



The use of regression for assessing a seasonal forecast model experiment

Rasmus E. Benestad¹, Retish Senan², and Yvan Orsolini³

¹Norwegian Meteorological Institute, Oslo, 0313, Norway

²European Centre for Medium-range Forecasts, Reading, UK

³Norwegian Institute for Air Research, Kjeller, Norway

Correspondence to: Rasmus E. Benestad (rasmus.benestad@met.no)

Received: 6 April 2016 – Published in Earth Syst. Dynam. Discuss.: 31 May 2016

Revised: 27 September 2016 – Accepted: 3 October 2016 – Published: 10 November 2016

Abstract. We show how factorial regression can be used to analyse numerical model experiments, testing the effect of different model settings. We analysed results from a coupled atmosphere–ocean model to explore how the different choices in the experimental set-up influence the seasonal predictions. These choices included a representation of the sea ice and the height of top of the atmosphere, and the results suggested that the simulated monthly mean air temperatures poleward of the mid-latitudes were highly sensitivity to the specification of the top of the atmosphere, interpreted as the presence or absence of a stratosphere. The seasonal forecasts for the mid-latitudes to high latitudes were also sensitive to whether the model set-up included a dynamic or non-dynamic sea-ice representation, although this effect was somewhat less important than the role of the stratosphere. The air temperature in the tropics was insensitive to these choices.

1 Introduction

The question of whether seasonal forecasting has useful skill is getting increasingly relevant with the progress in climate modelling. Another question is how we can learn more about such skills, and one strategy is to examine the models used in seasonal forecasting. These include state-of-the-art coupled atmosphere–ocean–land–surface models, built on our knowledge of physical processes and formulated in terms of computer code (Palmer and Anderson, 1994; Stockdale et al., 1998; Palmer, 2004; George and Sutton, 2006). They can be used for seasonal forecasting if a correct initial state is provided, and from which the subsequent evolution can be simulated. Their skill depends on several factors, such as the quality of the initial states, the representation of all relevant processes, and whether the seasons ahead truly are predictable in the presence of non-linear chaos (Palmer, 1996). Thus, in order to address the initial question of useful skill for seasonal predictions, we need to understand what is important and what is irrelevant for the outcome of the predictions, which includes choices about the model set-up. Here we look at seasonal forecast results for the air temperature. We know

that the atmosphere in the high latitudes is subject to non-linear dynamics, and that the effect of different factors may interfere and amplify or dampen each other (Charney, 1947; Gill, 1982; Lindzen, 1990; Held, 1993; Feldstein, 2003).

1.1 Background

It is well known that numerical weather prediction (NWP) has a limited forecast horizon because small initial errors will grow over time in a non-linear fashion (Lorenz, 1963). The case for seasonal forecasting is somewhat different, as it relies on slow changes in the ocean and cryosphere, which act as persistent boundary conditions. NWP and seasonal forecasting represent two types of predictability referred to as “type 1” and “type 2” (Palmer, 1996). Whereas NWP is more an initial value problem (type 1), the seasonal forecasts embeds a degree of the boundary value problem aspect (type 2). Furthermore, seasonal forecasts tend to present the statistics of the weather over a given interval, rather than the exact state at any instant. In other words, seasonal forecasts can be compared with predicting a change in the statistics of a sample of measurements, whereas weather forecasting is more

like predicting the details about one specific data point in that sample.

Models used for seasonal forecasting have traditionally involved a model for the atmosphere coupled to an ocean component, and were originally developed for the tropical region and the El Niño–Southern Oscillation (Anderson, 1995; Stockdale et al., 1998; Palmer and Anderson, 1994). Aspects, such as sea ice, the stratosphere, and snow cover, were not emphasised as they were not believed to play an important role for the seasonal weather evolution. More recent studies have looked at the potential influence of sea ice (Balmaseda et al., 2010; Petoukhov and Semenov, 2010; Overland and Wang, 2010; Francis et al., 2009; Deser et al., 2004; Magnusdottir et al., 2004; Seierstad and Bader, 2008; Benestad et al., 2010; Orsolini et al., 2012), especially after the recent dramatic downward trends in the sea-ice extent (Kumar et al., 2010; Boé et al., 2010; Holland et al., 2008; Wilson, 2009; Kauker et al., 2009; Stroeve et al., 2007, 2008). Other studies have involved the effect of snow cover on the atmospheric circulation (Cohen and Entekhabi, 1999; Ge and Gong, 2009; Ueda et al., 2003; Hawkins et al., 2002; Watanabe and Nitta, 1998; Orsolini et al., 2013) or the influence of stratospheric conditions on the lower troposphere (Baldwin and Dunkerton, 2001; Baldwin et al., 2003; Thompson et al., 2002). Few of these studies, however, have looked at how these different factors in combination may interfere with each other, nor has there been many sensitivity tests for investigating how the model set-up, with different combinations of the components representing these different aspects, affects the results. One question we would like to address is whether the response to these different factors adds linearly or if the response is a non-linear function of these factors. Furthermore, it is interesting to find out which of these factors are more dominant than others. Moreover, our objective was to try to understand which *processes simulated by the model* are more important, rather than what real signals there are in nature. In this sense, this was a so-called *perfect model study* (Day et al., 2014). We present the combination of an experimental design (Williams, 1970; Kleijnen and Standridge, 1988; Kleijnen, 2015) and analytical techniques that can address this question. The results were taken from a “synthesis” experiment with a moderately high-resolution earth system model. Hence, these numerical experiments constitute a kind of sensitivity study (Bürger et al., 2013).

2 Method and data

2.1 Model simulations

The model used in this study was the EC-Earth version 2.1 state-of-the-art earth system model (Hazeleger et al., 2010), which had been developed by a consortium of meteorological institutes/universities across Europe. The atmospheric component of the EC-Earth model was based on ECMWF’s Integrated Forecasting System (IFS) cycle 31R1 with a new con-

vection scheme and a new land surface scheme. The ocean component was based on version 2 of the NEMO model (Madec, 2008), with a horizontal resolution of nominally $1 \times 1^\circ$ and 42 vertical levels. The sea-ice model was the LIM2 model (Fichefet and Maqueda, 1997). The ocean–ice model was coupled to the atmosphere–land model through the OASIS 3 coupler (Valcke, 2006).

