Clim. Past Discuss., 3, S617–S625, 2007
www.clim-past-discuss.net/3/S617/2007/

Climate of the Past Discussions

Interactive comment on "How unusual was autumn 2006 in Europe?" *by* G. J. van Oldenborgh

G. J. van Oldenborgh

Received and published: 5 October 2007

I would like to thank the reviewer for the statistical input. As is clear from the manuscript I do not have a strong background in statistics, and these comments help to improve the paper.

In my opinion the goal of the manuscript is valuable and will contribute to increase our understanding of extreme events. I also find, however, that the manuscript can be technically improved; too frequently the formulations are imprecise, and in an analysis that is essentially statistical in nature, this raises some concerns. Perhaps more important, the manuscript virtually obviates uncertainties, and I think this is an aspect that should be considered with much more care. I try to explain my suggestions in the following.

I have emphasised the uncertainties more in the text.

To estimate return periods, the data prior to 2006 are assumed to be drawn from a certain distribution - this would be the null-hypothesis against which autumn 2006 is *confronted. However, this null hypothesis is formulated in a very loose manner. To assume that the data only show 'interannual variability' is not a clear description of this null-hypothesis. No spectrum can only show a peak at frequencies of 1 year and be zero otherwise. The author means probably - although it is not completely clear to me that the data are assumed to be Gaussian white noise, i.e. drawn from a Gaussian distribution and independent of one another. This process shows equal variability at all timescales, not only interannual, and the population autocorrelation function is zero.*

The gaussian hypothesis is hidden in the text, and it should be stated more clearly since it may be central to the results obtained. Were the distribution not gaussian, perhaps skewed to 'the right', extremes would be more probable, and the autumn 2006 might not such be clear outlier. A test for normality for the data prior to 2006 seems therefore necessary. The hypothesis of independence may also have some relevance, although perhaps a smaller one. As far as I understood, the data prior to 2006 are used to estimate a standard deviation, which is then used to estimate the return periods for the 2006 event assuming that the distribution is gaussian. However, if the data are not independent, the estimation of the standard deviation from the 'usual' estimator is not correct: this would yield an underestimation of the true standard deviation (as the data are not independent, the risk of not sampling the full spread is larger). There exist some estimators of the variance from autocorrelated data, and I think that a test of the possible effect of this autocorrelation will be also helpful.

In the extrapolation without taking climate change into account, I have clarified the 'only interannual variability' as suggested by the reviewer. Correcting for the autocorrelations does not add much to the argument, as it is known that there is a strong rise at the end of the record due to global warming, and this is addressed in the next section on physical grounds.

I also added some comments on the normality of the data. In Fig. 3 it can be seen that in De Bilt, the normal approximation overestimates the tail of the distribution. A comparison of a map of return times computed with the GPD extrapolation with the return times based on a normal assumption shown in Fig. 5 shows that the normal assumption does not lead to a systematic overestimation of the return times. This agrees with the observation that the autumn temperatures are negatively skewed in most grid points. (In contrast, summer and winter temperatures are very non-Gaussian in Europe.)

When trying to discount the effects of 'global warming' (equation 1) , some additional problems may be lurking. First, no estimation of the uncertainties is given. If instead of the best estimate of the regression coefficient, one would take the upper limit of the 95% confidence interval, the return periods for the 2006 event would certainly diminish. Both probabilities — that of a higher regression coefficient and that of higher 2006 temperatures under the null-hypothesis — are coupled. But how large can the influence uncertainty in the regression be? Compounding this problem is the fact that the regression in eq. 1 is performed with autocorrelated data, which again increases the uncertainty bounds, and by the fact that the residuals of the regression will be also probably autocorrelated, with the problem mentioned before of the estimation of their standard deviation. Furthermore, the estimation of a regression coefficient when the residuals are autocorrelated is problematic — a missspecified statistical model — and here a test a standard test for autocorrelated residuals (for instance Durbin-Watson) could be of help. Perhaps a more serious problem here is that the standard deviation of the residuals in eq. 1 is estimated from the calibration period, and this is again an underestimation of the true standard deviation of the residuals. A better estimation could be achieved by estimating the regression in a calibration period and estimating the residuals in a validation period.

