Letters to referees #1 and #2.

Letter to Referee #1

We thank the Referee #1 for the comments, finding anwers related to the moist energy norm was insigthful. We answer the questions raised by Referee #1 and indicate what action was taken in order to correct the problem in the following:

- *The title does not make much sense to most readers. Something about EPPES, or related, should be there.*
  Title changed to "Optimization of NWP model closure parameters using total energy norm of forecast error as a target". **Title changed.**

- *Should have "dry" in front of total energy norm throughout the paper.*
  Total energy norm is now referred to as dry total energy norm. **Text amended.**

- *Discuss or even speculate how much the moisture part can influence the results and conclusions.*
  The moisture part of the total EN (calculated as in e.g. Barkmeijer, 2001) seems to be in the same order of magnitude as the temperature term when using w=1. When using the uniform weight the tropical lower troposhere is dominating the term.

  Even using the dry total energy as the target for optimization improves the humidity profile when compared to the default model. We speculate that including the moisture term would have a slight effect on the final parameter distributions, with a (small) added influence on the model performance with respect to the humidity fields. But without constructing a weighting function for the moisture part we cannot accurately predict what the magnitude of the impact would actually be. **Text added, Chapter 5, p. 13 lines 3-8.**

- *Also when kinetic energy is used, please explain why not use the dry total energy.*
  The division of the dry total energy norm into kinetic energy, and temperature and surface pressure terms is done to better understand the model response to the change of the closure parameters. We want to study the total norm itself, but also learn about the individual contributions. This has been emphasized now. **Text amended, Chapter 3.2, p. 7 lines 9-12.**

- *As this is based on the previous work using other norms, it would be nice to show some comparison results, which can demonstrate the superiority of the energy norm.*
  We have added a scorecard comparison of tropical RMSEs of the energy norm and geopotential height target criterion experiments. **Text changed, Chapter 5, p. 12 lines 26-28. Added Fig. 9.**

- *I guess the energy norm can also computed over a limited area and a selected vertical range. I know many people try to find a universal number for a model parameter over the whole globe, but I guess we may have to*

use different numbers for different areas. Some discussion may be useful, especially in connection with the regional degradations.
Yes, the energy norm can be also computed non-globally. Also, the EPPES estimation can be applied to target improvements in limited area(s). Thus, if the aim is to find a model that performs at the best possible forecast skill in a certain area(s), EPPES could be used to optimize a closure parameter set with this in mind. However, we feel that extending the discussion into the technicalities of modeling the physical processes themselves is beyond the scope of this paper. **Text added to emphasize the limited area optimization possibilities, Chapter 5, p. 13 lines 9-13.**

- *All readers need to read previous EPPES papers in order to read this paper. Is EPPES really well-known?*
  We have extended the description of the EPPES algorithm. **Text added, Chapter 2, p. 4 lines 10-12, 24-27, p. 5 lines 1-16.**

- *Eq (2). Should there be a delta_p or delta_sigma in the vertical summation to give proper weights to different model layers? At least some comments should be offered on why they can use the same weight for different layers for the total energy computation.*
  Yes, there should be. The term was omitted since we use dp=1 throughout the atmosphere. Since this was an initial experiment we wanted to also have a contribution to the cost function from the surface pressure term. Including proper dp weights would make the kinetic energy and temperature terms about 30 times larger. Also, the weights between the levels are quite uniform, thus we feel that having dp=1 does not produce substantial errors in the atmospheric weighting. Although we do realize that the upper atmosphere is a bit overweighted in our treatment. The small surface pressure term could also just be omitted (like in e.g. Orrell, 2001). **Equation 2 amended and text added, Chapter 3, p. 6 lines 8-10.**

- *"The impact of initial state and parameter perturbations separately... (not shown)." Why not? It is quite interesting.*
  We have added a figure showing the individual spread contributions. **Text added, Chapter 3.2, p. 7 lines 19-26. Added Fig. 2.**

- *Fig.1. What is the unit for energy norm?*
  $[v]^2 * [A] * [p] = ... = J/kg * m^2 * Pa$. **Unit added, Fig. 1.**