The synthesis experiments consisted of a set of 12 coupled model simulations. Six of these simulations used the L62 vertical resolution for the atmospheric component, which extended up to 5 hPa, while the other six used the higher resolution L91 version, which extended up to 0.01 hPa. These two sets of experiments were designed to determine the sensitivity of model results to a better representation of the stratosphere. Further, to evaluate the role of sensitivity to the representation of sea ice, the LIM2 sea-ice model was implemented as a standard thermodynamic–dynamic model (DyIce) and as a thermodynamic-only model (NoDyIce). Finally, sensitivity to initial conditions was tested by introducing perturbations to initial conditions corresponding to positive/negative NAO SST (North Atlantic Oscillation sea surface temperature) anomaly patterns over the North Atlantic (Melsom, 2010). All simulations started on 1 January 1990 and lasted 90 days. The initial conditions used in this experiment came bundled with the earlier (test) versions of EC-Earth (up to V2.1) and were based on ERA-Interim. An overview of the model simulations is listed in Table A1.

2.2 The analysis

Here the experiments and analysis used an approach known as “factorial design” (Yates and Mather, 1963; Fisher, 1926; Hill and Lewicki, 2005; Wilkinson and Rogers, 1973; Benestad et al., 2010), where a factorial regression was used to assess which influence each of the choices in the model set-up has on the forecasts. It is a technique that can analyse sets of factors which are considered to have potential effects on the outcome in experiments, where an analysis of variance (ANOVA; Wilks, 1995) provides estimates for error bars and the level of statistical significance. Hence, factorial regression offers an alternative to traditional ways for estimating statistical significance used in meteorology and climate sciences, such as difference tests between two ensembles. Factorial regression can be applied to data that are generated by a process that involves two or more factors (set-up options or categories) and are difficult to quantify due to their discrete nature (e.g. some factors may either be present or absent). It has been used to analyse the effect of introducing different crop varieties in agriculture (e.g. Baril et al., 1995; Vargas et al., 1999, 2006; Voltas et al., 2005). It is based on the concept “factorial experiment”, or factorial design, in statistics, which involves two or more factors, each of which can be assigned a category or a discrete value. This kind of analysis takes all possible combinations of levels over all such factors including their interactions into account.

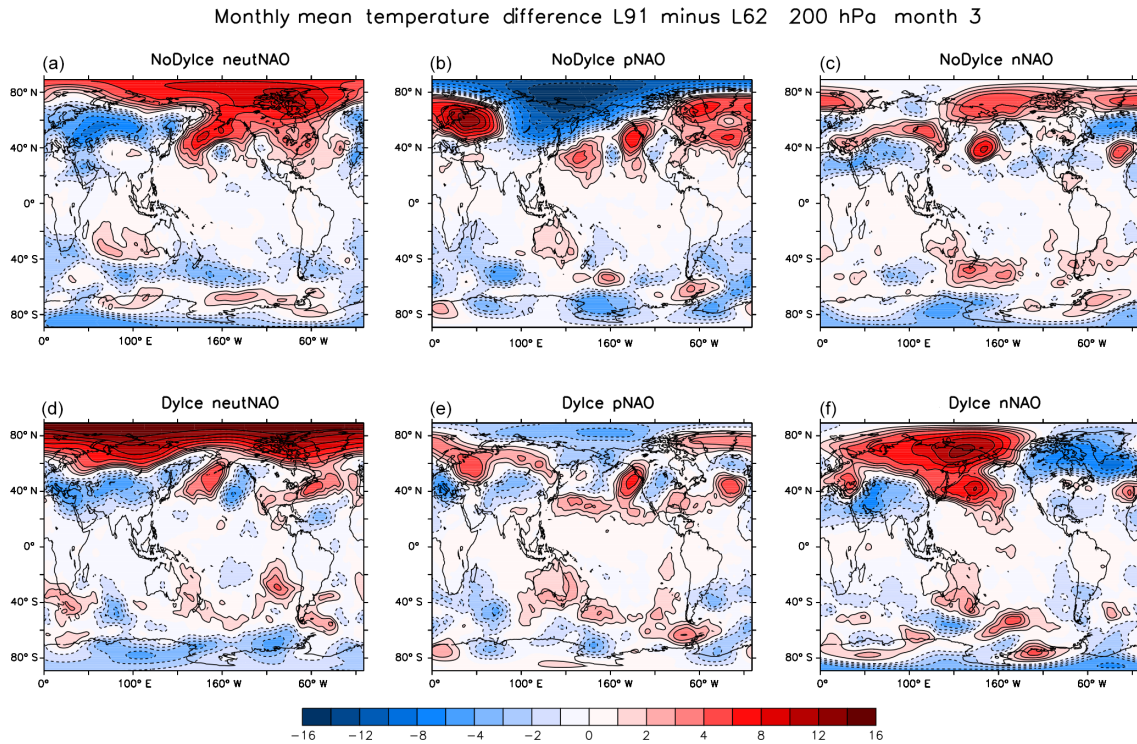


Figure 1. Map of monthly mean air temperature difference at 200 hPa between the high-top and low-top experiments for month 3.

The model response to different initial conditions or different model set-ups with different options for three configurations (SST perturbation, model top, and sea-ice model) was investigated, and a comparison was made between the different experiments in terms of vertical and horizontal cross sections of temperature anomalies. If the final response ΔT is a linear function of sea ice, SST, and stratospheric effects, then it can be expressed as a sum of these different contributions $\Delta T = x_1 C(\text{sea ice}) + x_2 C(\text{SST}) + x_3 C(\text{stratosphere})$, where $C(\dots)$ signifies the difference in outcome due to different choices in terms of one option setting. The factorial regression provided an estimate of the coefficients x_i and their error estimates. In a non-linear case, this linear expression is unlikely to provide a good description, and the regression analysis will yield large errors and low statistical significance.

We did not know the relative strength of the different factors in terms of an input; however, the factorial regression quantified the differences between output from different combinations of subsets. It was also used to estimate the probability that the response in the different combinations of these subsets would be due to chance. The results from the factorial regression were subsequently used to explore the combined effect of several factors.

The Walker test was used to assess the false discovery rate of the p values found in the factorial regression (Wilks, 2006). The test involved comparing the minimum p value p_n from the local tests with $p_W = 1 - (1 - \alpha)^{1/K}$ for K locations and the statistical significance level α . If $p_n < p_W$ then the

expected fraction of local null hypothesis with incorrect rejections is smaller than the number of statistically significant local p values.