The normality of the time series after subtracting a linear regression with global warming has been checked. I assumed that this would be clear from the plot, and from the comparison with the GPD extrapolation, but it is now also noted explicitly in the text that the skewness is slightly negative.

The autocorrelation of the 'weather' residuals of Eq (1) is slightly negative for 1-4 years.

This shows that the large majority of the autocorrelations that marred the original extrapolation are due to global warming, and the remaining 'weather noise' can be treated as white noise. This is also in agreement with many publications that show that over land decadal variability is small.

The reviewer noted an error in figure 7 in that I neglected the errors on the regression against global mean temperature when assessing the confidence intervals. This figure has therefore been re-computed using a simultaneous estimate of the regression against T'_{global} and the return time (using a bootstrap). This lowers the lower bound of the 95% CI to 125 years, mainly due to the possibility of a larger trend which lowers the anomaly above the curve.

Figure 8 only show the central values, which are not affected by this error. The Gaussian approximation gives an underestimation of the return period, as the residuals of Eq. (1) are negatively skewed over most of the region. I computed the lower boundaries as in Fig. 7 and added the main results in the text.

A good-of-fit analysis would be also necessary, i.e. a test of whether the relationship between global and local temperatures is linear. Indication of a non-linear fit might also increase the probability that the residual in 2006 falls, after all, within the pre-2006 probability distribution.

In regard of the possibility of non-linear effects of global warming, this is the point I want to make with the paper. In the current climate models these nonlinearities are small in autumn, but the observed temperatures in 2006 suggest the possibility that the effects are more nonlinear in nature. The number of data points is too small to do a more formal statistical analysis of these nonlinearities, I think.

In summary, there exists a variety of technical issues that have not been considered, and which almost without exception could contribute to increase the probability of occurrence of the the 2006 event under the null-hypothesis. Perhaps, after all, this event remains quite improbable, but these issues should be quantified.

CPD

3, S617–S625, 2007

Interactive Comment

I am aware that all these points are interconnected and that a 'clean' estimation is probably quite convoluted. I would suggest to explore some type of Montecarlo approach in which these aspects are all incorporated simultaneously. For instance, the author may want to check whether a bootstrap estimation of the probability distribution of the data prior to 2006 yields very different results - and wide confidence intervals - as by simply estimating the standard deviation of a white noise processes. For the results derived from the regression 1, perhaps a bootstrap approach for the estimation of the regression coefficient in a calibration period, following for each bootstrap sample by estimation of the standard deviation of the residuals in an independent period, could perhaps be more correct than the solution presented in the manuscript. Perhaps some of these methods have already been use in the manuscript - see caption figure 3- but there have not been clearly explained. These are just some suggestions, and other approaches may be feasible and more accurate. A supplementary material covering these technical issues should be very helpful for the reader.

I have tried to address the major concerns of the reviewer, including the wrong error estimates of the return times of the weather noise (Fig. 7), using a bootstrap. Unfortunately time and knowledge constraints mean I can not rewrite the article in a completely rigourous statistical way, with well-defined null-hypotheses and tests.

The errors on the standard deviation of $\mathcal{O}(0.1\mathrm{K})$ are smaller than other problems with the time series, such as inhomogeneities in observation techniques, the effects of which are clearly visible in Fig. 6 as isolated spots with trends that are very different from surrounding areas.

When the model gets more complex to include some regional physical processes (eq. 2-4) , the statistical problems may increase. Certainly, the predictors are autocorrelated - this widens the again the confidence interval of the regression coefficients. The underestimation of the residual variance in the calibration period becomes now a more serious issue. The variance of the residuals in calibration and validation will probably now differ more strongly, since more predictors allow for a better artificial fit in the

CPD

3, S617–S625, 2007

calibration. To illustrate this potential problem, one may consider the case with 50 predictors: the residual variance will be vanishing small, and the 2006 event will appear as very unlikely, but this would be just an artifact. The physical basis of the regression model 2-4 may be debated, and other authors may exclude some of them and include others, although I in my view the choice of predictors seems reasonable. However, the skill of the model must be estimated in a period independent of the calibration. Only in this way one can judge the real validity of the predictors, and the residual variance be estimated.

