Geosci. Model Dev. Discuss., 8, C1822–C1832, 2015
www.geosci-model-dev-discuss.net/8/C1822/2015/

Geoscientific Model Development Discussions

Interactive comment on "A global empirical system for probabilistic seasonal climate prediction" *by* J. M. Eden et al.

J. M. Eden et al.

jonathan.eden@knmi.nl

Received and published: 30 August 2015

Response to the comments of Anonymous Referee #1

Main comments: The title describes exactly the content of the paper: describing a statistical model and evaluating the seasonal probabilistic scores over the world. When I read this title, my first mind was "Yet another comparison of seasonal hindcast scores between numerical models and statistical methods." For reasons I will not develop here, I consider this type of contest as fair as comparing a car race with a horse race. But the content and the philosophy of the manuscript are completely different. The authors reduce strongly the score overestimation by selecting individually predictors amongst a few well known indices (not trying multiple combinations). They also use a progressive learning approach, closer to what would be a true forecast. They present

their product as a complement of numerical operational forecasts, not as a challenger. This convinced me to propose this manuscript as suitable for publication, with a few minor corrections:

Response: We thank the Reviewer for their comments the effort made to understand and appreciate the content of our paper. We are particularly encouraged that the emphasis placed on the avoidance of overfitting and using as few predictors as possible has been acknowledged. A full response to the Reviewer's comments is given below.

Additional comments

p2 lines 2 and 3: relies ... reliable

Response: Sentence changed to "...is dependent on the availability of reliable forecasts".

p3 line 10: of course, the model inadequacy wrt the true world induces systematic errors, but the main problem is elsewhere; if a model had just a cold bias, but would successfully predict the sequence of cold and warm seasons (which is measured by time correlation), one would be satisfied of it; conversely if a model has been carefully tuned and has a bias close to zero, but a very weak time correlation wrt observed seasons, one is not satisfied at all. So the critical point is the fact that, because a model has the wrong equations (assuming that the real world follows a small set of equations), its predictability is low. It is possible that, in addition, this model has biases, but the link bias-predictability is not very tight in practice.

Response: We have revised this sentence to avoid the ambiguity highlighted by the Reviewer. We now refer to "...errors and biases..." rather than "...systematic errors..." in order to make clear that we are describing all model errors. However, we do not feel it is appropriate here to begin a discussion on the sources of model errors. The Reviewer's point about the large portion of model skill being in its ability to simulate the correct sequence of weather events but this is largely driven by the representation of

the initial conditions and the degree of data assimilation performed during the simulation. For clarity, and in order to make our sentence applicable to all climate models, we have chosen the following revision:

"However, the development of dynamical systems is a continuous challenge; climate models are inherently complex and computationally demanding and often contain considerable errors and biases that limit model skill in particular regions and seasons."

p4 line 20: the non transferrability of empirical relations is a major criticism in the case of big climate change (e.g. 2100 RCP8.5). In the case of seasonal hindcasts spanning over the last 30 years, the stationarity hypothesis is acceptable. The major criticism is that these empirical relations could be partly based on coincidences of big events in the past which might not repeat in the future (I mean the coincidences, not the big events). There is thus a dilemma: longer time series, better robustness, lesser stationarity.

Response: In this paragraph (Section 1, paragraph 4) we seek to outline the importance of considering external forcings, not only in climate change simulations and decadal prediction, but also in seasonal forecasting. We assert that, in general, the inclusion of a predictor describing greenhouse gas forcing gives the system greater transferability to a perturbed climate.

p 10 line 16: why random sampling ? The reference forecast for calculating a skill score should be the forecast which minimizes this score in the absence of information: -average of the available past observations for RMSESS -distribution of the available past observations for CRPSSS

Response: We have chosen to sample randomly from the climatology in order to maintain an ensemble vs ensemble approach to the generation of skill scores.