3 Results

Figure 1 shows the difference in the forecasts' associated stratosphere, more specifically between the low-top (L62) and high-top (L91) versions of the atmosphere for month 3. It presents horizontal transects at the 200 hPa level, and shows the monthly mean temperature starting with a 2-month lead time. The left panels show results with no initial perturbation (neutral NAO conditions), the middle panels show results from model simulation with initial conditions set at a positive phase of NAO, and the right panels show results for which the initial conditions were the negative phase of the NAO. All the panels show that there were differences between the low- and high-top results, and the difference between the low- and high-top model simulation was most pronounced at negative and positive NAO-type initial conditions (not shown). Hence, the forecasted air temperature was sensitive to the inclusion of the upper part of the atmosphere, and the effect can be seen extending throughout the entire vertical extent of the atmosphere (not shown). The differences between the upper and lower rows show the effect of dynamic vs. non-dynamic sea-ice representation. With a non-dynamic sea ice, the inclusion of a stratosphere resulted in stronger vertical dipole patterns at certain longitudes and for positive NAO initial conditions.

For the negative NAO initial conditions, the dynamical sea-ice representation amplified the differences between the L91 and L62 model simulations.

Figure 1 suggests that the effect of including the stratosphere and the representation of sea-ice matter for the mid-latitude regions to the polar regions, and the choice of the vertical levels had less impact in the tropics. The response suggests mid-latitude wave-like structures in the 200 hPa temperatures, albeit with a tendency of a coherent anomaly over the North Pole. The choice of the sea-ice representation had a visible impact on the simulation of the monthly mean temperature after 3 months, seen as the difference between upper and lower panels. The horizontal picture at 200 hPa (Fig. 1) suggests radically different wave structure for the negative NAO phase, however, whereas for the respective “positive” and “neutral” NAO states, the differences were seen in both regional details and in magnitude. The exact geographical structure in these maps is not the important point here, as the longitude of action will depend on the initial condition. The important information here is the pronounced response in the mid-latitudes to high latitudes.

In summary, it is apparent from Fig. 1 that the effect of different model aspects such as the choice of model top and sea-ice representation influenced the model forecasts. Furthermore, we see that the influence varied with the initial SST conditions, and that different sea-ice representation introduced changes in the forecast of similar magnitude as the influence of the model top. It is difficult to compare these effects with that of the initial conditions merely from Fig. 1; however, we compared the effect from these different aspects through the means of a factorial regression. The ANOVA for the factorial regression yielded a set of coefficients β describing the association between the temperature and the model set-up choice, as well as the associated error bars' ε and p values.

Figure 2 presents the coefficients and the error estimates from the factorial regression. The top panel shows the mean air temperature for the model forecasts with a model set-up of dynamical sea-ice component, no perturbation in the SST, and 62 vertical levels (low top). Panels b–e show the differences in the forecasts due to different choices in the model set-up in terms of the regression coefficients β , and panels f–i show error estimates for these coefficients. Regions with large values estimated for the coefficients and large errors suggest a high sensitivity but also that the response cannot readily be attributed to the given factor. In other words, the level of both the signal and the noise is high. The magnitude of the error was mainly below 3 K except for around 100° E near the 100 hPa level, and generally smaller than the influence of the variable. The results suggested that the forecasts were sensitive to both the representation of the sea ice and the inclusion of the stratosphere, as well as the initial conditions. The analysis also suggested that the magnitude of the effect of the sea-ice representation and the model top was similar to those of the different SST perturbation near 60° N.

Furthermore, the error estimates associated with the three factors (SST perturbation, sea-ice representation, and atmosphere top) exhibited similar magnitudes and spatial structure. A comparison between the different panels in Fig. 2 suggests that the different choices for model set-up had similar magnitude on the predicted outcome for all these factors.

The previous results have indicated a high sensitivity to the various choices in the model set-up; however, we need to examine the relationship between the regression coefficients and error estimates in order to infer whether any has a systematic effect on the model predictions. Figure 3 shows the ratio response to error for sea ice (upper), positive NAO SST perturbation (second from the top), negative NAO SST perturbation (third), and the stratosphere L91 (bottom). Only a small region had a response that was greater in magnitude than the error estimate for the sea ice, whereas for the SST perturbations and the stratosphere, the regions where the response-to-error ratio had a magnitude greater to unity were more extensive. Both large negative and positive values indicate that the signal is stronger than the noise $|\beta/\varepsilon| > 1$, as β may be both positive and negative, whereas ε is positive.

The factorial regression gave the highest number of low p values for the stratosphere (L91), followed by the SST perturbation (not shown). For most of the 60° N vertical transect, the sea-ice representation did not yield a large response compared to the error term. Furthermore, for a global statistical significance level of $\alpha = 0.05$ and $K = 3840$, the threshold value for the Walker test was $p_W = 1.3 \times 10^{-5}$. The minimum p value for sea ice was 0.01, for SST perturbation $p_n = 9.2 \times 10^{-4}$ and the stratosphere $p_n = 1.6 \times 10^{-4}$. In other words, the 12-member experiment was not sufficient to resolve the response in the air temperature forecast at 60° N for month 3 to the different set-up options; however, they did suggest that the model top had the greatest impact on the forecast. The lack of a clear dependency between the sea-ice representation and the forecast was also found for the summer in Benestad et al. (2010), and the obscure links between the factors and the response may be explained by the presence of strong non-linear dynamics, where one given factor may result in different forecasts depending on other influences.

The question of degree of non-linearity can be addressed by comparing the sum of the influence from the different factors with simulations with and without a set of factors combined; i.e. we check for the equivalency:

$$\begin{aligned} DyIce\ pNAO\ L91 - NoDyIce\ nNAO\ L62 \\ \approx (DyIce - NoDyIce)\ nNAO\ L62 \\ + NoDyIce\ (pNAO - nNAO)\ L62 \\ + NoDyIce\ nNAO\ (L91 - L62). \end{aligned} \quad (1)$$

Here, the left-hand side of Eq. (1) (Fig. 4a) shows the difference between the simulation with high top, dynamic sea ice, and positive NAO perturbation (*DyIce pNAO L91*) and that with low top, non-dynamic sea ice, and negative NAO

