Geospatial Health
Health Applications in Geospatial Science

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

pagepress

Appendix

Geospatial Health #796

Space and time predictions of schistosomiasis snail host population dynamics across hydrologic regimes in Burkina Faso

1. Remotely sensed estimates of environmental covariates

Remotely sensed data were used for the regionalization of predicted population dynamics of snail intermediate hosts throughout the country. The choice of covariates for inclusion in the spatial models of snail ecology considered both model identification results from Perez-Saez *et al*. (2016) as well as the availability of relevant remotely sensed products providing estimates at suitable spatial and temporal resolutions. Among the 6 environmental covariates investigated in 2016 (air temperature, water temperature, conductivity, level, precipitation amounts and number of intense rainfall events), only precipitation and air temperature were retained due to their systematic appearance in the model identification process and the availability of the corresponding remotely sensed products. For both covariates, available remotely sensed products were selected among possible candidates, corrected against ground measurements and their uncertainty characterized by fitting the appropriate error models. Following common approaches in hydrology for rainfall-runoff models (Hong *et al*., 2006; Moradkhani *et al*., 2009), the latter step is necessary to account for the propagation of uncertainty in measured and estimated environmental covariates into the outputs of snail demography models. Covariate extraction and aggregation to monthly steps were done using Google Earth Engine API (https://developers.google.com/earth-engine). Given different temporal coverages of the remotely sensed products, the extraction was made from January 2003 to December 2017.

Precipitation

Among possible remotely sensed products regarding precipitation the Climate Prediction Center's Rainfall Estimate, v. 2.0 (RFEV2), a decadal estimate at ≈ 10km spatial resolution (Xie *et al*., 1997) was retained as it has been singled out as the product with the highest correlation with rain gauge measurements over Burkina Faso (Dembélé *et al.*, 2016). Indeed, the estimates of monthly precipitation correlate well (R2 = 0.92) to ground measurements obtained from the 10 Meteorological agency stations of Burkina Faso (Direction Générale de la Météo, Ouagadouguo, Burkina Faso) for the period 2006-2016. The agreement between the data observed at station *i*, and satellite data was further evaluated in terms of the average monthly bias $B_m(i)$:

$$B_m(i)=\frac{1}{n_m}\sum_{y_m}\left(P_{obs,i}(t)-P_{est,i}(t)\right), \quad \text{Eq. 1}$$

where m $\in$ [1, 12] is the month index; γ_m the set of dates t that have month m; $P_{obs,i}$ and $P_{est,i}$ the observed and estimated monthly precipitation sums at station i; and n_m the number of non-missing datapoints in γ_m. The bias correction was extended to the rest of the dataset by geographical differentiation analysis (GDA) (Cheema *et al.*, 2012). The estimated monthly bias $B_m^*(i)$ in the precipitation estimates at each pixel i was computed through inverse distance weighting (IDW) of the observed biases $B_m(j)$ at each gauging stations j as:

$$B_m^*(i,p)=\sum_{j=1}^{n_s} d_{ij}^{-p}\left(\sum_{j=1}^{n_s} d_{ij}^{-p}\right)^{-1} B_m(j) \qquad \text{Eq. 2}$$

where p is a parameter; ns = 10 the number of available gauging stations; and d_{ij} the Euclidean distance between pixel i and station j (Babak *et al.*, 2009). Parameter p was calibrated through leave-one-out cross-validation by minimizing the sum of squared differences between the observed and IDW-predicted biases at each station

$$\frac{1}{n_s}\sum_{j=1}^{n_s}\left(B_m(j)-B_m^*(j,p)\right)^2 .$$

The uncertainty in the satellite precipitation estimates was quantified by fitting the candidate probability density function to the observed monthly errors. Here, we used the multiplicative error model on the unbiased precipitation estimates according to McMillan *et al.* (2011) and Tian *et al.* (2013):

$$\hat{P}_{est,i}(t)=P_{obs,i}(t)\times e^{\epsilon_{i,t}} \qquad \text{Eq. 3}$$

$$\epsilon_{i,t}=\log\left(\frac{\hat{P}_{est,i}(t)}{P_{obs,i}(t)}\right) \qquad \text{Eq. 4}$$

where $\hat{P}_{est,i}(t)$ is the unbiased rainfall estimate; and $\epsilon_{i,t}$ the multiplicative error term.