* * * * *

Response to the comments of Anonymous Referee #2

Main comments: The manuscript describes and assesses the skill of a global empirical

system for probabilistic seasonal climate prediction. The manuscript is well organized and provides an original and novel contribution for the field of seasonal prediction, being a valuable benchmark for future assessment of dynamical seasonal prediction systems. The development of empirical systems is an important and complementary contribution to dynamical prediction systems. However, I feel a number of improvements are required in order to make the manuscript ready for publication. Please see below a list of major remarks and additional remarks that I recommend to be addressed prior to publication of this manuscript.

Response: We thank the Reviewer for their comments and in particular for identifying where our manuscript is unclear in describing the methodology used. Our revision takes into account all comments and includes many changes designed to offer greater transparency and clarity to the reader. In particular, we direct the Reviewer to Section 2 in the revised manuscript, which has been restructured considerably.

Major comments: 1) Lack of methodological information to allow repeatability: The presented methodology in the manuscript is mainly descriptive. To allow repeatability of the described empirical model it is necessary to include in the manuscript the equations used to define the model, including a description of model parameters, predictor and predictand variables, and explain how model parameters were estimated. Currently the methodological description is limited to indicate that the developed global empirical system is based on multiple linear regression. A substantially improved description of the developed empirical model with the inclusion of the required equations for the production of probabilistic prediction is needed.

Response: The content of Section 2 is split into sub-sections and we now include a set of equations detailing the development of our empirical model. We begin Section 2 with equation (1) to describe the assumption that the predictand is a function of both external forcing and internal variability components. We seek to make it clear that the external forcing (represented by global CO2-equivalent) is the primary predictor in our system; the inclusion of other predictors vary according to location and season. This

is made clear in equation (2). The fitting of and application of the regression model following the selection of predictors is shown in equations (5) and (6). The generation of a forecast ensemble for probabilistic prediction is shown in equation (7).

2) More precise methodological description needed: In the current model description it unclear which seasons are included in the lagged analysis. Additionally, the procedure of removing the impact of CO2 equivalent signal from modeled time series needs to be explained in details because this procedure is currently unclear. Including the equation used to perform this procedure will help this clarification.

Response: As made clear in Section 2, paragraph 2, predictor information is taken from "the previous three-month season at a lead time of one month (e.g. the forecast for the season March-April-May is estimated using predictors from November-December-January)." Section 2.2 in the revision deals specifically predictor selection and model fitting. We have chosen to define the predictor selection scheme as a two-step process, with the first step built largely on existing knowledge of physical processes. The second step is fully quantitative; we remove the linear trend associated with CO2EQV from the predictand and all other predictors passed from step one and identify the predictors that exhibit a significant correlation with the predictand following the detrending. The procedure for removing the linear CO2EQV trend is made clear in equations (3) and (4).

3) Improved figures are required: All multiple panel figures are currently excessively small in size. For this reason it is not possible to clearly see the results, particularly for the described statistical significance. All multiple panel figures need to be improved (i.e. enlarge all individual panels) to allow the reader to clearly appreciate the presented evidences.

Response: Unfortunately, the volume of plots in this manuscript necessitates the use of multi-panel figures and thus smaller plots than would be ideal. To compensate, all Figures have been changed to .pdf format in the revised manuscript. This allows the

online user to zoom on each plot to a much greater extent and view the necessary detail.

4) Text requires clarity improvement: The text needs to be carefully revised in order to improve clarity. Please see below a number of additional remarks indicating where clarification is required.

Response: The comments of the Reviewer specified below have been taken into account. Our revision includes a much clearer outline of the prediction system, including predictor selection and model fitting. Further details are given in the responses to the Reviewer's comments below.

Additional comments:

Abstract: The acronyms NGOs and ENSO are not defined. All acronyms need to be defined when first used. Please revise the entire manuscript to make sure all acronyms are defined when first used.

Response: Changes made in revision.

Abstract, line 17: using correlation and skill scores. Please be more precise. Correlation of what with what? Which skill scores?