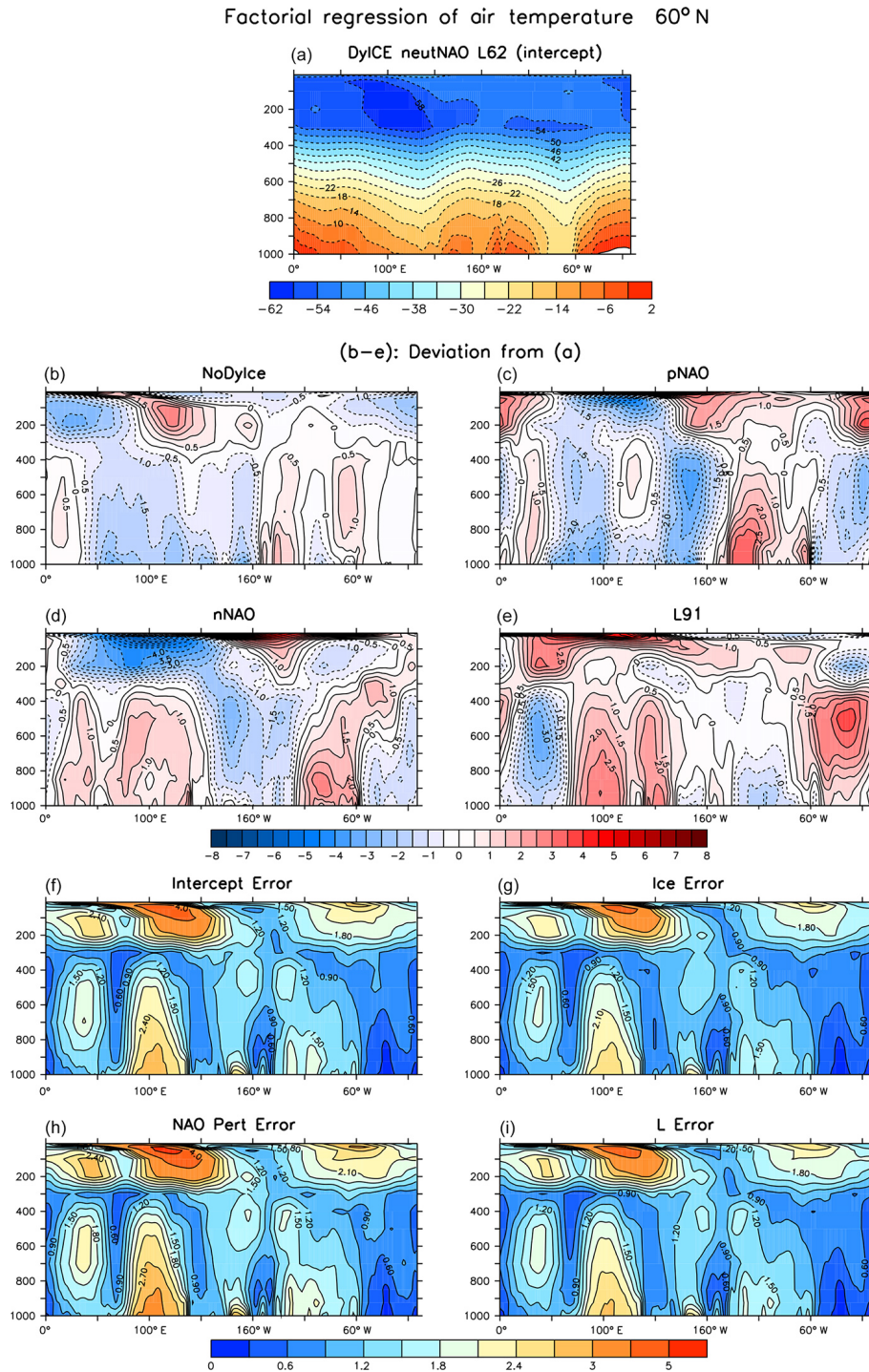


Figure 2. Coefficients and error estimates from the factorial regression of air temperature at 60° N. These results describe the systematic differences associated between the different choices in the model set-up.

(NoDylce nNAO L62). We compared Fig. 4a with the sum of the differences from individual factors (right-hand side of Eq. 1, Fig. 4b), and the comparison showed that the non-linear model response was mainly confined to the mid-latitudes to high latitudes, especially in the Northern Hemi-

sphere (Fig. 4c), e.g. along the 60° N transect presented in Fig. 3.

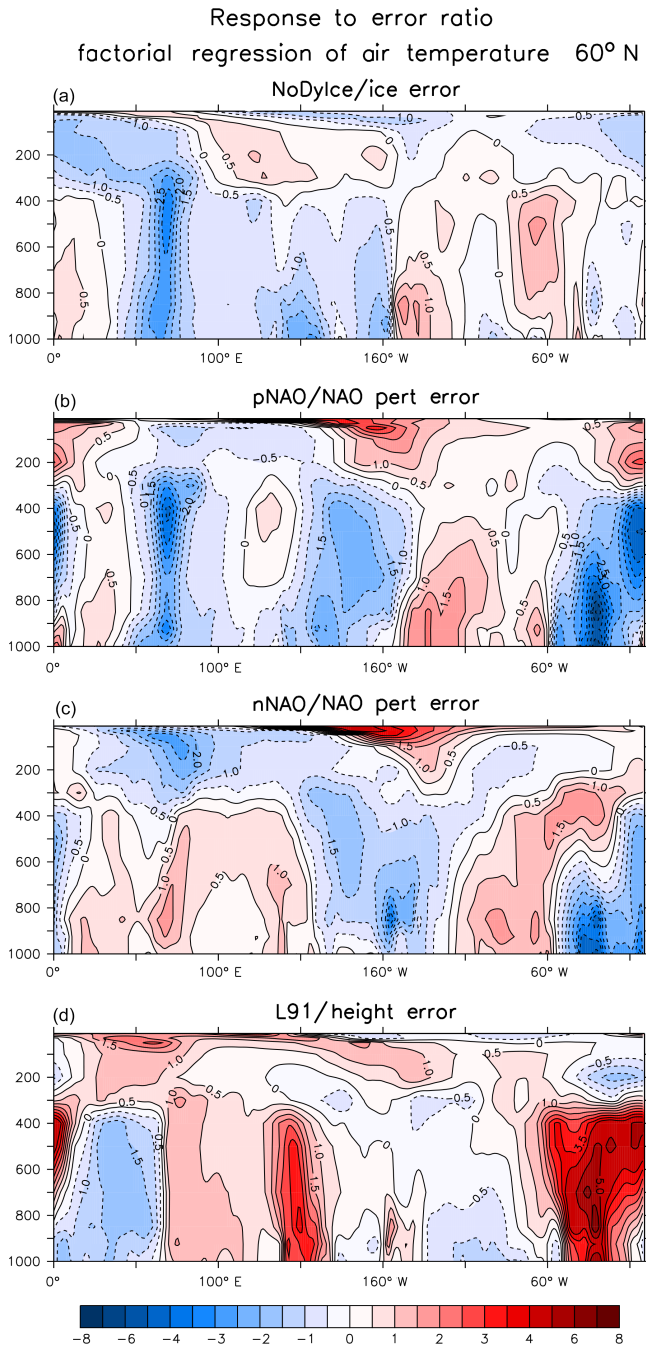


Figure 3. The ratio of the factorial regression coefficients to the error estimate for different factors: (a) sea-ice representation, (b) positive NAO SST perturbation, (c) negative NAO SST perturbation, and (d) the model top L91/stratosphere (bottom).

