperature term has the least spread. The mean is quite flat with respect to latitude, but at higher latitudes the model deficiencies start to appear, especially in the Northern Hemisphere. The ensemble spread of the temperature term remains relatively small at all latitudes, and is governed by the initial state perturbations in the extratropics and by parameter variations in the tropics.

The mean error in the kinetic energy term has also multiple maxima: one in the midlatitudes in each hemisphere, and one in the tropics. The ensemble spread is large at all latitudes. Parameter perturbations dominate the spread in the tropics and extratropics, while initial state perturbations dominate in the southern midlatitudes. In the northern midlatitudes, initial-state and parameter perturbations generate roughly the same amount of ensemble spread.

4 Results

4.1 Parameter evolution

The evolution of the parameter subset from 1 January to 31 March 2011 (2011JFM) is shown in Fig. 3. The parameter perturbation distribution mean μ (continuous line), width (± two times the standard deviation; thin dashed lines), and default parameter values (thick dashed line) are presented. A vertical column of markers represents a set of 50 parameter values evaluated at the corresponding date, and the marker shading is indicative of the importance weight in the distribution update. Two of the parameters (CAULOC and CPRCON) shift fairly quickly to higher parameter values, followed by saturation. CMFCTOP and ENTRSCV, however, change more conservatively throughout the evaluation period. The posterior distribution mean μ and standard deviation after the final iteration are given in Table 2.

4.2 Validation

4.2.1 Skill scores

To validate the parameter distributions, the model is run applying the parameter posterior mean values. Three time periods are covered: (i) the dependent period of 2011JFM, (ii) an independent period of April 2011 (2011A), and (iii) an independent period of January to March 2010 (2010JFM). We first study how the optimized model compares with respect to the target criterion. Figure 4 represents the energy norm differences between the default and optimized model for the three time periods and up to forecast day 10. The mean difference (continuous line) and the 95 % confidence interval of the mean (gray vertical bars; the bar width is two times the standard deviation of the differences divided by the square
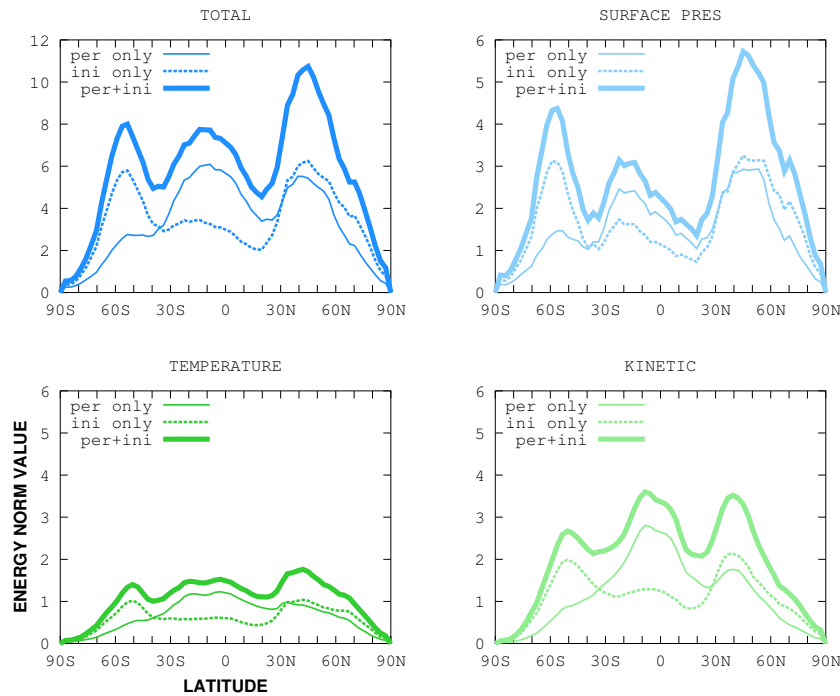
Figure 2. Ensemble spread (two times the standard deviation; unit J/kg m^2 Pa) at forecast day three averaged over 30 ensembles. Spread of dry total energy norm (total), and surface pressure (surface pres), temperature and kinetic energy (kinetic) terms. Experiments with only parameter perturbations active (thin continuous lines), only initial state perturbations active (dashed lines), and both sources of uncertainty active (thick continuous line).