Response: Sentence changed to: ". . .validated against observations using deterministic (correlation of seasonal means) and probabilistic (continuous rank probability skill score) metrics."

Page 3944, line 4: by limiting the effects of model biases. What do you mean here? Do you mean empirical forecasts produced with empirical models do not have biases by construction? Please clarify.

Response: Sentence changed to: ". . .by limiting the effects of dynamical model biases."

Page, 3948, line 18: during the predictor period. What is the predictor period? Please be more precise.

Response: As this section is a description of the predictors only, we agree that this sentence is slightly ambiguous. Sentence therefore changed to: “Finally, as a proxy for soil moisture, which has been shown to impact on local temperature (e.g. van den Hurk et al., 2012), we also consider accumulated rainfall (CPREC) as a potential predictor.”

Page 3949, line 12: using data since 1901. A comment on data availability in the early 1900 is needed here.

Response: The following text has been added to Section 2.2, paragraph 4.

“It is also important to note that, insetting the earliest hindcast to 1961, we seek to limit the impact of poor quality available predictand and predictor data in the early 20th Century. Additionally, to ensure robustness, the multiple linear regression model requires complete predictand-predictor time series of at least thirty years in the fitting period for a forecast to be produced.”

Page 3949, lines 18-20: Please provide equation to explain precisely what was the procedure implemented here to make the described removal.

Response: The original text was slightly confusing to the reader. Please see rewritten section 2.2 in which we show equations to make clear how the linear trend of CO2EQV is removed from the predictand and predictors.

Page 3949, lines 21-26: The described procedure is unclear. Is each predictor tested separately/independently? Please provide equations to show more precisely what has been done.

Response: Please see rewritten section 2. We make it clear that the regression model outlined in equation (2) is implemented independently at each grid point and for each season. In Section 2.2, we offer a clearer explanation of the selection procedure.

“This is achieved by correlating the detrended predictand with each detrended predictor and identifying those predictors that exhibit significant (at the 90% level) correlation.”

Page 3949, lines 25-26: Predictor inclusion is determined independently for each hindcast. What does this precisely mean? Please rephrase and better explain.

Response: We offer better clarification on what is meant here with new text in Section 2.2.

“The predictor selection procedure, in addition to being location-specific, is also implemented independently for each hindcast. In other words, for a given grid point, a given predictor would only be included in the regression model for hindcasts with fitting periods during which it demonstrates predictive potential, allowing for the maximum value to be taken from predictor information in the fairest way.”

Page 3950, line 20: We parameterize this trend. . . The described procedure is unclear. Please provide equations to show more precisely what has been done.

Response: For clarity, the first sentence in Section 3.1, paragraph has been changed:

“The surface air temperature (SAT) shows a clear trend almost everywhere, which is assumed to be proportional to the forcing of greenhouse gases, described by CO2EQV.”

Page 3951, line 1: previous year CO2EQV. Why previous year if you are considering seasonal averages? ShouldnâĂšt it be previous season?

Response: Yes, this is an error. Change made in revision.

Page 3951, lines 4-5: when natural variability is small compared to the forced signal. Please further expand and explain precisely what you mean by this sentence.

Response: We offer clarification on this point in the revision (Section 4.1, paragraph 1).

“Correlation between SAT and CO2EQV is in general strongly positive across the majority of the globe, and particularly so when the response of SAT to the internal variability of the climate system is known to be small compared to the response to the signal associated with anthropogenic forcing, for example in the northern hemisphere during

spring (MAM) and summer (JJA) and throughout the tropics at all times of year.:

Page 3951, line 24: PDO, IOD and AMO indices. At this point it is unclear how predictors are selected. Please further explain and provide precise information on the selection procedure.

Response: Please see revised Section 2, and particularly Section 2.2 in which a clearer explanation of the predictor selection procedure is given.

Page 3952, lines 21-23: The correlation is also strong. . . Unfortunately it is not possible to see these described features. Figure panels are too small. Please enlarge figure panels.

Response: As mentioned above, all Figures will be changed to .pdf format in the revised manuscript.