4 Discussion

The set of sensitivity experiments shows that seasonal forecasts at mid-latitudes to high latitudes are sensitive to a number of factors concerning the model set-up, and that the choice of subjective and subtle options can have as strong an

Monthly mean temperature 200 hPa month 3

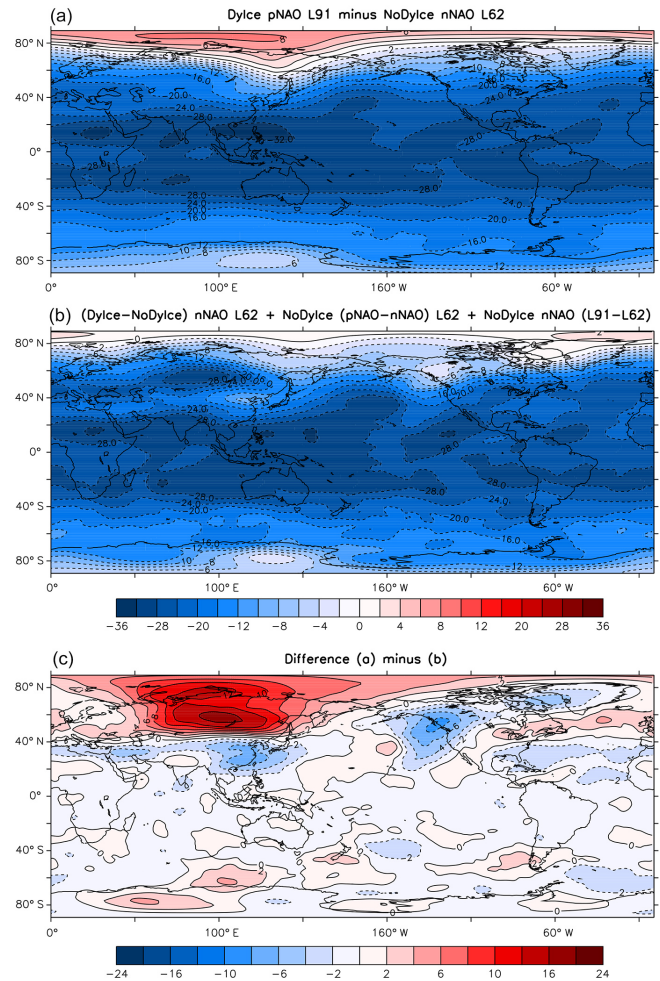


Figure 4. Monthly mean air temperature at 60° N. (a) Difference between Dylce pNAO L91 and NoDylce nNAO L62. (b) Sum of the differences: NoDylce (pNAO-nNAO) L62, (Dylce-NoDylce) nNAO L62, and NoDylce nNAO (L91-L62). (c) Difference between (a) and (b).

effect on the monthly mean temperature poleward of the mid-latitudes as the initial conditions. A factorial design experiment allows us to assess the relative magnitudes of different model height with that of different sea ice or different SST perturbations. We can also test the response in the model to see if it is close to being a linear superposition of the different single factors, or if the model response is highly non-linear. The statistical significance was estimated based on the factorial regression. The magnitude of the effect of the sea ice, SST perturbations, and the model top height were roughly similar, although the response to the sea ice was somewhat weaker than the others. The lower ratio of estimate-to-error also reflected the degree of non-linearity, and the relatively higher p values associated with the sea ice may be due to a greater degree of non-linearity in the response to the sea-

ice representation. The experiment nevertheless suggested that stratospheric conditions are important for mid- to high-latitude seasonal forecasting. This experiment was only carried out for the northern hemispheric winter, and may change with season. The stratosphere decouples in the summer, and there was a hint of a weaker influence from the model top in the Southern Hemisphere when it was summer.

There is previous work in which model sensitivity and uncertainty have been assessed (e.g. Rinke et al., 2000; Wu et al., 2005; Pope and Stratton, 2002; Jacob and Podzun 1997; Knutti et al., 2002; Dethloff et al., 2001); however, most of these assessments have been carried out for climate simulations as opposed to seasonal forecasts. In seasonal forecasting, the emphasis has been more on multi-model forecasts and their spread (Weisheimer et al., 2009), rather than the configuration of single models. However, Jung et al. (2012) discussed the effect of the spatial resolution on seasonal forecast based on an experimental design with a single model. The use of factorial regression was also discussed by Rinke et al. (2000) in conjunction with climate simulations, and Benestad et al. (2010) used it in a study of seasonal predictability and the effect of boundary conditions associated with sea ice and initial conditions. This study applied factorial regression to a new set of model configuration options, including the model top, the representation of sea ice, and initial conditions. In this case, we emphasised the individual factors rather than their interaction because of the limited sample of model runs. An inclusion of these interactive factors can give an indication of the effects of changing more than one option at the time (given a sufficient sample), e.g. how the combination of different vertical extent, sea ice model, and initial conditions results in a different outcome. However, we addressed this issue separately in this study by comparing the different terms in Eq. (1), which indeed suggested that the results from changing more than one factor give a non-linear response. These aspects require more efforts to form a better understanding, both in terms of larger ensemble experiments and understanding of the physics involved. However, the objective here was to try to find potential additional explanations for why seasonal forecasting has been associated with such low skill in mid-latitudes, in addition to the higher degree of non-linear dynamics in connection to weather patterns.

