Visualization of 2D uncertainty in decadal climate predictions

Michael Böttinger[1], Holger Pohlmann[2], Niklas Röber[1], Karin Meier-Fleischer[1], and Dela Spickermann[1]

[1]Deutsches Klimarechenzentrum GmbH, Hamburg, Germany
[2]Max Planck Institute for Meteorology, Hamburg, Germany

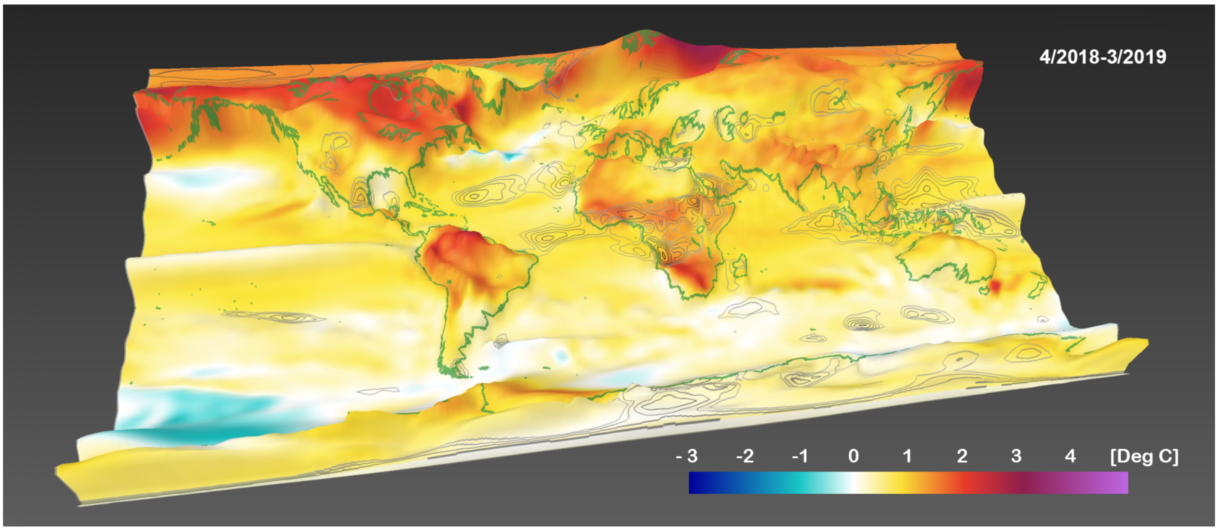


Figure 1: Visualization of predicted 2m temperature anomaly (colors), forecast skill (isolines) and ensemble spread (height).

Abstract

In recent years, climate prediction systems based on coupled climate models are used for investigating the climate predictability on a decadal time scale. Based on ensemble simulation techniques applied and hindcast experiments carried out first, the predictive skill of a system can be derived. The ensemble simulations used for the decadal climate predictions enable the issuing of probabilistic information along with the quantities predicted. In this work, we focus on the concurrent visualization of three related 2D fields: the forecast variable, here the 2m temperature anomaly, along with the corresponding predictive skill and the ensemble spread. We show exemplary solutions produced with three different visualization systems: NCL, Avizo Green and ParaView.

Categories and Subject Descriptors (according to ACM CCS): J.2 [Computer Applications]: Earth and atmospheric sciences—I.3.8 [Computing Methodologies]: Computer Graphics—Applications

1. Introduction

Due to the chaotic features of weather, meteorological data can be highly variable in space and time. Deterministic weather forecasts based on atmosphere models are only reliable for short periods of time. The ability of weather models to correctly meet spatiotemporal weather developments decreases with increasing forecast periods. Although deterministic forecasts are only possible for short time scales, the *climate* – the statistical features of the weather on longer time scales – can be well simulated with *coupled models* of the climate- or Earth system, respectively, which include interactions and feedback among its different components.

Climate projections for the coming centuries such as, e.g., discussed in the Assessment Reports of the Intergovernmental Panel on Climate Change (IPCC) [oCC14], are started randomly from a long control simulation since the initial conditions are unimportant on centennial time scales. On shorter time scales, however, the initial conditions determine climate predictions significantly. Therefore, decadal climate predictions are initialized from actual observations.