Due to climate-dependent differences in estimation errors, the available rain-gauge stations were subdivided into three latitude regions accounting for the South-to-North precipitation gradient by performing a k-means clustering for both the latitude and mean annual precipitation of each pixel (Figure A1). For each month (January through December) and each latitude region, the normal and the asymmetric Laplace (ASL) distributions were fitted to the precipitation estimation errors by maximum likelihood estimation, and the one with the best (smallest) Akaike information criterion (AIC) was retained. Initially proposed for quantile regression, the ASL distribution reads:

$$f(x;\mu,\sigma,p)=\frac{p(1-p)}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\left[p-I(x\leqslant\mu)\right]\right) \qquad \text{Eq. 5}$$

where $\mu \in (-\infty, +\infty)$ is the location parameter; $\sigma > 0$ the scale, $p \in (0, 1)$ the skew; and $I(\cdot)$ is the indication function (Yu *et al.*, 2005). As opposed to the normal distribution, it is suitable for describing data that present more peaked and skewed distribution than the former would assume (Yu *et al.*, 2005).

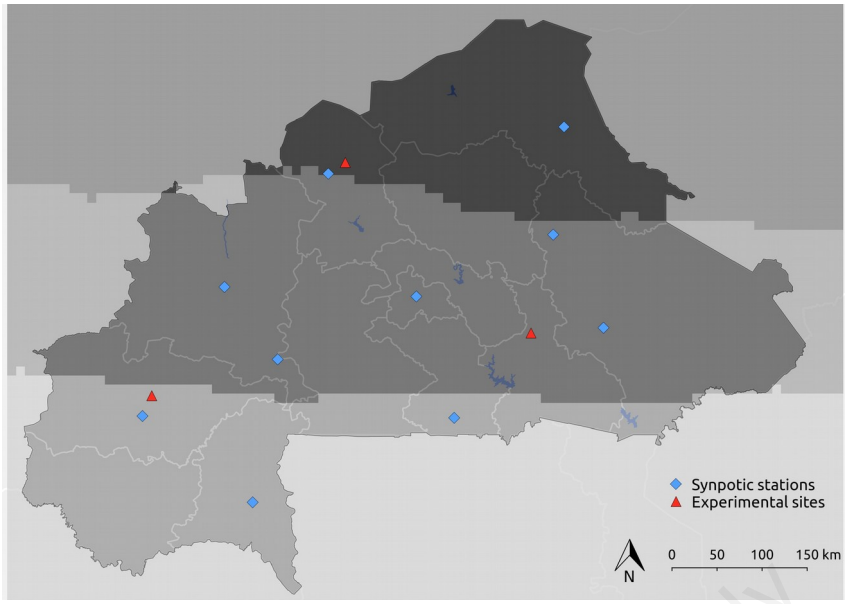


Figure A1. Map of precipitation gauging stations used for the ground-truthing of the remotely sensed estimates. Grey tones indicate the latitude and precipitation clusters used for the partitioning of stations for error model fitting.