These experiments involved global coupled atmosphere–ocean models that are used for operational seasonal forecasting, especially for the El Niño–Southern Oscillation (ENSO); however, our analysis focused on the mid-latitudes. The results nevertheless allow for a comparison between the tropics and higher latitudes. They suggest that the outcome of the predictions in the mid-latitudes is sensitive to the choice of the top of the atmosphere and the representation of sea ice, but the low latitudes are insensitive to these factors. Hence, they support the hypothesis that the lack of seasonal prediction skill reported in the mid-latitudes may be linked to non-optimal model configuration. Further insight from these experiments moreover includes (1) that subjective choices in terms of model set-up (vertical levels and type of sea-ice representation) have an effect on the outcome of the seasonal forecasts in the high latitudes, (2) that factorial regression can be used as a means to describe the effect of different model options, and (3) that the effect of these different choices results in a non-linear response. These aspects have rarely been discussed in the past, perhaps because they do not have a strong effect on the simulation of processes in the tropics (e.g. ENSO).

5 Conclusions

A set of sensitivity tests revealed that seasonal predictability of the temperature at the mid-latitudes to high latitudes was as sensitive to subjective choices regarding the model set-up as the initial SST conditions. Hence, these results illustrate the difficulties associated with seasonal forecasting at the higher latitudes with an effect on the forecast skill. The tropical temperatures were insensitive to these choices, and the sea-ice representation and the stratosphere do not have a visible effect on, e.g., ENSO forecasts.

6 Data availability

The data presented here are available from <http://www.figshare.com>, doi:10.6084/m9.figshare.4131375s.

Appendix A

Experiment	Description
1. <i>DyIce neutNAO L62</i>	EC-Earth with L62 vertical resolution and no perturbations to initial conditions and a thermodynamic–dynamic LIM2 sea-ice model
2. <i>NoDyIce neutNAO L62</i>	Same as above but with thermodynamic-only sea-ice model
3. <i>DyIce neutNAO L91</i>	Same as 1. above but with L91 vertical resolution
4. <i>NoDyIce neutNAO L91</i>	Same as 2. above but with L91 vertical resolution
5. <i>DyIce pNAO L62</i>	Same as 1. above but with perturbation to initial condition corresponding to a positive NAO SST anomaly pattern over the North Atlantic
6. <i>NoDyIce pNAO L62</i>	Same as 5. above but with thermodynamic-only sea-ice model
7. <i>DyIce pNAO L91</i>	Same as 5. above but with L91 vertical resolution
8. <i>NoDyIce pNAO L91</i>	Same as 6. above but with L91 vertical resolution
9. <i>DyIce nNAO L62</i>	Same as 5. above but with perturbation to initial condition corresponding to a negative NAO SST anomaly pattern over the North Atlantic
10. <i>NoDyIce nNAO L62</i>	Same as 9. above but with thermodynamic-only sea-ice model
11. <i>DyIce nNAO L91</i>	Same as 9. above but with L91 vertical resolution
12. <i>NoDyIce nNAO L91</i>	Same as 10. above but with L91 vertical resolution

Acknowledgements. We are grateful to Wilco Hazeleger and the EC-Earth community for providing a stand-alone version of the EC-Earth model, and Simona Stefanescu at the ECMWF for all her assistance. Comments from two reviewers have also improved this paper. This work was carried out under the SPAR project (“Seasonal Predictability over the Arctic Region – exploring the role of boundary conditions”); project 178570, funded by the Norwegian Research Council and the Meteorological Institute) and SPECS (EU Grant Agreement 3038378), and the model simulations used computational resources at NOTUR – the Norwegian Metacenter for Computational Science. The data used in this analysis can be obtained by contacting the authors.