root of number of cases) are shown. The first thing to note is that the energy norm at forecast day three for 2011JFM is improved at the 95% confidence level, implying that the EPPES algorithm is able to find a model that is improved with respect to the target criterion. In fact, there is an improvement at all ranges. The energy norm improvement is statistically significant also for forecast ranges beyond 2 days in the independent sample 2011A, and beyond 5 days in the 3-month sample 2010JFM.

Next, the model is validated against the standard headline score of 500 hPa geopotential height. In addition to the RMSE (root mean square error), the anomaly correlation coefficient (ACC) is also shown. ACC is a verification quantity which is sensitive to the forecast patterns. Notation is the same in Figs. 4 and 5. Positive values for both RMSE and ACC indicate where the optimized model is performing better than the default one. The RMSE scores for all three data sets are improved at the 95 % confidence level for all forecast ranges. Interestingly, the mean RMSE scores of the independent sample of 2011A are improved more than in the dependent sample. ACC scores in the dependent sample are improved for forecast ranges longer than 2 days, and statistically significantly at forecast ranges of 2.5–8 and 9.5–10 days. The ACC scores are also improved from forecast day five onwards for the independent sample of 2011A, although this does not hold at the 95 % confidence level. For

the second independent sample the ACC is mostly neutral with some statistically insignificant improvements for forecast ranges beyond 7 days.

4.2.2 Scorecard

A more general validation of the model changes with the optimized parameters is provided by a scorecard (Fig. 6a–c). It is a concise but comprehensive presentation of a large number of scores for various geographical regions, variables, levels, and forecast ranges. The notation is such that green (red) colors indicate the optimized model scoring better (worse) than the default model. Small and large arrowheads pointing up (down) indicate the result is significant at the 95 or 99 % confidence level, respectively, for the optimized (default) model to score better. White boxes indicate the models performing equally well.

The main features of Fig. 6a–c are as follows. First, RMSE scores of all forecast fields (with exception of temperature at 100 hPa) in the Northern Hemisphere are improved beyond a forecast range of 2 days. In the Southern Hemisphere the same holds at forecast ranges longer than 3.5 days. ACC scores in the Northern Hemisphere closely follow those of the RMSE, whereas, in the Southern Hemisphere, wind fields at the 2.5–4.5 day range and cloud cover at upper levels differ from their respective RMSE improvements. There is a general improvement in RMSE scores for the tropics, with the
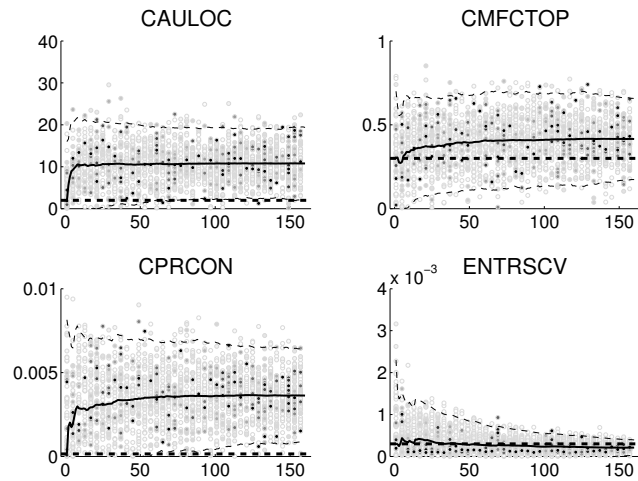
Figure 3. Evolution of parameter subsets from 1 January 2011 to 31 March 2011. The distribution mean μ (continuous line), $\pm$ two times the standard deviations (thin dashed lines), and default parameter value (thick dashed line). A vertical column of markers represents parameter values evaluated at the corresponding date, the marker shading is indicative of the weighting in the distribution update. For clarity only every fourth ensemble is plotted.