By comparing predictions of these model systems for past initialization dates in so-called hindcast experiments, observed past data can be used to determine the forecast skill achieved. The forecast skill is a spatial pattern that changes with time. For areas with high skill, the uncertainty in the predictions due to internal variability can be reduced. The ensemble spread represents the internal climate variability simulated by the model and provides us with additional useful information. It can also be interpreted as the probability of future values occurring in a certain range. For a quick visual analysis of the forecast variable, the corresponding skill, as well as of the ensemble spread, a visualization is required that concurrently shows the temporal evolution of all three 2D fields.

2. Related work

2.1. Uncertainty visualization

During the last two decades, the need to evaluate data along with its uncertainty has gained in importance in many scientific disciplines. Consequently, uncertainty visualization has become an active research topic in the visualization community. Using examples taken from different application areas, [BOL12] and [BHJ*14] give a general overview of the current state-of-the-art developments in uncertainty visualization. Specifically for ensemble data, [OJ14] define two categories of visualization approaches: *feature-based* and *location-based* methods. In the first category, features are extracted from individual ensemble members and then visually combined. In the latter method, statistical properties of the ensemble are computed for each grid point and the resulting fields are visualized. [PWB*09] and [SZD*10], e.g., use both techniques and also make use of *linked views* in order to enable the interactive visual analysis of statistical properties together with the forecast variables.

2.2. Decadal climate predictions

Decadal climate prediction has been recognized to be potentially important for society and decision making. Therefore, an exercise of retrospective predictions (hindcasts) over the past 50 years was performed for the 5th Coupled Model Intercomparison Project CMIP5 [TSM12] to be analyzed in the 5th Assessment Report (AR5) [KP13] of the Intergovernmental Panel on Climate Change (IPCC). For decadal predictions, both the initial conditions of the climate system and changing radiative forcing are important.

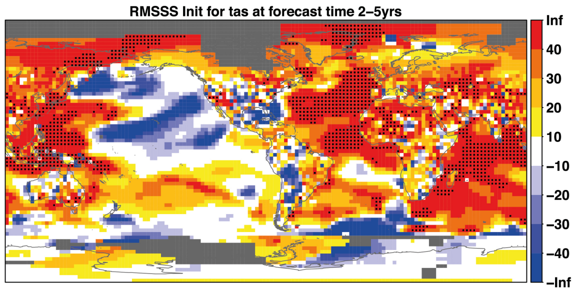


Figure 2: 2D visualization of two quantities using the overlay technique (Figure 11.4 from IPCC AR5 [KP13]). Here, a root mean square skill score is shown by color-filled grid cells. The overlayed black dots mark regions with a statistical significance of 95%.

2.3. Visualizations used in the domain

The UK's Met Office coordinates an informal exchange of near-real-time decadal predictions [SSB*13], which are regularly issued over the internet (http://www.metoffice.gov.uk/research/climate/seasonal-to-decadal/long-range/decadal-multimodel). However, also due to the lack of a standard for skill estimates, the predictions are issued without any skill information. Estimates for uncertainties can only be obtained by the visual comparison of the single-model predictions with each other or with the multi-model mean. Although an overlay display of the uncertainties, the multi-model skill, or the skill of the individual systems would be very useful to evaluate and interpret the results, only single-model results are currently presented side-by-side with the multi-model mean.

However, combined visualizations of physical variables and corresponding uncertainty information have been used for many years in the climate community. Here, the dimension reduction is achieved by the overlay technique: filled contours are mostly used for the physical variable, and line-based techniques such as contour lines or stippling are overlayed to display the statistical information. Examples in literature are mostly restricted to two fields that are visually combined, see e.g. Figure 2 ([KP13]).

3. Data and Methodology

The decadal climate predictions of the Max Planck Institute for Meteorology (MPI) for CMIP5/AR5 are based on the Earth System Model *MPI-ESM* [SGE*13]. Within the MiKlip project (http://www.fona-miklip.de/en/), the decadal prediction system has been improved in various aspects [PMK*13]. The initialization is now based on the ORAS4 ocean reanalyses [BMW13] for the ocean, and additionally on ERA40 [UKS*05] until 1989 and ERA-Interim [DUS*11] thereafter for the atmosphere. The model resolution is T63/L47 in the atmosphere and 1.5 degrees/L40 in the ocean. An ensemble of 10 members is produced with yearly initialization using lagged days around the 1st of January between 1961 and 2015.

