Department for Environment Food & Rural Affairs
Department for Business, Energy & Industrial Strategy
Met Office Hadley Centre
Environment Agency

UKCP18 Technical Note: Clipping and Baseline advice on Land Strand 1 data in UKCP18

Headline Summary

This note provides technical detail on specific aspects of the methods used in the probabilistic projections. This may affect how you use and interpret the data. In short:

- Pre-prepared probability distribution functions (pdfs) and cumulative distribution functions (cdfs) are available as data and graphical plots, for each of five future emissions scenarios. In these, the data are clipped at 1% and 99% probability levels, i.e. climate changes have been capped at the extreme top and bottom ends of the distribution. This was also done in UKCP09.
- Underlying the pdfs and cdfs is a sample of 3000 realisations of annual anomalies from 1961-2000. These are clipped in each individual year. You may wish to construct your own pdfs or cdfs from the realisations. When doing this, we recommend pooling data from a period of consecutive years, in order to reduce sampling uncertainties. Since the values at which clipping occurs are different from year to year, such pooled distributions may show slightly broader regions of increased probability density near their extremes, in comparison to the steps visible in the pre-prepared distributions discussed above.
- In general, relative probabilities for outcomes below the 5% probability level, or above the 95% probability level, should not be regarded as reliable. For such outcomes, it is better to consider the cumulative probability of a lower or higher outcome.
- You will find that the sum of annual anomalies during the baseline period (1981-2000) is close to but not precisely zero. The effect is very small compared to the range of anomalies, but you may wish to consider adjusting the data.

Version 1.0 | March 2019

1. Introduction

The probabilistic projections from Land Strand 1 consist of a sample of 3000 realisations of annual anomalies for five alternative future emissions scenarios (RCP2.6, 4.5, 6.0 and 8.5, and SRES A1B), expressed relative to a baseline period of 1981-2000. Data are provided for a set of core climate variables from 1961-2100, on a 25km version of the OSGB British National Grid and for three sets of aggregated regions. See the guidance on "Data availability, access and formats", available from the UKCP18 website.

The production of the Land Strand 1 data is described in the Land Projections report (see UKCP18 website). The methodology involves extensive statistical processing, in order to represent uncertainties by combining results from several ensembles of climate model simulations.

The results can be used to estimate probability levels for future climate anomalies that represent climate modelling uncertainties, the effects of internal climate variability, and errors associated with statistical emulation and scaling techniques used in the calculations. The results are conditional upon the modelling inputs and statistical choices used in the methodology.

This short document provides guidance on two issues, clipping and baselines.

2. Clipping advice

In general, we expect confidence in the projected probability levels to be higher in the bulk of the distribution (typically between the 5 and 95% probability levels) than in the tails. For example, extreme outcomes can often be more sensitive to limitations in the statistical assumptions made, such as how to represent residual uncertainties in regression relationships. In each year, we therefore clip the 3000 realisations at the 1% and 99% probability levels, by resetting lower values to that of the 1% level and higher values to the 99% level (also referred to as 'winsorization'). This avoids provision of extreme values that may be sensitive to methodological assumptions and therefore less credible. Similar clipping was performed in the UCKP09 projections.

The annual Strand 1 data is available as probability density functions (pdfs) and cumulative distribution functions (cdfs), from the User Interface (UI) and the CEDA data catalogue. Graphical plots are also available from the UI. These are calculated from the sample realisations, employing two techniques to reduce the effects of sampling noise. Firstly, raw annual data is pooled over a larger time window centred on the year in question to give a more representative distribution. Here we use an 11-year window, so the pdfs and cdfs for the year 2090, for example, are estimated from pooled annual sample realisations from the years 2085 to 2095. Secondly, we apply kernel density estimation (KDE) techniques to reduce the effects of sampling noise. The UKCP18 probability distributions are fitted to the sample data using the python scipy kernel density estimation package 'gaussian_kde' documented in https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.gaussian_kde.html. The standard normal density function is assumed for the kernel, and the bandwidth is specified using Scott's Rule, which for univariate data is specified by $sn^{-1/5}$, where *s* is the sample deviation of the sample data and *n* the sample size (D.W. Scott, Multivariate Density Estimation: Theory, Practice, and Visualization, Second Edition, John Wiley & Sons, 2015, DOI:10.1002/9781118575574).

In the pdf data, the probability densities below the 1% level and above the 99% level are set to the 1% and 99% values respectively, calculated prior to clipping. Similarly for the cdf data, all anomaly values below the 1% probability level are set equal to the 1% value prior to clipping, and all anomaly values above the 99% probability level are set to the 99% value. The cdf and pdf data and graphical plots both include this clipping. As an example, the first panel of Figure 1 shows a typical precipitation pdf estimated using KDE on unclipped sample realisations, and then clipping the distribution at the 1% and 99% levels. The second panel shows the corresponding lower tail of the cdf. All values in the cdf for percentiles less than or equal to 1% are set here equal to -51.7%, the precipitation anomaly for the 1% probability level.