- *Fig.2. Where is the shading scale? May need to use colors. Units?*
  The importance weights are not the primary focus of the figure, but rather included as a curiosity factor. Thus, adding a shading legend would, in our mind, take the focus to the wrong place. The units are in fractions, i.e. black dot 51/51 (the parameter value dictated solely the distribution update), white dot 0/51 (the parameter value did not effect the distribution update). **No changes.**

- *Fig.3. Units?*
  As in Fig. 1. **Unit added, Fig. 3.**

- *Fig.4. Units? Are these large or small differences?*
  Units for RMSE is m, for ACC fraction (0,01 equals to 1%). The ECHAM5

average RMSE score for three day forecast of z500 over the sampling period is about 27 m. The change is thus about 2%. **RMSE unit added to caption of Fig. 4.**

- *Fig.5. Too small.*
  Agreed. **Figure enlarged.**

- *Fig.6. Why not dry total energy?*
  We felt it would be of more interest to show a quantity which is easier to relate to real phenomena (accuracy improvements in wind speeds). **No changes.**

- *Fig.7 Too small.*
  Agreed. **Figure enlarged.**

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.

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.

Letter to Referee #2

We also thank the Referee #2 for the comments, the manuscript was clarified in many places due to your critisism. The questions poised by Referee #2 are answered and relevant corrections presented in the following:

- *The estimation procedure itself is not rigorously explained within this paper. It is advisable to read (Ollinaho et al.,QJRMS 2013) first in order to understand the details. The relation to that paper should be made more clear from the beginning.*
  The EPPES methodology has now been explained more thoroughly. **Text added, Chapter 2, p. 4 lines 10-12, 24-27, p. 5 lines 1-16.**

- *It could be mentioned that the total energy norm is not only used for seeking the fastest growing modes (as cited) but also for forecast sensitivity studies based on adjoints or forecast ensembles. (this kind of application is related even closer to this approach).*
  Forecast sensitivity studies are now cited. **Text added, Chapter 1, p. 3 lines 18-20.**

- *The paper presents the temporal evolution and final value of the standard deviation of the estimated parameters (Table 2, Figure 2). However the meaning of this uncertainty measure should be explained in more detail within this paper. Is it some objective measure of the estimated parameter or can it be interpreted only in the context of this estimation procedure (to draw a reasonable a priory ensemble).*
  The posterior distribution width is indeed an objective measure of the parameter uncertainty. The prior uncertainty should not play a very important role in this, although the distribution width will converge slower if the prior distribution width is too wide or too narrow. **Text added, Chapter 2, p. 4 lines 22-23, p. 5 lines 18-21.**

- *The value of the parameter 'w' in equation (3) should be given, as well as a more concise reasoning for its choice. Has the value of w any influence on the final estimate of parameter uncertainties or on the convergence of the scheme?*
  The value of w does influence the estimation procedure; it controls how many of the ensemble members influence the hyperparameter update, i.e. w acts to scale the pdf of the analysis field errors. This is done to prevent i) a too narrow error pdf, where only the ensemble member closest to the analysis would influence the distribution update, and ii) a too wide error pdf, giving all members equal likelihood. **Text amended and added, Chapter 3, p. 6 lines 21-22, p. 7 lines 1-2.**

- *Using the energy norm as a target has been shown to be superior to using geopotential height. This is contributed to the fact that deviations of model parameters from the analysis are constrained at all levels and not only at 500 hPa. It would be nice if this mechanism could be explored in more detail, for instance by showing zonal averages (pressure - latitude slices) of total energy contributions.*
  We feel the article is already quite heavy with figures. Thus only a description of the vertical structure of the improvements was added. The

figure is nonetheless attached below. **Text added, Chapter 4.2.3, p. 11 lines 15-19.**