The predictions for the global 2m temperature are stored as NetCDF files for each ensemble member on a monthly basis. First, anomalies are calculated relative to the mean climatology of the period 1961-2010. Second, the anomalies are low-pass filtered with a one-year running mean. Third, the ensemble mean, the spread – which is defined here as the ensemble standard deviation for a certain lead time –, and the skill are calculated. For the evaluation of the skill of the hindcasts we use the Pearson's correlation coefficient (e.g. [Wil11]), defined here as

$$cor_t = \frac{\sum x_{it} o_i}{\sqrt{\sum x_{it}^2}\sqrt{\sum o_i^2}}$$

with x_{it} being the anomaly for the ensemble mean hindcast at a given lead time t and certain initialization i, and with o_i being the corresponding anomaly of the observation. This verification method is also used e.g. by [GKS*13].

The derived predictive skill and the ensemble spread are two different aspects of the forecast uncertainty. Both vary with time. Areas with high standard deviation can be interpreted as areas in which small changes in the initial conditions can result in completely different temperature values, and hence the probability for a good prediction – the skill – is low. Regions with high skill values are therefore almost only present in areas with low ensemble spread even though low ensemble spread does not guarantee for a high skill.

For a better understanding of the prediction system and its statistical properties, we aim to concurrently visualize the time-dependent 2-dimensional prediction data consisting of three different variables. While the combination of two 2D maps is straightforward with the techniques used in the domain, the combination of three 2D maps (here: prediction, skill and ensemble spread) is challenging because overlayed elements can occlude each other.

4. Results

We used three different visualization solutions to create exemplary visualizations for the different tools and techniques: NCL, Avizo Green and ParaView. All of these systems directly support NetCDF model data. For displaying the predicted 2m temperature anomaly, we always used a symmetric red-white-blue colormap to highlight positive (red) and negative anomalies (blue).

NCL (NCAR Command Language) [UCA14] is an interpreted language specifically for analyzing and visualizing geo-scientific data. The software was developed at the National Center for Atmospheric Research (NCAR) and is freely available (http://www.ncl.ucar.edu/). The threshold of 0.5 has arbitrarily been chosen for the isocontours, representing a very high confidence in prediction skill. This allows to apply a second line-based technique for an overlay

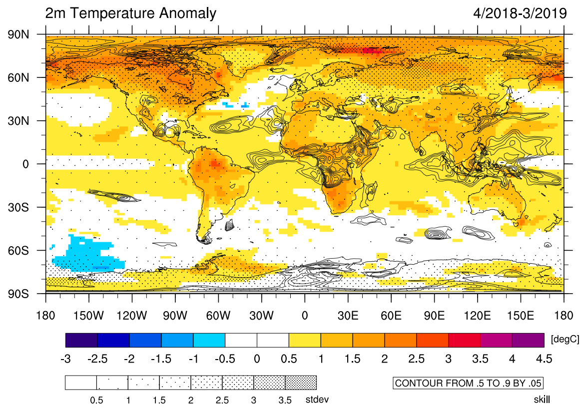


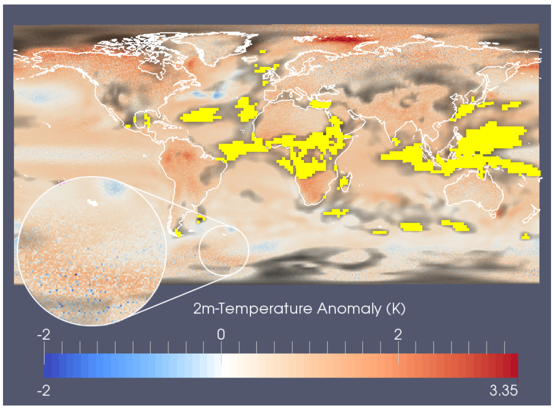
Figure 3: 2D visualization of three quantities using the overlay technique realized with NCL.

display of the ensemble spread because in our case high skill values can only be achieved in regions with low standard deviation. As shown in Figure 3, we applied a stippling technique using seven different density levels for the visualization of the standard deviation.