I do not do any extreme value statistics on the residues of the Very Simple Model Eqs (2)-(4). This model is a rough first-order approximation of the influence of circulation on temperature and neglects many nonlinear effects. It is only used to quantify (in this first-order approximation) the effects of circulation on the observed temperature extremes. I have tried to make this more clear.

The main problem with the model is the correlation between the global warming term and the memory term, which is also subject to global warming. All other cross-correlations are lower than 0.5 (at De Bilt).

If I would have used it as a forecasting model I would have shown a cross-validated skill. However, it is just a first-order description of well-known relationships. The difference between Fig. 10 and a cross-validated one is small over the 58 years, as the number of parameters is still relatively small.

I have some mixed feelings about section 5 - application to climate simulations. I think it can be very valuable to explore whether climate simulations show a change in the probability distribution of extreme local temperatures that is not related to simple changes in the mean. This is indeed shown in Figure 14, but I am not so sure that the preceding paragraphs are really necessary. Is the information about the the estimation of the trends in global, or even regional, temperature by the different models relevant here? I would delete these paragraphs and, on the other hand, expand the more *interesting last paragraph in section 5 and concentrate in the changes of extreme value statistics over time within the simulations, if any. This would amount to explain more clearly Figure 14. Some of the issues raised for the observations reappear here - for instance the problem of a 'gaussian fit' for the estimation of return periods.*

On the model section, I think the paragraphs on the trends are very necessary, as the trend and extremes are interdependent. As the improved error analysis of Eq. (1) again shows, the observed temperatures can either be due to strong shift of the whole PDF, or a nonlinear effect which causes a an increase of the probability of positive extremes. The climate models do not show either of these effects, which should both be discussed.

The agreement between the data points and the Gaussian approximation in Fig. 14 shows that it is a good approximation. A skewed distribution would have shown disagreement in the tail, as the negative skewness in Fig. 7 shows up by the fit being above most points in the tail of the PDF. I have added a remark to the discussion about this.

Some particular points:

Abstract: 'current circulation' means current mean atmospheric circulation?

Yes, added.

Page 813, 'pre-instrumental reconstruction indicate' $\rightarrow$ 'pre-instrumental reconstructions indicate'.

OK.

Page 814, define the acronyms GHCN/CAMS

Google. . . OK.

Page 815, Cold Ocean/Warm Land pattern. Another factor that may contribute to enhance European temperature changes relative to the global mean is the larger warming *in the Northern Hemisphere*

Yes, but this can be explained by the greater proportion of land in the northern hemisphere, assuming land warms faster than ocean.

Page 815, 'the lower bound of the 95% CI' . How has this been estimated?

With a non-parametric bootstrap, added to the text.

Page 815, 'even in a linearly warming climate'. This is inaccurate, I think the author means 'in a local climate linearly related to global temperature'

'. . . even taking into account a shift in the probability density function proportional to global warming.'

Page 816, 'd'?M is a memory term of past circulation' I think it is a memory term of past local temperature.

Yes.

Page 816 'A_s, A_w, B are fixed by the interannual variability'. The author means probably that these parameters are most strongly influenced by the interannual variability.

Yes.

*Page 817, âd'?in Fig. 11 the *estimated* contribution of the various terms *for the 2006* event*

Yes.

Figure captions. Include units in all figures.

OK.

Caption 2 September-November 2006

This figure has been deleted.

CPD

3, S617–S625, 2007

Interactive Comment

Caption 8, âď?As in Figure 3âď?, but Figure 3 has a very different appearance. Please expand the caption fully.

This was an incorrect cross-reference, I meant Fig. 5.

Caption 13. As suggested above, I would delete these paragraphs and corresponding figures, but in case they are retained, please mention the method that was used to estimate the linear trends.

OK.

Interactive comment on Clim. Past Discuss., 3, 811, 2007.