Figure 4. Energy norm differences (unit $\mathrm{J/kg\,m^2\,Pa}$) between default and optimized model. Top panel: dependent sample (January–March 2011), middle panel: independent sample of April 2011, bottom panel: independent sample of January–March 2010. Mean difference (continuous line) and 95 % confidence interval of the mean (grey bars).

exception of geopotential height at the forecast range of 3–7 days at 1000 and 850 hPa levels, temperature at the 100 hPa level, and surface temperature. The ACC scores for the tropics are affected similarly to the RMSE scores; the exception being cloud fraction, which is negatively affected at nearly all forecast ranges.

4.2.3 Geographical validation

Next, the geographical distribution of the energy norm differences between the optimized and default models are presented. The kinetic energy mean forecast difference for day three forecasts from 2011JFM is shown in Fig. 7. Positive values indicate where the optimized model is better than the default model. The main improvements are concentrated in the tropics (Southeast Asia, the western coasts of Africa and South America). A weakly positive region is close to the Atlantic storm track. The Atlantic and Indian oceans around 40° S are somewhat degraded.

Figure 8a–c illustrate the zonally averaged mean energy norm difference in the dependent sample (2011JFM) for forecast ranges of 3, 6, and 10 days (Fig. 8a–c, respectively). The total energy norm (dark blue), and surface pressure (light blue), temperature (dark green) and kinetic energy (light green) terms are presented. The mean error (continuous black line), and the 95 % confidence interval of the mean (width of the colored area) are also shown.

At forecast day three (Fig. 8a), most of the improvements in the dry total energy take place in the tropical belt, but there is also a favorable impact on the northern midlatitudes (north of 45° N). A forecast degradation is seen in the

Southern Hemisphere (25–50° S). In the tropics, the surface pressure term displays oscillations arising from orographically induced noise as the analysis data are at higher resolution than the forecasts, and the term stays negative excluding the high latitudes (south of 55° S and north of 45° N). The temperature term displays a broad positive signal for all latitudes. Improvements in the tropics are dominated by the kinetic energy, with positive impacts for all latitudes expect 25–50° S. Figure 9 represents the vertical distribution of the zonally averaged total energy norm (EN) differences between the default and optimized model. Positive values indicate where the optimized model is performing better. The tropical total EN improvements seen in Fig. 8a are located between 850 and 150 hPa layers. The biggest improvements are found in the upper troposphere centered around 200 hPa, and lower in the troposphere around 700 hPa. The largest extratropical improvements occur between 400 and 300 hPa pressure levels. The southern hemispheric degradation is situated near the tropopause above 100 hPa.

At longer forecast ranges, the improvements are spread from the tropics to the midlatitudes and grow larger. By forecast day six (Fig. 8b), the largest values are at midlatitudes and are dominated by the kinetic energy term, and later by the surface pressure term (Fig. 8c). Note the different scale in the panels of Fig. 8a–c.

5 Discussion

The EPPES methodology was able to find a parameter set corresponding to an improved model with respect to the

Figure 5. The 500 hPa geopotential height difference. Left panels: RMSE (default minus optimized model; unit m), right panels: ACC (optimized minus default model). Top panels: dependent sample (January–March 2011), middle panels: independent sample of April 2011, bottom panels: independent sample of January–March 2010. Mean difference (continuous line) and 95 % confidence interval of the mean (gray bars).

Figure 6a. A forecast validation scorecard for 180 forecast cases between 1 January and 31 March 2011 for the Northern Hemisphere. Forecast performance is color-coded as follows: green is good for the optimized model while red is good for the default model. Small (large) arrowheads indicate the 95 % (99 %) level of statistical significance of the score difference. The first column indicates the area, second the variable, third the pressure level, and the fourth and fifth columns RMSE and ACC scores for forecast days 1–10.

Figure 6b. As Fig. 6a, but for the Southern Hemisphere.

Figure 6c. As Fig. 6a, but for the tropics.

target criterion, and thus demonstrates that the algorithm works as intended. This improvement is not confined to the sampling period, as it is also present in the independent sample 2011A, and to some extent also in the 2010JFM sample.