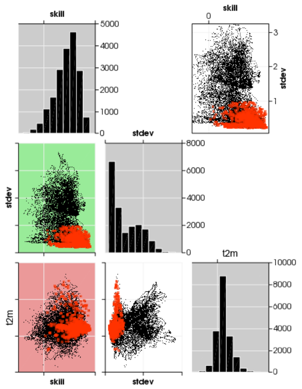
The commercial 3D visualization system Avizo Green offers various state-of-the-art visualization techniques. We used the height field method to display the ensemble spread, and color coding for displaying the surface temperature anomaly. For the visualization of the predictive skill we extracted isolines for skill values greater than or equal to 0.5. For mapping the isolines onto the height field, an intermediate step was necessary: first, a series of textures with the color coding, the isolines and the continental outlines was created, which were then texture mapped onto the according height field geometries. The result is shown in Figure 1. Finally, an animation was created to account for the time dependence of the data.

Using the free available software ParaView [Kit15], we performed a detailed *interactive* visual data analysis. Figure 4 shows four screens of this analysis process, with Figure 4a displaying a 2D visualization comparable to the results shown earlier. The primary variable displayed is the 2m temperature anomaly (t2m), which is visualized using a so-called *uncertainty surface*. A cold/warm color table is used to map the respective temperature values, while the standard deviation (stdev) is employed to perturb the color coding. The inset in Figure 4a shows a close-up of this color perturbation. This technique enables us to show the *ensemble temperature anomaly*. Local mean values can be recognized by the mean color, while the local spread can be identified by a closer look at the color perturbations. Areas with high skill values are additionally marked with a dark color.

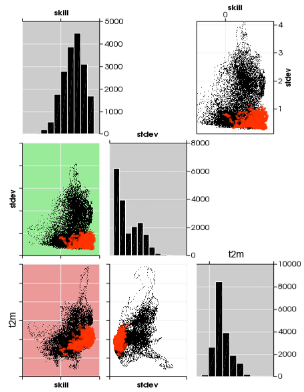
Furthermore, a selection using a parallel coordinates plot has interactively been made, highlighting (in yellow) areas with skill values above 0.5 and standard deviation below 1.0;



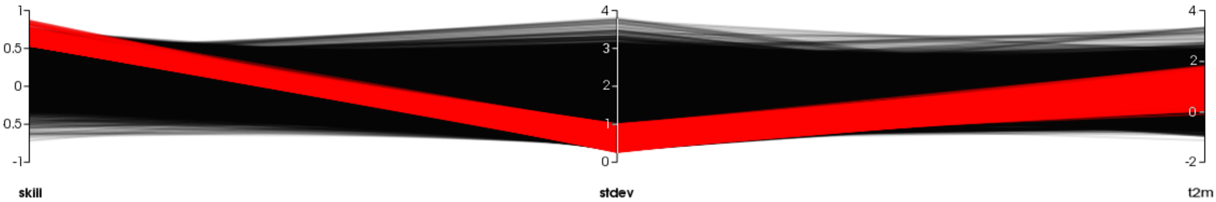
(a) 2m temperature anomaly, skill and stdev at time: 54.



(b) Scatterplot at time: 0.



(c) Scatterplot at time: 108.



(d) Parallel Coordinates Plot at time: 54.

Figure 4: Visual data analysis of decadal climate predictions.

refer to Figure 4d. The values were arbitrarily chosen, however, they represent a good threshold to show data with little uncertainty. The selection itself was made half way through the simulation at time step 54, and is the same in all views. In ParaView, the selection is based on the grid, and remains the same during an animation of time. This allows us to compare the data values at those points at other time steps. For comparison, Figures 4b and 4c show scatterplots and the distribution of this selection at time steps 0 and 108.

Scatterplots are – similar to parallel coordinate plots – a powerful tool to display correlations and dependencies between individual variables. It can clearly be seen that the distribution of points is much more compact towards the end of the simulation. This is visible in the entire data as well as in the data points selected. Where at time step 0 almost all temperature anomaly values are present in the selection, at the end it is only a small range. Also, the shape of the distribution changes and clearly shows a direct correlation between skill and ensemble spread.