Edited by: B. Kravitz

Reviewed by: two anonymous referees

References

- Anderson, D. L. T.: The Legacy of TOGA: Seasonal Climate Prediction, in: *Predictability*, edited by: ECMWF, Vol. 2, Seminar Proceedings, ECMWF, 1995.
- Baldwin, M. P. and Dunkerton, T. J.: Stratospheric Harbingers of Anomalous Weather Regimes, *Science*, 294, 581–584, 2001.
- Baldwin, M. P., Stephenson, D. B., Thompson, D. W. J., Dunkerton, T. J., Charlton, A. J., and O’Neil, A.: Stratospheric Memory and Skill of Extended-Range Weather Forecasts, *Science*, 301, 636–640, 2003.
- Balmaseda, M. A., Ferranti, L., Molteni, F., and Palmer, T. N.: Impact of 2007 and 2008 Arctic ice anomalies on the atmospheric circulation: Implications for long-range predictions, *Q. J. Roy. Meteor. Soc.*, 136, 1655–1664, doi:10.1002/qj.661, 2010.
- Baril, C. P., Denis, J.-B., Wustman, R., and van Eeuwijk, F. A.: Analysing Genotype by Environment Interaction in Dutch Potato Variety Trials Using Factorial Regression, *Euphytica*, 84, 23–29, doi:10.1007/BF01677553, 1995.
- Benestad, R. E., Senan, R., Balmaseda, M., Ferranti, L., Orsolini, Y., and Melsom, A.: Sensitivity of summer 2-m temperature to sea ice conditions, *Tellus A*, 63, 2, 324–337, doi:10.1111/j.1600-0870.2010.00488.x, 2010.
- Boé, J., Hall, A., and Qu, X.: Sources of spread in simulations of Arctic sea ice loss over the twenty-first century, *Clim. Change*, 99, 637–645, doi:10.1007/s10584-010-9809-6, 2010.
- Bürger, G., Sobie, S. R., Cannon, A. J., Werner, A. T., and Murdock, T. Q.: Downscaling Extremes: An Intercomparison of Multiple Methods for Future Climate, *J. Climate*, 26, 3429–3449, doi:10.1175/JCLI-D-12-00249.1, 2013.
- Charney, J. G.: The Dynamics of Long Waves in a Baroclinic Westerly Current, *J. Meteorol.* 4, 135–163, 1947.
- Cohen, J. and Entekhabi, D.: Eurasian snow cover variability and northern hemisphere climate predictability, *Geophys. Res. Lett.*, 26, 345–348, doi:10.1029/1998GL900321, 1999.
- Day, J. J., Hawkins, E., and Tietsche, S.: Will Arctic Sea Ice Thickness Initialization Improve Seasonal Forecast Skill?, *Geophys. Res. Lett.*, 41, 7566–7575, doi:10.1002/2014GL061694, 2014.
- Deser, C., Magnusdottir, G., Saravanan, R., and Phillips, A.: The Effects of North Atlantic SST and Sea Ice Anomalies on the Winter Circulation in CCM3. Part II: Direct and Indirect Components of the Response, *J. Clim.*, 17, 877–889, 2004.
- Dethloff, K., Abegg, C., Rinke, A., Hebestadt, I., and Romanov, V. F.: Sensitivity of Arctic Climate Simulations to Different Boundary-Layer Parameterizations in a Regional Climate Model, *Tellus A*, 53, 1–26, doi:10.1034/j.1600-0870.2001.01073.x, 2001.
- Feldstein, S. B.: The Dynamics of NAO Teleconnection Pattern Growth and Decay, *Q. J. Roy. Meteor. Soc.*, 129, 901–924, doi:10.1256/qj.02.76, 2003.
- Fichefet, T. and Maqueda, M. A. M.: Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics, *J. Geophys. Res.*, 102, 12609–12646, doi:10.1029/97JC00480, 1997.
- Fisher, R.: The Arrangement of Field Experiments, *Journal of the Ministry of Agriculture of Great Britain*, 33, 503–513, 1926.
- Francis, J. A., Chan, W., Leathers, D. J., Miller, J. R., and Veron, D. E.: Winter Northern Hemisphere weather patterns remember summer Arctic sea-ice extent, *Geophys. Res. Lett.*, 36, L07503, doi:10.1029/2009GL037274, 2009.
- Ge, Y. and Gong, G.: North American Snow Depth and Climate Teleconnection Pattern, *J. Clim.*, 22, 217–233, doi:10.1175/2008JCLI2124.1, 2009.
- George, S. E. and Sutton, R. T.: Predictability and Skill of Boreal Winter Forecasts Made with the ECMWF Seasonal Forecast System II, *Q. J. Roy. Meteor. Soc.*, 132, 2031–2053, 2006.
- Gill, A. E.: *Atmosphere-Ocean Dynamics*, International Geophysics Series. San Diego, California, Academic Press, 662 pp., 1982.
- Hawkins, T. W., Ellis, A. W., Skindlov, J. A., and Reigle, D.: Interannual Analysis of the North American Snow Cover-Monsoon Teleconnection: Seasonal Forecasting Utility, *J. Clim.*, 15, 1743–1753, 2002.
- Hazeleger, W., Severijns, C., Semmler, T., Stefanescu, S., Yang, S., Wang, X., Wyser, K., Dutra, E., Baldasano, J. M., Bintanja, R., Bougeault, P., Caballero, R., Ekman, A. M. L., Christensen, J. H., van den Hurk, B., Jimenez, P., Jones, C., Kållberg, P., Koenigk, T., McGrath, R., Miranda, P., Van Noije, T., Palmer, T., Parodi, J. A., Schmith, T., Selten, F., Storelvmo, T., Sterl, A., Tapamo, H., Vancoppenolle, M., Viterbo, P., and Willén, U.: EC-Earth: A Seamless Earth-System Prediction Approach in Action, *B. Am. Meteor. Soc.*, 91, 1357–1363, doi:10.1175/2010BAMS2877.1, 2010.
- Held, I. M.: Large-Scale Dynamics and Global Warming, *B. Am. Meteor. Soc.*, 74, 228–241, 1993.
- Hill, T. and Lewicki, P.: *Statistics: Methods and Applications?: a Comprehensive Reference for Science, Industry, and Data Mining*, ISBN 1884233597, Tulsa, OK, USA, StatSoft, 2005.
- Holland, M. M., Serreze, M. C., and Stroeve, J.: The sea ice mass budget of the Arctic and its future change as simulated by coupled climate models, *Clim. Dynam.*, 34, 185, doi:10.1007/s00382-008-0493-4, 2008.
- Jacob, D. and Podzun, R.: Sensitivity Studies with the Regional Climate Model REMO, *Meteor. Atmos. Phys.*, 63, 119–129, doi:10.1007/BF01025368, 1997.
- Jung, T., Miller, M. J., Palmer, T. N., Towers, P., Wedi, N., Achuthavarier, D., Adams, J. M., Altshuler, E. L., Cash, B. A., Kinter, J. L., Marx, L., Stan, C., and Hodges, K. I.: High-Resolution Global Climate Simulations with the ECMWF Model in Project Athena: Experimental Design, *Model Cli-*