Figure 4 illustrates how the optimized model stays closer to the verifying analyses than the default model. The energy norm is optimized at day three but the improvements are also maintained at longer forecast ranges, and the optimized

model seems to outperform the default model the longer the forecast lead time is. This indicates that the optimization procedure has managed to reduce the model error since the forecasts are launched from the same initial conditions. Figure 8a indicates that the model error reduction primarily affects the evolution of kinetic energy in the tropical region in the forecasts of up to 3 days. This is likely to be because the set of four parameters optimized here mostly impacts
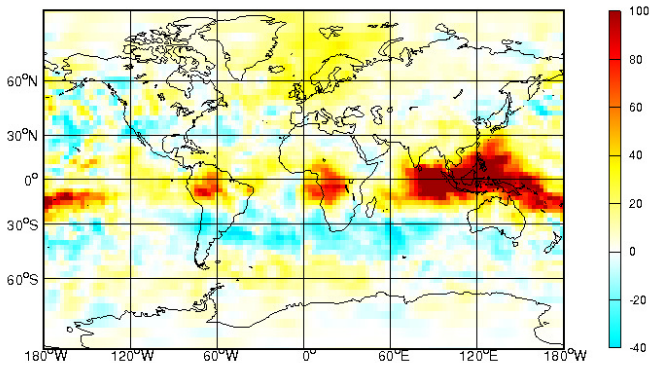
Figure 7. Forecast day three kinetic energy mean difference (unit J/kg m^2 Pa) of the optimized and default model from January to March 2011. Positive values indicate improved day three forecasts after parameter optimization.
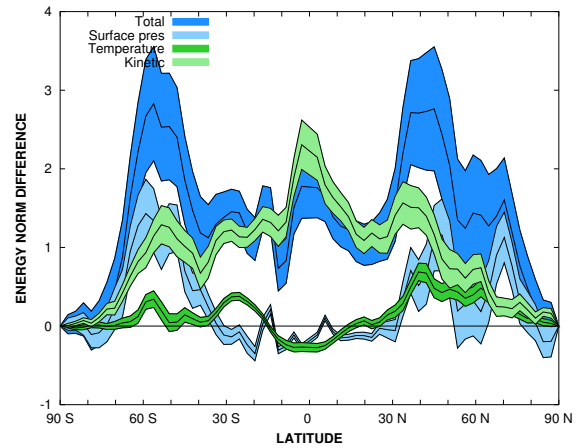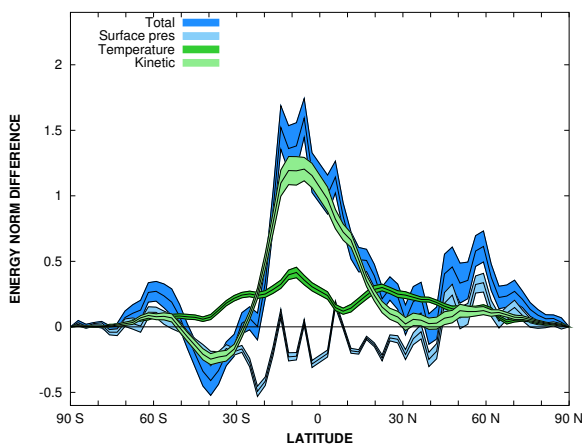

Figure 8b. As Fig. 8a, but for forecast day six.

Figure 8a. Zonally averaged and areal-weighted energy norm difference (unit J/kg m^2 Pa) between the default and optimized models from January to March 2011 for forecast day three. Dry total energy norm (dark blue), and surface pressure (light blue), temperature (dark green) and kinetic energy (light green) terms individually. Continuous black line indicates the mean error, and width of the colored area represents the 95 % confidence interval of the mean.
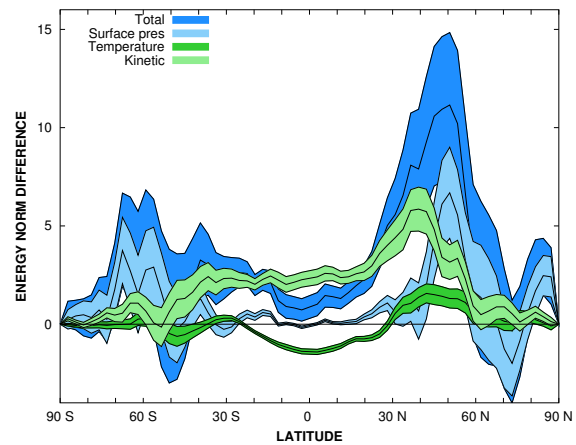