5. Discussion

In this work, we have used three different visualization systems and location based techniques to visualize two different fields with uncertainty information together with predicted 2m temperature anomalies. All of the visualization systems used allowed us to visualize the three different 2D fields in a combined figure. However, all of these solutions have their specific strengths and weaknesses.

While static visualizations can already be meaningful for single time steps or temporal means, the temporal development is usually not taken into account. The vector graphics visualization shown in Figure 3 is an ideal quantitative example intended for print media, but due to the stippling technique applied, it is not very well suited for an animated version of the visualization. The animated version of the 3D visualization created with Avizo (see Figure 1) works much better to qualitatively show the spatiotemporal patterns in the data. Without interaction or a movement of the viewing angle, though, the structure of the ensemble spread height field might be hard to perceive. Finally, interactive visual data analysis techniques, such as *linked views* and *brushing*, as shown in Figure 4, enable the user to interactively explore the data and study the relations and dependencies between different variables in much more detail. Combined with ParaView's time animation capabilities, the temporal development of all variables at selected grid points can be studied. However, since various different displays have to be evaluated and understood together, the results achieved with ParaView are complex and not necessarily intuitive for non-experts. On the other hand, these techniques can be very valuable for experts because complex interdependencies in the data can be discovered more easily.

References

[BHJ*14] BONNEAU G.-P., HEGE H.-C., JOHNSON C., OLIVEIRA M., POTTER K., RHEINGANS P., SCHULTZ T.: Overview and state-of-the-art of uncertainty visualization. In *Scientific Visualization*, Hansen C. D., Chen M., Johnson C. R., Kaufman A. E., Hagen H., (Eds.), Mathematics and Visualization. Springer London, 2014, pp. 3–27. 2

[BMW13] BALMASEDA M. A., MOGENSEN K., WEAVER A. T.: Evaluation of the ecmwf ocean reanalysis system oras4. *Quarterly Journal of the Royal Meteorological Society 139*, 674 (2013), 1132–1161. 2

[BOL12] BRODLIE K., OSORIO R. A., LOPES A.: A review of uncertainty in data visualization. In *Expanding the Frontiers of Visual Analytics and Visualization*, Dill J., Earnshaw R., Kasik D., Vince J., Wong P. C., (Eds.). Springer Verlag London, 2012, pp. 81–109. 2

[DUS*11] DEE D. P., UPPALA S. M., SIMMONS A. J., BERRISFORD P., POLI P., KOBAYASHI S., ANDRAE U., BALMASEDA M. A., BALSAMO G., BAUER P., BECHTOLD P., BELJAARS A. C. M., VAN DE BERG L., BIDLOT J., BORMANN N., DELSOL C., DRAGANI R., FUENTES M., GEER A. J., HAIMBERGER L., HEALY S. B., HERSBACH H., HÓLM E. V., ISAKSEN L., KÅLLBERG P., KÖHLER M., MATRICARDI M., MCNALLY A. P., MONGE-SANZ B. M., MORCRETTE J.-J., PARK B.-K., PEUBEY C., DE ROSNAY P., TAVOLATO C., THÉPAUT J.-N., VITART F.: The era-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society 137*, 656 (2011), 553–597. 2

[GKS*13] GODDARD L., KUMAR A., SOLOMON A., SMITH D., BOER G., GONZALEZ P., KHARIN V., MERRYFIELD W., DESER C., MASON S., KIRTMAN B., MSADEK R., SUTTON R., HAWKINS E., FRICKER T., HEGERL G., FERRO C., STEPHENSON D., MEEHL G., STOCKDALE T., BURGMAN R., GREENE A., KUSHNIR Y., NEWMAN M., CARTON J., FUKUMORI I., DELWORTH T.: A verification framework for interannual-to-decadal predictions experiments. *Climate Dynamics 40*, 1-2 (2013). 3

[Kit15] KITWARE: ParaView. Visualization Software, February 2015. 3

[KP13] KIRTMAN B., POWER S.: Near-term climate change: Projections and predictability. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker T., Qin D., Plattner G.-K., Tignor M., Allen S., Boschung J., Nauels A., Xia Y., Bex V., Midgley P., (Eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013. 2