- mate, and Seasonal Forecast Skill, *J. Climate*, 25, 3155–3172, doi:10.1175/JCLI-D-11-00265.1, 2012.
- Kauker, F., Kaminski, T., Karcher, M., Giering, R., Gerdes, R., and Voßbeck, M.: Adjoint analysis of the 2007 all time Arctic sea-ice minimum, *Geophys. Res. Lett.*, 36, L03707, doi:10.1029/2008GL036323, 2009.
- Kleijnen, J. P. C. and Standridge, C. R.: Experimental Design and Regression Analysis in Simulation: An FMS Case Study, *Eur. J. Oper. Res.*, 33, 257–261, doi:10.1016/0377-2217(88)90168-3, 1988.
- Kleijnen, J. P. C.: Design and Analysis of Simulation Experiments; Second Edition, Springer, ISBN 978-3-319-18087-8, 2015.
- Knutti, R., Stocker, T. F., Joos, F., and Plattner, G.-K.: Constraints on Radiative Forcing and Future Climate Change from Observations and Climate Model Ensembles, *Nature*, 416, 719–723, doi:10.1038/416719a, 2002.
- Kumar, A., Perlwitz, J., Eischeid, J., Quan, X., Xu, T., Zhang, T., Hoerling, M., Jha, B., and Wang, W.: Contribution of sea ice loss to Arctic amplification, *Geophys. Res. Lett.*, 37, L21701, doi:10.1029/2010GL045022, 2010.
- Lindzen, R. S.: Dynamics in Atmospheric Physics, Cambridge, UK, Cambridge University Press, 324 pp., 1990.
- Lorenz, E.: Deterministic Nonperiodic Flow, *J. Atmos. Sci.*, 20, 130–141, 1963.
- Madeo, G.: NEMO ocean engine, Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No. 27, 300 pp., 2008.
- Magnusdottir, G., Deser, C., and Saravanan, R.: The Effects of North Atlantic SST and Sea Ice Anomalies on the Winter Circulation in CCM3. Part I: Main Features and Storm Track Characteristics of the Response, *J. Clim.*, 17, 857–875, 2004.
- Melsom, A.: Perturbing the ocean initial state from NAO regression, Met. no. report, 1/2010, 8 pp., 2010.
- Orsolini, Y. J., Senan, R., Benestad, R. E., and Melsom, A.: Autumn atmospheric response to the 2007 low Arctic sea ice extent in coupled ocean-atmosphere hindcasts, *Clim. Dynam.*, 38, 2437–2448, doi:10.1007/s00382-011-1169-z, 2012.
- Orsolini, Y. J., Senan, R., Carrasco, A., Balsamo, G., Doblas-Reyes, F. J., Vitart, F., Weisheimer, A., and Benestad, R. E.: Impact of snow initialization on sub-seasonal forecasts, *Clim. Dynam.*, 41, 1969, doi:10.1007/s00382-013-1782-0, 2013.
- Overland, J. E. and Wang, M.: Large-scale atmospheric circulation changes are associated with the recent loss of Arctic sea ice, *Tellus A*, 61, 1, doi:10.1111/j.1600-0870.2009.00421.x, 2010.
- Palmer, T. N.: Predictability of the Atmosphere and Oceans: from Days to Decades, in: Decadal Variability, edited by: Anderson, D. L. T. and Willebrand, J., Springer, NATO ASI series, 44, 83 pp., 1996.
- Palmer, T. N.: Progress towards Reliable and Useful Seasonal and Interannual Climate Prediction, *WMO Bull.*, 53, 325–333, 2004.
- Palmer, T. N. and Anderson, D. L. T.: The Prospect for Seasonal Forecasting – A Review Paper, *Q. J. Roy. Meteor. Soc.*, 120, 755–793, 1994.
- Petoukhov, V. and Semenov, V. A.: A link between reduced Barents-Kara sea ice and cold winter extremes over northern continents, *J. Geophys. Res.*, 115, D21111, doi:10.1029/2009JD013568, 2010.
- Pope, V. and Stratton, R.: The Processes Governing Horizontal Resolution Sensitivity in a Climate Model, *Clim. Dynam.*, 19, 211–236, doi:10.1007/s00382-001-0222-8, 2002.
- Rinke, A., Lynch, A. H., and Dethloff, K.: Intercomparison of Arctic Regional Climate Simulations: Case Studies of January and June 1990, *J. Geophys. Res.-Atmos.*, 105, 29669–26683, doi:10.1029/2000JD900325, 2000.
- Seierstad, I. A. and Bader, J.: Impact of a projected future Arctic Sea Ice reduction on extratropical storminess and the NAO, *Clim. Dynam.*, 33, 937–943, doi:10.1007/s00382-008-0463-x, 2008.
- Stockdale, T. N., Anderson, D. L. T., Alves, J. O. S., and Balmaseda, M. A.: Global Seasonal Rainfall Forecasts Using a Coupled Ocean-Atmosphere Model, *Nature*, 392, 370–373, 1998.
- Stroeve, J., Holland, M. M., Meier, W., Scambos, T., and Serreze, M.: Arctic sea ice decline: Faster than forecast, *Geophys. Res. Lett.*, 34, L09501, doi:10.1029/2007GL029703, 2007.
- Stroeve, J., Serreze, M., Drobot, S., Gearheard, S., Holland, M., Maslanik, J., Meier, W., and Scambos, T.: Arctic Sea Ice Extent Plummet in 2007, *Eos*, 89, 13–14, 2008.
- Thompson, D. W. J., Baldwin, M. P., and Wallace, J. M.: Stratospheric Connection to Northern Hemisphere Wintertime Weather: Implications for Prediction, *J. Climate*, 15, 1421–1428, 2002.
- Ueda, H., Shinoda, M., and Kamahori, H.: Spring northward retreat of Eurasian snow cover relevant to seasonal and interannual variations of atmospheric circulation, *Int. J. Clim.*, 23, 615–629, 2003.
- Valcke, S.: OASIS3 User Guide (prism_2–5), PRISM Support Initiative Report No. 3, 64 pp., 2006.
- Vargas, M., Crossa, J., van Eeuwijk, F. A., Ramírez, M. E., and Sayre, K.: Using Partial Least Squares Regression, Factorial Regression, and AMMI Models for Interpreting Genotype – Environment Interaction, *Crop Sci.*, 39, 955–967, doi:10.2135/cropsci1999.0011183X003900040002x, 1999.
- Vargas, M., van Eeuwijk, F. A., Crossa, J., and Ribaut, J.-M.: Mapping QTLs and QTL – Environment Interaction for CIMMYT Maize Drought Stress Program Using Factorial Regression and Partial Least Squares Methods, *Theor. Appl. Gen.*, 112, 1009–1023, doi:10.1007/s00122-005-0204-z, 2006.
- Voltas, J., López-Córcoles, H., and Borrás, G.: Use of Biplot Analysis and Factorial Regression for the Investigation of Superior Genotypes in Multi-Environment Trials, *Eur. J. Agron.*, 22, 309–324, doi:10.1016/j.eja.2004.04.005, 2005.
- Watanabe, M. and Nitta, T.: Relative Impacts of Snow and Sea Surface Temperature Anomalies on an Extreme Phase in Winter Atmospheric Circulation, *J. Clim.*, 11, 2837–2857, 1998.
- Weisheimer, A., Doblas-Reyes, F. J., Palmer, T. N., Alessandri, A., Arribas, A., Déqué, M., Keenlyside, N., MacVean, M., Navarra, A., and Rogel, P.: 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, L21711, doi:10.1029/2009GL040896, 2009.
- Wilkinson, G. N. and Rogers, C. E.: Symbolic Description of Factorial Models for Analysis of Variance, *Appl. Stat.* 22, 392–399, 1973.
- Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Orlando, Florida, USA, Academic Press, 649 pp., 1995.
- Wilks, D. S.: On Field Significance and the False Discovery Rate, *J. Appl. Meteorol. Climatol.*, 45, 1181–1189, doi:10.1175/JAM2404.1, 2006.
- Williams, J. D.: A Regression Approach to Experimental Design, *J. Exp. Educ.*, 39, 83–90, 1970.

Wilson, M.: Satellite altimetry quantifies the alarming thinning of Arctic sea ice, *Phys. Today*, 62, 19–21, 2009.

Wu, W., Lynch, A. H., and Rivers, A.: Estimating the uncertainty in a regional climate model related to initial and lateral boundary conditions, *J. Climate*, 18, 917–933, doi:10.1175/JCLI-3293.1, 2005.

Yates, F. and Mather, K.: Ronald Aylmer Fisher, 1890–1962, *Biographical Memoirs of Fellows of the Royal Society*, 9, 91–129. doi:10.1098/rsbm.1963.0006, 1963.