Figure 8c. As Fig. 8a, but for forecast day 10.

convective circulation in the tropics. After the 3-day optimization period, the tropical kinetic energy improvements spread by nonlinear model dynamics into the midlatitudes (Fig. 8b), and begin also to appear as improvements in the distribution of potential energy via the surface pressure term. Note, that there is a tropical maximum in the kinetic energy distribution at day six (Fig. 8b). The interpretation of this maximum is that the reduced model error continues to operate in the tropics and feeds more realistic kinetic energy evolution via better tropical circulation throughout the 10-day forecast range.

Ollinaho et al. (2013b) estimated the same ECHAM5 model parameters as here with the EPPES methodology but using a mean-squared forecast error of the 500 hPa geopotential height at forecast days 3 and 10 as a target criterion. Those experiments showed that the EPPES methodology is capable of optimizing a given target in an atmospheric GCM of full complexity. The posterior mean parameter values of Ollinaho et al. (2013b) are within two standard deviations of the values found here. In particular, the posterior mean of the parameter CAULOC assumes a very similar value using either of the two targets, while the parameter CPRCON results in a value almost 1.8 times higher using the 500 hPa height rather than energy norm as a target. However, the 500 hPa skill optimized model developed a significant bias above the 500 hPa level, visible for instance as inferior 100 hPa height skill scores compared with the default model. A scorecard presenting tropical RMSE scores of the two optimized
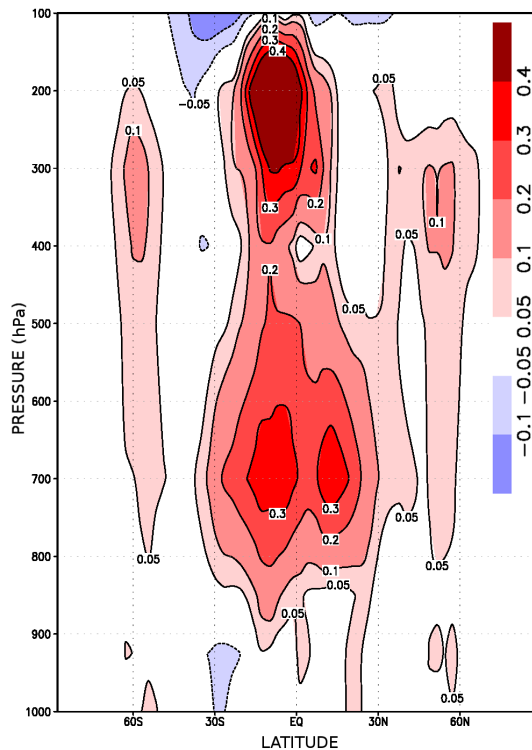
Figure 9. Pressure–latitude cross section of forecast day three zonal mean energy norm differences (unit $J/kg\,m^2\,Pa$) between default and optimized models from January to March 2011. Positive values indicate where the optimized model is performing better.

models is shown in Fig. 10. A comparison of the models reveals that the version optimized using the energy norm is superior especially with respect to the winds. One reason for this result is the ambiguity of 500 hPa skill as a target: the upper troposphere and lower stratosphere circulation is not properly constrained and there are many model realizations (i.e., the same model structure at the 500 hPa level but different closure parameter values) that fulfill the target.

Analysis of the model moisture fields implies that applying the moist energy norm (see e.g., Barkmeijer et al., 2001, for the formula) as the target criterion would further emphasize the tropics in the estimation process. The contribution of the moisture term to the total EN would be on the same order as the temperature term. We speculate that including the term into the cost function would have a small effect on the final parameter distributions. Although, without constructing a weighting function for the moisture part we cannot predict what the magnitude of the impact would actually be.