You may also wish to calculate and plot your own cdfs and pdfs from the samples of 3000 realisations, possibly for multi-decadal averages of annual data, or for pooled annual data covering a longer period, such as 2050-2069 or 2080-2099. It is not recommended however that you attempt to recalculate pdfs for single years, since the resulting pdfs will be impacted by sampling noise, and due to the clipping, unphysical spikes in probability density could be obtained at the 1% and 99% probability levels. For the graphical annual distributions available from the UI we use centred, pooled 11-year periods to estimate the pdf for a specified year, and you are likewise advised to use similar (or larger) pooling windows when calculating your own distributions from the sample data. Even pooling of clipped data over a decadal (or multi-decadal) window will potentially still show slightly broader regions of increased probability density near the extreme probability levels of the pooled distribution. This is because the clipping is performed separately for each individual year, and the relevant threshold values will vary from year to year.

An example is shown in Figure 2 for the winter precipitation response to RCP8.5 for the years 2085-2095 in Wales. The blue curve corresponds to the pdf obtained by applying the KDE technique described above to clipped sample realisations. The orange curve was obtained with an alternative, less smooth approach, estimating the cdf for 50 equally sized bins and centre-differencing to estimate the pdf. In both cases, a slight increase in relative probability near the anomaly value of +100% is obtained (a similar but reduced effect is found for the other tail, near -50%).

Such features should not be interpreted as a physically realistic increase in the relative chance of the relevant extreme outcomes, compared to those for nearby probability levels. We advise in general that relative probabilities for specific outcomes below the 5% level or above the 95% level should be regarded as being of reduced confidence. For outcomes in the tails of the distributions, it is better to consider the cumulative probability of outcomes below or above specific extreme probability levels (e.g. below the 5% or 10% levels, or above the 90% or 95% levels). For the example in Figure 2, the 5% and 95% levels are -30% and +84% respectively. It can be noted also that pdfs or cdfs plotted from the realisation data may look slightly less smooth than the pre-prepared UI graphics, unless you choose to apply similar smoothing techniques to the data as described above.

In the case of precipitation, there is a second contribution to clipping (see Appendix C of the Land Projections Report). This arises because precipitation is a bounded variable, which cannot be negative. Anomalies are expressed in the calculations as percentages of the baseline value. In cases where the projected change in the long-term average climate shows a strong drying, the statistical procedures used to represent inter-annual variability about the long-term average can occasionally generate unphysical dry anomalies exceeding -100%. These instances are reset to -100%, before the standard clipping at the 1% probability level is applied. This bounding issue mainly arises for scenarios, regions and periods characterised by high warming and large reductions in average precipitation, such as in south-eastern parts of England during summer, during the latter decades of the 21st century under the RCP8.5 emissions scenario. In such extreme cases, the frequency of occurrence of this second type of clipping can amount to 5-10% of individual values in the realisations. Where this occurs, the probability density for affected probability levels is set to the probability density found at the lowest probability level unaffected by the application of the -100% bound. For cdf data, the precipitation anomalies for affected probability levels are set to -100%. As an example, Figure 3 shows the pdf for the summer precipitation response in Wales for the individual year 2090, under the RCP8.5 scenario. We find for this variable that three percent of the distribution is clipped at -100%. You should also look out for this effect in self-prepared graphics derived from the clipped realisation data.

3. Baseline advice

The statistical methods of Land Strand 1 (see Land Projections Report) include use of a temporal filter. This is applied to anomalised data during the calculations of time-dependent changes derived from the global climate model simulations included in Land Strand 1. This is done to provide an approximate partitioning between signals of long-term climate change on time scales of ~20 years, and internal climate variability on shorter time scales. The chosen filtering method (a Butterworth filter with a cut-off of 20 years) allows data from outside any given 20 year period to influence low-pass signals within that period. This means that time series of low-pass filtered anomalies may not add up to precisely zero during the baseline period in question. In particular, this means that in a specific realisation, average anomalies during 1981-2000 will not usually add up to precisely zero. Those realisations subject to clipping will see an additional contribution to small changes in their average anomalies, providing a second driver for the occurrence of non-zero average anomalies during 1981-2000.

These effects are small. For example, in the case of surface temperature in administrative regions, the offset in the average anomaly in a typical realisation during 1981-2000 is about 0.07°C, while the mean offset across all 3000 realisations is typically ~0.02°C. These numbers can be compared with ranges of anomalies (of typically several degrees) projected during the 21st century. Nevertheless, some users may prefer to apply empirical adjustments to the downloaded data, in order to re-base the realisations to remove these small offsets.