Air temperature

The results by Perez-Saez *et al.* (2016) highlighted the link between air and water temperature for the ecological dynamics of both genera of snails across habitats. More particularly, air temperature was systematically selected in the ephemeral habitats, whereas water temperature was more often retained in the permanent habitat, especially for *B. pfeifferi*. Given the national scale of the analysis, predicting water temperature in the whole river network is well beyond the scope of this work. Instead, we used the mean maximal and minimal air temperatures (available from the Burkina Faso National Meteorological Agency) here. Different global scale modelling estimates of temperature at daily to monthly resolutions are available, including the MODIS Aqua/Terra temperature 8-day composites at 1-km spatial resolution (MYD11A2) as done by Wan (2015), and the 3-hour state variable outputs of the Global Land Data Assimilation system (GLDAS)-1, *i.e.* the 3-hour model output at 0.25º resolution (Rodell *et al.*, 2004). Given the discrepancies between the field measurements and remotely sensed temperature estimates, a temperature correction process following methods applied to mapping malaria transmission suitability at national and continental scales was used. Indeed, the spatio-temporal estimation of daily maximal and minimal air temperature based on remotely sensed data has been the subject of numerous studies in the field of malaria given its importance for in the ecology of the mosquito vector, and for the development of the parasite within it (Garske *et al.*, 2013; Weiss *et al.*, 2014). Here, we followed the method proposed by Garske *et al.* (2013), which consists in multivariate regression analysis allowing for random slopes and intercepts on the temporal covariates.

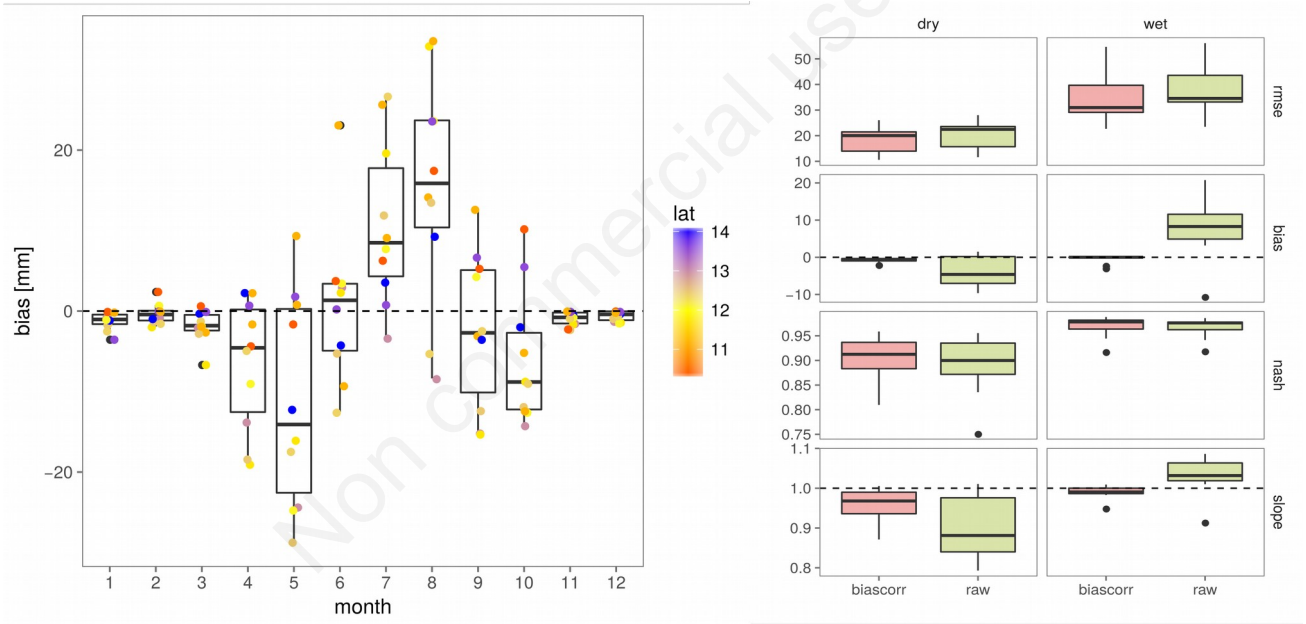
Specifically, the mean monthly maximal and minimal observed temperatures ($T^{o}_{max,}$ T^{o}_{min}) were first modelled as a linear combination with random slopes of the MODIS time series (T^{M}) of the mean daily temperature range (T_{day} - T_{night}), the MODIS daily 1-km resolution enhanced vegetation index (EVI) (Huete *et al.*, 2002), the estimated daylight hours (Forsythe *et al.*, 1995) and the precipitation (cf. preceding section). All covariate combinations were tested and the best one selected using the AIC separately for day and night temperatures. In a second step, the random slopes and intercepts in these best fitting models were estimated as a function of all possible combinations of fixed (non-temporal) covariates, which included longitude, latitude, mean annual EVI and mean annual precipitation (for details see Garske *et al.*, 2013). The models from step 2 with the lowest AIC were