Since the target criterion can be chosen quite freely, changes in specific regions can also be targeted for optimization with the EPPES algorithm. For instance, in the current experimentations with the IFS parameter variations have a rather small impact on calculated EN scores outside the tropics. Thus, a cost function constructed from the tropical EN

scores might only be more efficient for optimization purposes.

The choice of target criterion has to be considered carefully prior to the parameter estimation. Tuning of the physical processes could be done by e.g., focusing on the direct effects of the parametrizations only; i.e., cloudiness and precipitation in this study. However, this can lead to models where a (seemingly) good representation is reached at the expense of other model fields. Hence, a target criterion focusing on the model forecast skill in more general terms seems more practical when the goal of the tuning is a univocal model improvement. The total energy norm offers a potential target for parameter optimization since it takes into account the model changes in all model fields, and focuses on key features of the model.

6 Conclusions

This article explores the use of atmospheric dry total energy norm in improving NWP model forecast skill. EPPES (Järvinen et al., 2012; Laine et al., 2012) is utilized to estimate four ECHAM5 model parametrization closure parameters related to clouds and precipitation. The ensemble runs are generated using the ECHAM5 model to evolve the perturbed initial states generated by the ECMWF for their ensemble prediction system. Here, model error is represented (and thus ensuring sufficient spread of the ensembles) by perturbing the ECHAM5 closure parameters which are being estimated. The twice-daily 50 member ensembles are generated over a period of 3 months and each ensemble member is used in the sequential parameter distribution update according to their respective weights obtained by calculating the dry total energy norm of the 3-day forecast error against the ECMWF analyses.

We first study the impact of initial state and parameter perturbations on the ensemble spread in terms of energy norm of the 3-day forecast error in a sample of 30 forecasts using the default model. On average, the forecast departures from the analyses are largest at the Northern (winter) Hemisphere's midlatitudes. In the tropics, the ensemble spread is mostly due to parameter variations, whereas at higher latitudes initial state perturbations either dominate or are equally important as parameter perturbations.

The optimization is performed in a 3-month period (January–March 2011), and the optimized model is validated with respect to the optimization criterion, typical head-line scores, and a comprehensive scorecard. First, the optimized model is an improvement with respect to the target criterion. Moreover, the improvement is propagated to 3–10-day forecasts. Second, head-line scores are improved in dependent and independent samples. Third, the scorecard shows improvements on a broad range of individual scores, such as clearly improved tropical winds. The improvements of the energy norm are found to stem from better representation

Figure 10. Comparison of forecast validation score cards for the tropics. Left column: model optimized with dry total energy norm as target criterion, right column: model optimized with geopotential height mean squared error (MSE) at the 500 hPa level as target criterion. In total, 180 forecast cases between 1 January and 31 March 2011. Forecast performance is color coded as follows: green is good for the optimized model while red is good for the default model. Small (large) arrowhead indicates the 95 % (99 %) level of statistical significance of the score difference. The first column indicates the area; second, the variable; third, pressure level; and the fourth and fifth columns the RMSE scores for forecast days 1–10.

of tropical kinetic energy in short (up to 3-day) forecasts. This improvement spreads in 3–6-day forecasts to midlatitudes and starts to appear as a better representation of the potential energy distribution.

We conclude that the EPPES algorithm is a viable option in optimization of atmospheric GCMs of full complexity. The optimization target of the algorithm can be selected rather freely. The dry total energy norm seems promising in this respect. Please note that the EPPES codes used here and some examples are available online at http://helios.fmi.fi/~lainema/eppes/.

References

Barkmeijer, J., Buizza, R., Palmer, T. N., Puri, K., and Mahfouf, J.-F.: Tropical singular vectors computed with linearized diabatic physics, Q. J. R. Meteorol. Soc., 127, 685–708, doi:10.1002/qj.49712757221, 2001.

