Geospatial Health
Health Applications in Geospatial Science

Università degli Studi di Napoli Federico II
Dipartimento di Medicina Veterinaria e Produzioni Animali

Appendix

To what extent does climate explain variations in reported malaria cases in early 20th century Uganda?

Adrian M. Tompkins, Laragh Larsen, Nicky McCreesh, David M. Taylor

Geospatial Health 2016; 11(s1):407
doi: 10.4081/gh.2016.407

Raining data filling

The daily rainfall data are complete, with the exception of Masindi station, where data are missing for April-December 1932 and December 1934. For these limited periods the rainfall was simply taken for the equivalent day of the year before, and the results of the simulations for Masindi should be disregarded for these and the subsequent two months.

Temperature data filling

The problem of missing temperature data is particularly severe during the mid to late 1930s and the early 1940s, where for some stations continuous periods of several years of data are missing. Fortuitously, during every single month during the period of interest, data values from a minimum of 5 stations are always available, permitting both temporal and spatial correction methods to be employed. We therefore examined a range of correction methods for filling the missing data at the 5 stations.

Climatology: the temperature was filled with the station's monthly mean value for all available data, considered a benchmark methodology.

Persistence: the standardised temperature anomaly was used for the closest available month at the same station preceding or succeeding the missing data point.

Mean or weighted spatial correction: the standardised anomalies were used for all other stations for which data are available for the specific month. The mean correction simply averages all available stations values across Uganda, while the latter method applies a distance-weighting to emphasise closer stations, using an e-folding weighting decay scale of 200 km.

Nearest station correction: this used the standardised anomaly from the closest station with data available for the month in question.

Persistence-mean/weighted: this correction combined the persistence method with the mean or weighted correction, using a weighting factor based on the time in months to nearest available data at the station in question and an e-folding decay of 3 months. Thus if data are available at the station the month before or after, the persistence value was used, but as the time gap increases the corrected value tends towards the spatial weighted value.

Standardised anomalies were converted into corrected values using the mean and standard deviation for each month and station. The correction was calculated on an extended period of 1901-1960 in order to ensure that the statistics were calculated using the longest dataset possible. We tested the methods using a cross validation Monte Carlo technique, whereby for each of the stations in turn, 24 consecutive available monthly values were set to missing, starting from Jan 1901 in the first experiment, February 1901 in the second, and so on until the final experiment removed 1959-60. If some of the data are already missing within the 24-month period, the algorithm extended the window to ensure that exactly 24 months were removed. This created a total of 3580 artificial datasets (5 stations by 696 start months), which were identical apart from the 24 additional missing values. Each correction algorithm was then applied to all 3580 datasets, and statistics were compiled regarding their performance in correcting the 24 missing values. Note that the removed data were chosen at consecutive rather than random months in order not to favour the persistence method, keeping in mind that the actual missing data points nearly always occur in coherent blocks rather than single missing months.

The results of the correction technique were studied in terms of mean bias, root mean square (RMS) error and the mean absolute error (MAE) (Appendix Table 1). The mean bias of the correction methods was for the most part negligibly small, with the largest magnitude occurring with the persistence methods, which still only amounted to a 0.15 K. Thus all methods were deemed to be acceptable in terms of the bias for malaria modelling purposes. The relative performance of the methods was identical for both error metrics (mean absolute and root mean square) and showed that using the nearest station or persistence method is actually worse than the benchmark use of the climatological value. In contrast, both methods that used an average of all available stations reduced the error by approximately 0.05%. The best results are obtained when the spatial methods were combined with the persistence methodology, taking advantage of the additional information available at the start and end of a missing data period, reducing the RMS by 0.1 K (approximately 11%). There was a small additional advantage of applying the distance weighting but it was not statistically significant and was sensitive to the choice of the weighting decay. Thus for this study we applied the simpler mixed persistence-mean spatial method.

Reconstructing daily temperature data

There have been complex stochastic methods available for converting monthly climatological data to daily values using so-called weather generators (Srikanthan and McMahon, 1999; Schuol and Abbaspour, 2007). While this is particularly challenging for the highly non-Gaussian precipitation field, temperature variability is less and as larvae and parasite development times are multiple days, we argue that the correct daily autocorrelation of temperature is not crucial as long as the temporal temperature variance is represented, which is important for nonlinear processes Paaijmans *et al*. (2009). We therefore employed a simple method whereby the monthly averages were first linearly interpolated and then Gaussian white noise was added with a standard deviation equal to that observed in the recent daily station record (1996-2011). This latter step assumes that the day to day temperature variability has not changed over the past century, and thus the experiments were repeated with a 30% increase/decrease in the temperature variability, which was not found to impact the key conclusions drawn in the modelling study (not shown).

Appendix Table 1. Mean bias, mean absolute error and the root mean square error of seven different correction methods for missing monthly temperature data (see text for methodology details).

Correction method	Bias (K)	MAE (K)	RMS (K)
Climatology	0.042	0.62	0.90
Mean spatial	0.033	0.58	0.86
Weighted spatial	0.036	0.57	0.85
Nearest station	0.0064	0.72	1.07
Persistence	-0.15	0.72	0.98
Persistence mean	-0.0044	0.56	0.80
Persistence weighted	-0.0015	0.55	0.80

MAE, mean absolute error; RMS, root mean square error.

References

Paaijmans KP, Read AF, Thomas MB, 2009. Understanding the link between malaria risk and climate. Proc Natl Acad Sci USA 106:844-9.

Srikanthan R, McMahon T, 1999. Stochastic generation of annual, monthly and daily climate data: a review. Hydrol Earth Syst Sc 5:653-70.

Schuol J, Abbaspour K, 2007. Using monthly weather statistics to generate daily data in a SWAT model application to West Africa. Ecol Model 201:301-11.