Page 3952, line 25: Lagged correlation between PREC and the predictors is shown in Fig. 2. What type of lag are you considering? Previous season predictor with next season PREC? Please be more precise.

Response: It is confusing to the reader to refer to the correlations as "lagged" when the forecast lead time time and the time difference between the predictand and predictors sets is made clear in section 2.

"In this case, this is defined as previous three-month season at a lead time of one month (e.g. the forecast for the season March-April-May is estimated using predictors from November-December-January)."

Page 3954, line 10: causal hindcast estimates. What do you mean by causal? Please rephrase of further explain.

Response: The explanation of the causal approach and the preference of this over a leave-one-out approach is made clear in Section 2.2

"The model is calibrated and validated in a hindcast framework using a causal ap-

proach: hindcasts are produced for 1961-2010 using data since 1901 prior to the hindcast start date. The causal approach was chosen instead of a leave-one-out framework in order to replicate the set of observational data that would have been available for each hindcast were it produced in real time."

Page 3954, lines 18-19: the incremental correlation attained by including additional predictors (second to eight lines). It is unclear if panels on lines 2-8 of Fig. 3 are for individual predictors or for a sequential cumulative addition of predictors. Please explain more precisely what is shown here.

Response: Revised text in Section 4.1 offers clarification.

"Hindcasts were produced with each predictor added in turn and verified against observations. Figure 3 shows the correlation between observations and a hindcast constructed using CO2-equivalent only (top line), the incremental correlation attained by including additional predictors cumulatively (second to eighth lines), and the observation-hindcast correlation following the inclusion of all predictors."

Page 3955, line 3: full correlation. What does this mean? Does it mean the correlation for the model that incorporate all predictions (i.e. bottom row in Fig. 3)? Please be more precise.

Response: Revised text in Section 4.1, paragraph 2 offer clarification.

"The correlation of observations with hindcasts estimated using CO2EQV (Figure 3, top line) only is much lower than that with hindcasts estimated using as a function of all potential predictors (Figure 3, bottom line)."

Pages 3965 and 3966, Figure 1 and 2: Figure panels are too small. It is currently difficult to appreciate the evidences presented in these figures. Please enlarge figure panels. The caption indicates one month lead time. Please be more precise in defining what is meant by one month lead time here. Does this mean the previous season predictor values are used to predict next season SAT and PREC?

Response: As mentioned above, all Figures will be changed to .pdf format in the revised manuscript. Again, in order to avoid confusing the reader, the reference to lead time is removed from the figure caption. The time difference between the predictand and predictors is made clear in Section 2.

Pages 3967 and 3968, Figures 3 and 4: For which period has this correlation been computed? Please provide this information in the figure caption. It is also unclear if the first 8 rows of this figure show the correlation skill considering only one predictor (i.e. the individual predictors indicated on the left side of each row). Please make sure the correct description is provided in the text and figure caption. Likewise, it is unclear if the last row of this figure shows the correlation skill considering all 8 predictors indicated in lines 1 to 8 above in the multiple linear regression model. Please make sure the correct description is provided in the text and figure caption. And unfortunately, because figure panels are too small, it is not possible to see the stippling indicated in the figure caption. Please enlarge figure panels to allow identification of statistically significant results. Make sure to provide references and or equations for the RMSESS and CRPSS shown in Figure 4.

Response: As mentioned above, all Figures will be changed to .pdf format in the revised manuscript. A reference for the skill scores is given in Section 2.2, paragraph 5.

Pages 3969 and 3970, Figure 5 and 6: Figure panels are too small. It is currently difficult to appreciate the evidences presented in these figures. Please enlarge figure panels. Figure 5 is apparently for SAT but caption indicates PREC. Please correct.

Response: As mentioned above, all Figures will be changed to .pdf format in the revised manuscript. Figure captions have been corrected accordingly.

Interactive comment on Geosci. Model Dev. Discuss., 8, 3941, 2015.