Figures

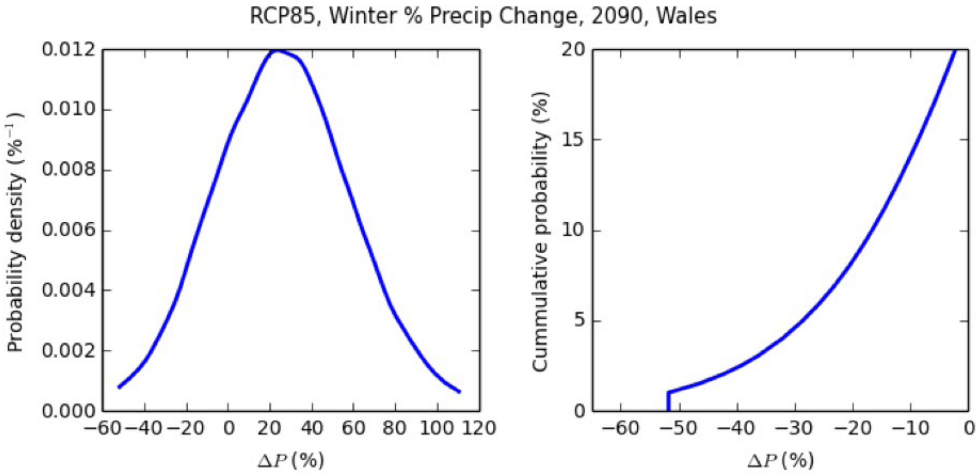

Figure 1. Left panel: The UKCP18 pdf for winter precipitation response for RCP8.5 forcing for Wales, based on pooled sample realisations for the 11-year window 2085-2095. Clipping has been applied at the 1% and 99% probability levels. Right Panel: The corresponding cdf for the data in the left panel, focussing on the lower tail only. Clipping at the 1% probability level is applied.

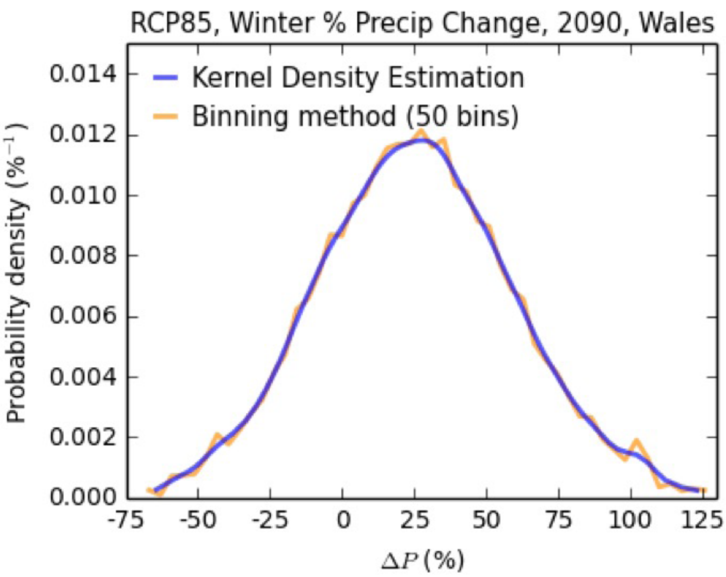

Figure 2. Pdfs for winter precipitation response for RCP8.5 forcing for Wales, based on a pooled sample of clipped realisations, for the 11-year window 2085-2095. Here the input sample data has been clipped at the 1% and 99% probability levels prior to the calculation of the pdfs, whereas in Figure 1 the pdfs are initially calculated from unclipped sample data, and clipping is applied subsequently. The blue curve corresponds to estimations using the kernel density estimation technique described in the text. The orange curve was obtained with an alternative, less smooth approach, estimating the cdf for 50 equally sized bins, and centre-differencing to estimate the pdf.

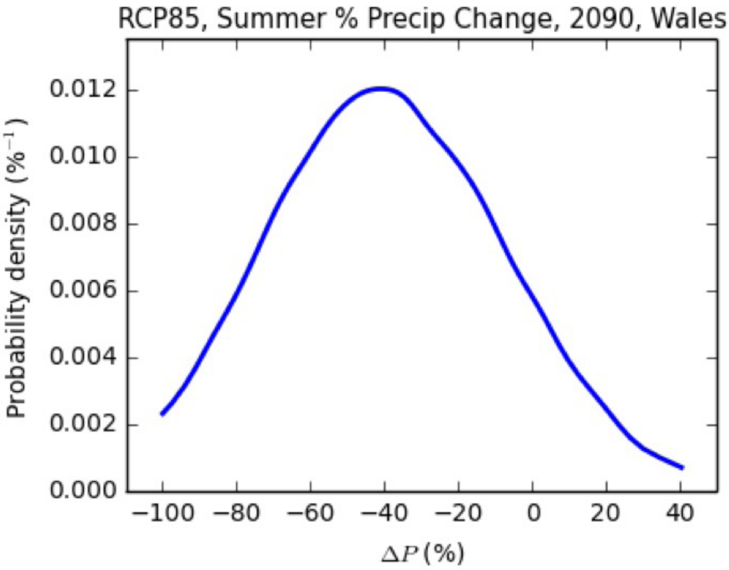

Figure 3. The UKCP18 pdf for summer precipitation response for RCP8.5 forcing for Wales, based on pooled sample realisations for the 11-year window 2085-2095. The three percent of the realisations projected to be less than -100% are clipped at this value. Clipping has also been applied at the 99% probability level.