

Assessment of Daily Rainfall Variability in Climate Model Simulations Using Estimations of Areal Rainfall **Carol McSweeney (CRU)**



Aims and Objectives

Variability in daily precipitation simulated by coarse scale climate models can be difficult to validate against station records because the level of variability in spatially averaged rainfall is not comparable to that of point rain-gauge observations.

Whilst for some regions, this can be overcome by using a large number of stations to give an areal average, for many regions the density or distribution of observations may not be sufficient to give a 'true' areal mean, and the result may mean that our baseline climate is biased.

The objective of this work is to estimate some of the properties of 'true' areal means, using only a small sample of stations. These estimates of areal precipitation variability can then be used to validate climate model daily precipitation simulations.

Tim Osborn (CRU) Phil Jones (CRU)

Nick Reynard (CEH Wallingford) Kevin Hiscock (UEA)



1. Methodology: Estimating Areal Precipitation for a Grid-box

(a) Dry-day probability (*P(dry)*)

Box (i) 'Effective n' (b) Gamma Parameters of Wet-day values (α and β)

Objective:

to estimate the dry-day probability in an *n*-station mean where *n* is large enough to represent a 'true areal mean' (...and where a wet day is any value of 0.3 mm or more...)

Approach:

For a set of *n* stations the concept of 'effective n' (box i) allows those stations to be treated as independent stations, for which simple laws of probability can be used to calculate *P(dry)* in the n-station mean...

$P(dry)_n = (P(dry)_1)^n$

When correlation co-efficient *r* is replaced with a measure of correlation between wet and dry day coincidences (*r(wet/dry*) (box ii), this provides a good estimate of the dryday probability in the *n*-station mean (Fig 1).

Figure 1 demonstrates that this approach is effective when applied to randomly selected clusters of *n* stations from 3 station datasets from the UK, China and Zimbabwe



For *n* stations with a mean interstation correlation r, we can describe using 'effective n' (n') their equivalent number of independent stations (*r*=0)...



Box (ii) Wet and dry day 'correlation': 'r(wet/dry)'

For this application we are less interested in the correlation between rainfall **amounts (r)**, and more interested in the 'correlation' between wet and dry day incidences...

The level of dependence between 2 stations can be quantified on a 0-1 scale by comparing the actual number of coincident dry-days with the maximum number (when 2 stations are totally dependent $P(dry)_2 = P(dry)_1$ and the minimum number (when 2 stations are independent $P(dry)_2 = (P(dry)_1)^2)$.

 $r(wet/dry) = \frac{(P(dry)_2(Obs) - (P(dry)_1)^2)}{(P(dry)_1)^2}$ $(P(dry)_1 - (P(dry)_1)^2)$

Extension to large values of N

To apply this to the case of **N** stations from a grid box (e.g. 1000), instead of *n* stations, we require:

(a) the average dry-day probability at a station in the grid box

calculated either by using a simple average of the *P(dry)* at all available stations (assumes distribution is even and therefore representative), more complex interpolation approaches such as

Objective:

to estimate the gamma parameters (α and β) of wet day values in an *n*-station mean where *n* is large enough to represent a 'true areal mean'...

X UK

X China

X Zimbabwe



Approach:

As stations are averaged to give an areal mean, the distribution of wet days becomes more normal: shape parameter $\boldsymbol{\alpha}$ increases and scale parameter $\boldsymbol{\beta}$ decreases, as the distribution becomes less skewed towards low values and the range of values narrows. The degree of change in the distribution with spatial averaging is dependent on (a) *n*: the number of stations (b): the degree of dependence between those stations.





Figures 2 and 3 show these relationships for the shape and scale parameters, using randomly selected clusters of *n* stations from the UK, Zimbabwe and China. For the scale parameter, a strong relationship is evident, but the shape parameter shows a less clear relationship.

However, by definition, the product of the shape and scale parameters should equal the mean of the distribution (the mean wet-day amount). This means that if *β* can be estimated, *a* can be inferred from the parameters that are known. Figure 4 shows that the shape parameter can be estimated this way with reasonable accuracy.

0.0 0.2 0.4 0.6 0.8 1.0 Estimated P(dry)n

Figure 1: Estimated n-station mean P(dry) vs. Actual P(dry) in randomly selected clusters of *n* stations from UK, Zimbabwe and China datasets.

Theissen's polygons, or information from existing climatologies. (b) the grid box mean inter-station r(wet/dry)

• Calculated using correlation decay curves in order to give expected correlations for a representative range of separation distances in the grid box, and avoid biasing if available stations are clustered.

This methodology can also be extended to values of **N** which are large than the *n* stations available in the same way as has *P(dry)*, using estimates of average station gamma parameters and grid box mean *rwet* estimated from correlation decay curves.

 $\alpha_{n}(\text{Estimated})$

Figure 4: Estimated n-station α vs. Actual n-station α in randomly selected clusters of *n* stations from UK, Zimbabwe and China datasets. (NB excludes cases where the mean daily amount is less than 0.3mm.

2. Application: Using Areal Precipitation Estimates for GCM Validation



Carol McSweeney c.mcsweeney@uea.ac.uk