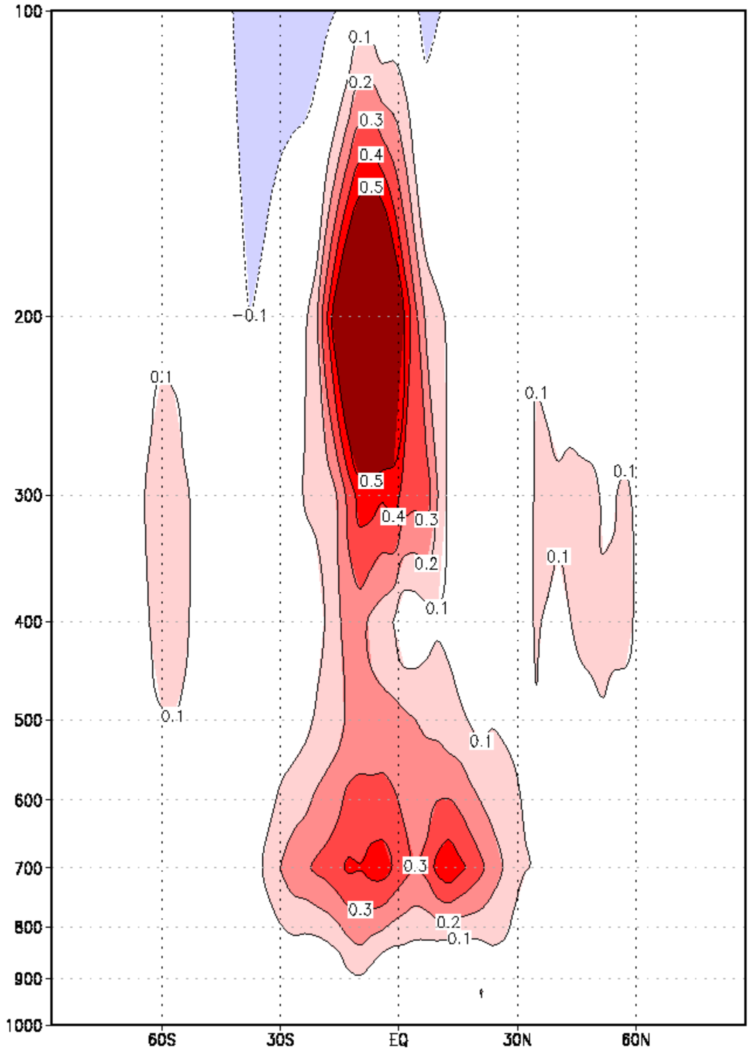

Figure 1: Pressure-latitude cross-section of the energy norm. Differences between the default and optimized models. Positive (negative) values indicate where the optimized (default) model is better.

- *I do not understand the discussion of 'ambiguity' in terms of bias in the Discussion. Any bias, even if it changes sign within the model domain, will give a contribution to the squared analysis minus forecast differences used in the energy norm and thus will be penalised.*
  Text removed, Chapter 5, p. 13.

- *Also the term 'ambiguity of 500 hPa skill as a target' does not seem appropriate. The problem is not that 'many model realisations fulfill the target',*

but that these model realisations lead to inferior scores (other than those enforced to be superior by choice of the target).
Our reasoning behind the ambiguity in this is as follows: Geopotential height is a summary quantity and is thus insensitive to the vertical profiles of the quantities which define it (temperature and humidity to great extend). Therefore a geopotential height profile negatively biased close to the ground and positively biased higher up could still lead to a good 500 hPa geopotential height (z500) RMSE score. Two different temperature and humidity profiles could therefore lead to same z500 RMSE scores. Moreover, "wrong" atmospheric states can lead to similar z500 RMSE scores as atmospheric states close to reality. To avoid confusion, we have clarified the text to emphasize that the same structure would only be observable at 500 hPa level. **Text clarified, Chapter 5, p. 13 lines 2-3.**

- *Only if all scores regarded to be relevant were included in the cost function (with appropriate weights) it could be assured that all scores would be improved (on average). This is probably no practical approach as not all desired properties may be addressed within the EPPES approach.*
  We agree that calculating the cost function from all relevant fields would be impractical. The energy norm implementation is experimented here in order to find a relatively simple cost function, which nonetheless would lead lead to a univocally improved model. **Text added, Chapter 1, p. 3 lines 12-14.**

- *The citation (Ollinaho et al.,2012) (QJRMS) should be (Ollinaho et al.,2013).*
  Text amended.