retained for the prediction of monthly day and night temperature. The same approach for quantifying the uncertainty in the errors in the estimated precipitation was applied to the residuals of the multivariate temperature regression models, except that the error models were fit per station and separately for the rainy (June through September) and dry (October through May) seasons, instead of a monthly basis as done for the precipitation data.

2. Uncertainty models for remotes-sensing covariates

Precipitation

RFEV2 has a systematic bias in, generally over-estimating rainfall during the dry season and under-estimating it during the wet season (Figure A2). Of importance for snail population dynamics is the general over-estimation of rainfall in the month of May which can lead to erroneous predictions of snail population bursts. The bias correction methodology successfully reduced bias as excepted, but also slightly reduced the overall error between observed an estimated precipitation (Figure A2). The normal ASL distribution was retained in 5/36 (31/36) month/latitude error partitions, and captured adequately the distribution of precipitation estimation errors (Figure A3).



A **B**

Figure A2. Bias correction of precipitation remote sensing estimates. A) The monthly bias of remote sensing precipitation estimates is non-stationary with systematic over-estimation in the dry season (October through May) and over-estimation in the rain season (June-August). No systematic relationship can be drawn between bias and latitude (point colour). B) Agreement between the estimated and observed precipitation in terms of root mean squared error (RMSE) bias, Nash–Sutcliffe coefficient and linear regression slope with 0 intercept (slope) before bias correction (green) and after (red) partitioned by season (rainy *vs* wet). Bias correction successfully reduces bias as expected and also reduces the estimation error slightly.