[oCC14] ON CLIMATE CHANGE I. P.: Long-term climate change: Projections, commitments and irreversibility pages 1029 to 1076. In *Climate Change 2013 - The Physical Science Basis*. Cambridge University Press, 2014, pp. 1029–1136. Cambridge Books Online. 2

[OJ14] OBERMAIER H., JOY K.: Future challenges for ensemble visualization. *Computer Graphics and Applications, IEEE 34*, 3 (May 2014), 8–11. 2

[PMK*13] POHLMANN H., MÜLLER W. A., KULKARNI K., KAMESWARRAO M., MATEI D., VAMBORG F. S. E., KADOW C., ILLING S., MAROTZKE J.: Improved forecast skill in the tropics in the new miklip decadal climate predictions. *Geophysical Research Letters 40*, 21 (2013), 5798–5802. 2

[PWB*09] POTTER K., WILSON A., BREMER P.-T., WILLIAMS D., DOUTRIAUX C., PASCUCCI V., JOHNSON C.: Ensemble-vis: A framework for the statistical visualization of ensemble data. In *Data Mining Workshops, 2009. ICDMW '09. IEEE International Conference on* (Dec 2009), pp. 233–240. 2

[SGE*13] STEVENS B., GIORGETTA M., ESCH M., MAURITSEN T., CRUEGER T., RAST S., SALZMANN M., SCHMIDT H., BADER J., BLOCK K., BROKOPF R., FAST I., KINNE S., KORNBLUEH L., LOHMANN U., PINCUS R., REICHLER T., ROECKNER E.: Atmospheric component of the mpi-m earth system model: Echam6. *Journal of Advances in Modeling Earth Systems 5*, 2 (2013), 146–172. 2

[SSB*13] SMITH D. M., SCAIFE A. A., BOER G. J., CAIAN M., DOBLAS-REYES F. J., GUEMAS V., HAWKINS E., HAZELEGER W., HERMANSON L., HO C., ISHII M., KHARIN V., KIMOTO M., KIRTMAN B., LEAN J., MATEI D., MERRYFIELD W., MÜLLER W. A., POHLMANN H., ROSATI A., WOUTERS B., WYSER K.: Real-time multi-model decadal climate predictions. *Climate Dynamics 41*, 11-12 (2013), 2875–2888. 2

[SZD*10] SANYAL J., ZHANG S., DYER J., MERCER A., AMBURN P., MOORHEAD R.: Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *Visualization and Computer Graphics, IEEE Transactions on 16*, 6 (Nov 2010), 1421–1430. 2

[TSM12] TAYLOR K. E., STOUFFER R. J., MEEHL G.: An Overview of CMIP5 and the Experiment Design. *Bull. Amer. Meteor. Soc. 93* (2012), 485–498. 2

[UCA14] UCAR/NCAR/CISL/VETS, BOULDER, COLORADO: The NCAR Command Language (Version 6.2.1). Visualization Software, 2014. 3

[UKS*05] UPPALA S. M., KALLBERG P. W., SIMMONS A. J., ANDRAE U., BECHTOLD V. D. C., FIORINO M., GIBSON J. K., HASELER J., HERNANDEZ A., KELLY G. A., LI X., ONOGI K., SAARINEN S., SOKKA N., ALLAN R. P., ANDERSSON E., ARPE K., BALMASEDA M. A., BELJAARS A. C. M., BERG L. V. D., BIDLOT J., BORMANN N., CAIRES S., CHEVALLIER F., DETHOF A., DRAGOSAVAC M., FISHER M., FUENTES M., HAGEMANN S., HÓLM E., HOSKINS B. J., ISAKSEN L., JANSSEN P. A. E. M., JENNE R., MCNALLY A. P., MAHFOUF J.-F., MORCRETTE J.-J., RAYNER N. A., SAUNDERS R. W., SIMON P., STERL A., TRENBERTH K. E., UNTCH A., VASILJEVIC D., VITERBO P., WOOLLEN J.: The era-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society 131*, 612 (2005), 2961–3012. 2

[Wil11] WILKS D. S.: *Statistical methods in atmospheric sciences. Third Edition. International Geophysics Series, Vol. 100.* Elsevier, 2011. 3