Errico, R. M.: Interpretations of the total energy and rotational energy norms applied to determination of singular vectors, Q. J. R. Meteorol. Soc., 126, 1581–1599, doi:10.1002/qj.49712656602, 2000.

Farrel, B.: Optimal excitation of neutral Rossby waves, J. Atmos. Sci., 45, 163–172, doi:10.1175/1520-0469(1988)045<0163:OEONRW>2.0.CO;2, 1988.

Gelaro, R., Buizza, R., Palmer, T. N., and Klinker, E.: Sensitivity Analysis of Forecast Errors and the Construction of Optimal Perturbations Using Singular Vectors, J. Atmos. Sci., 55, 1012–1037, doi:10.1175/1520-0469(1998)055<1012:SAOFEA>2.0.CO;2, 1998.

Järvinen, H., Räisanen, P., Laine, M., Tamminen, J., Ilin, A., Oja, E., Solonen, A., and Haario, H.: Estimation of ECHAM5 climate model closure parameters with adaptive MCMC, Atmos. Chem. Phys., 10, 9993–10002, doi:10.5194/acp-10-9993-2010, 2010.

Järvinen, H., Laine, M., Solonen, A., and Haario, H.: Ensemble prediction and parameter estimation system: the concept, Q. J. R. Meteorol. Soc., 138, 281–288, doi:10.1002/qj.923, 2012.

Laine, M., Solonen, A., Haario, H., and Järvinen, H.: Ensemble prediction and parameter estimation system: the method, Q. J. R. Meteorol. Soc., 138, 289–297, doi:10.1002/qj.922, 2012.

Lorenz, E. N.: Available potential energy and the maintenance of the general circulation, Tellus, 7, 157–167, doi:10.1111/j.2153-3490.1955.tb01148.x, 1955.

Lorenz, E. N.: Energy and numerical weather prediction, Tellus, 12, 364–373, doi:10.1111/j.2153-3490.1960.tb01323.x, 1960.

Lorenz, E. N.: Predictability: A problem partly solved, in Proceedings of seminar on predictability: Volume 1. ECMWF, Reading, UK, 1, 1–18, 1996.

Mitchell, H. L., Houtekamer, P., and Pellerin, G.: Ensemble Size, Balance, and Model-Error Representation in an Ensemble Kalman Filter, Mon. Weather Rev., 130, 2791–2808, doi:10.1175/1520-0493(2002)130<2791:ESBAME>2.0.CO;2, 2002.

Ollinaho, P., Bechtold, P., Leutbecher, M., Laine, M., Solonen, A., Haario, H., and Järvinen, H.: Parameter variations in prediction skill optimization at ECMWF, Nonlin. Processes Geophys., 20, 1001–1010, doi:10.5194/npg-20-1001-2013, 2013a.

Ollinaho, P., Laine, M., Solonen, A., Haario, H., and Järvinen, H.: NWP model forecast skill optimization via closure parameter variations, Q. J. R. Meteorol. Soc., 139, 1520–1532, doi:10.1002/qj.2044, 2013b.

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.

Palmer, T. N., Buizza, R., Molteni, F., Chen, Y.-Q., and Corti, S.: Singular vectors and the predictability of weather and climate, Phil. Trans. R. Soc. Lond. A, 348, 459–475, doi:10.1098/rsta.1994.0105, 1994.

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornblueh, L., Manzini, E., Rhodin, A., Schlese, U., Schulzweida, U., and Tompkins, A.: The atmospheric general circulation model ECHAM5, Part I: Model description, Tech. Rep. 349, Max-Planck-Institut für Meteorologie, 2003.

Schirber, S., Klocke, D., Pincus, R., Quaas, J., and Anderson, J.: Parameter estimation using data assimilation in an atmospheric general circulation model: From a perfect towards the real world, J. Adv. Model Earth Syst., 5, 58–70, doi:10.1029/2012MS000167, 2013.

Wilks, D. S.: Effects of stochastic parametrizations in the Lorenz '96 system, Q. J. R. Meteorol. Soc., 131, 389–407, doi:10.1256/qj.04.03, 2005.