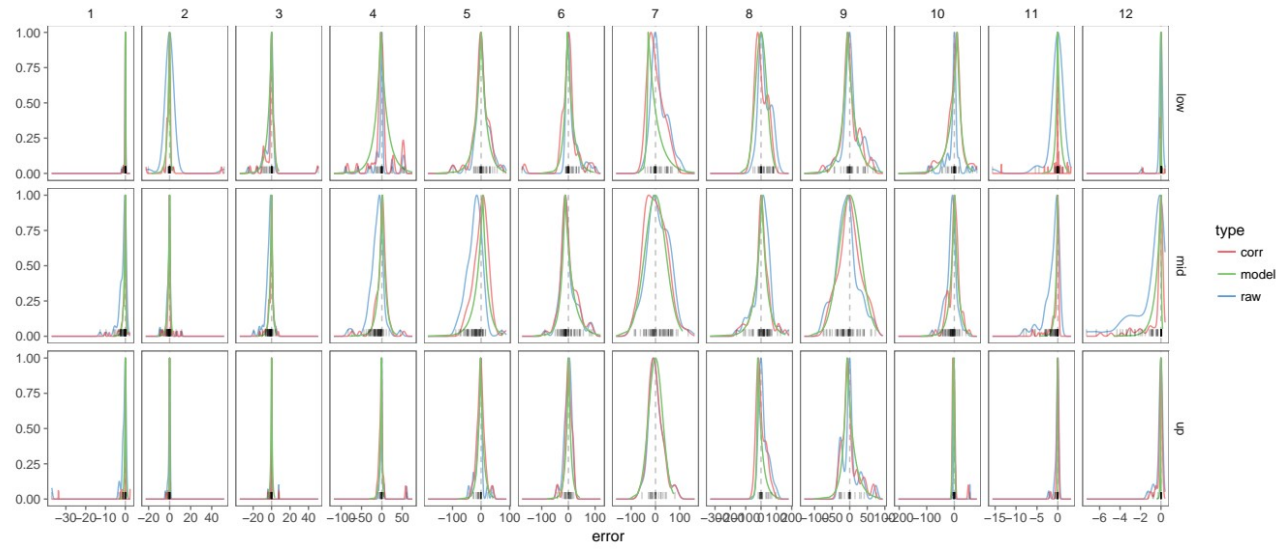


Figure A3. Error models of remotely sensed precipitation estimates. The frequency distribution of the errors in the raw dataset (black ticks and blue density lines) are shown against the normalized density of the bias-corrected dataset (red) and the best-fitting error model (blue) - either the normal or the ASL distribution, by latitude/precipitation zone (low ≈ latitude <13, mid =13 < latitude < 14, high = latitude >14).

Temperature

The best fitting temperature models were in good agreement with observed day and night temperature at all three monitoring stations (R2>0.95, Figure A4). Interestingly, both the GLDAS and MODIS temperature estimates were retained in the best-fitting models. As for the precipitation estimates, both the normal and the ASL distribution were retained (respectively in 9 and 3 of 12 data partitions by day/night and season), and represent well the observed error distributions (Figure A5).

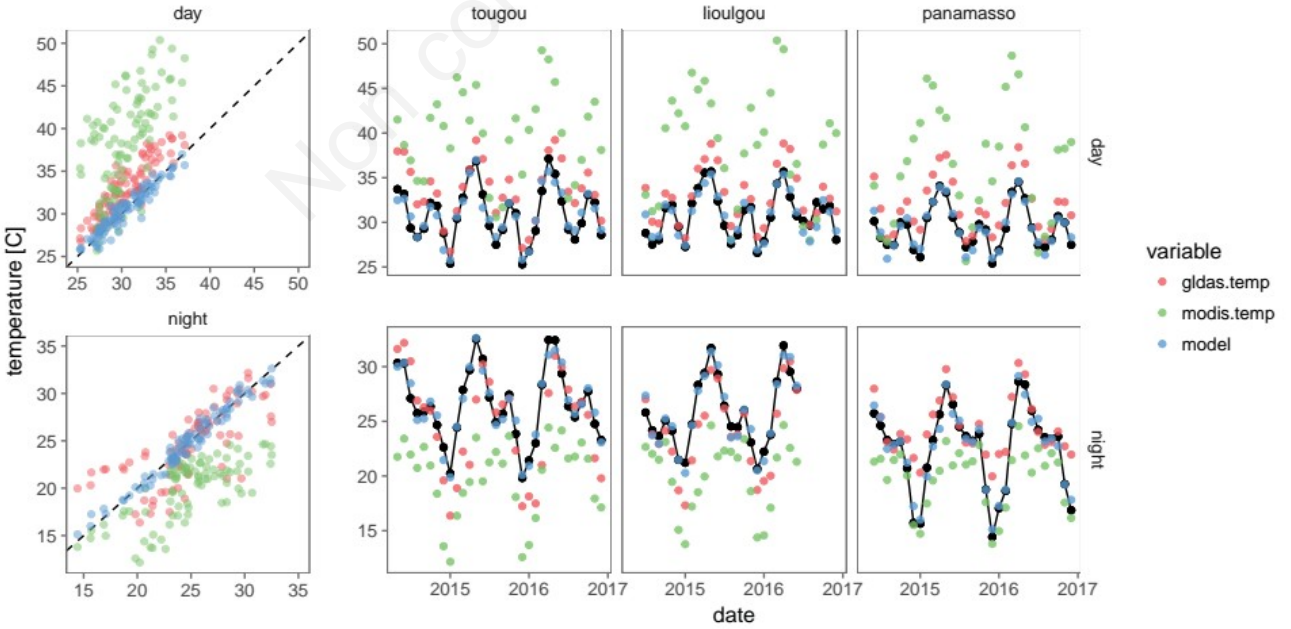


Figure A4. Multivariate regression of day and night temperatures. Day and night temperatures are modelled as a function of time-varying (precipitation, vegetation indices, daylight hours) and fixed (latitude, longitude, covariate means) covariates.

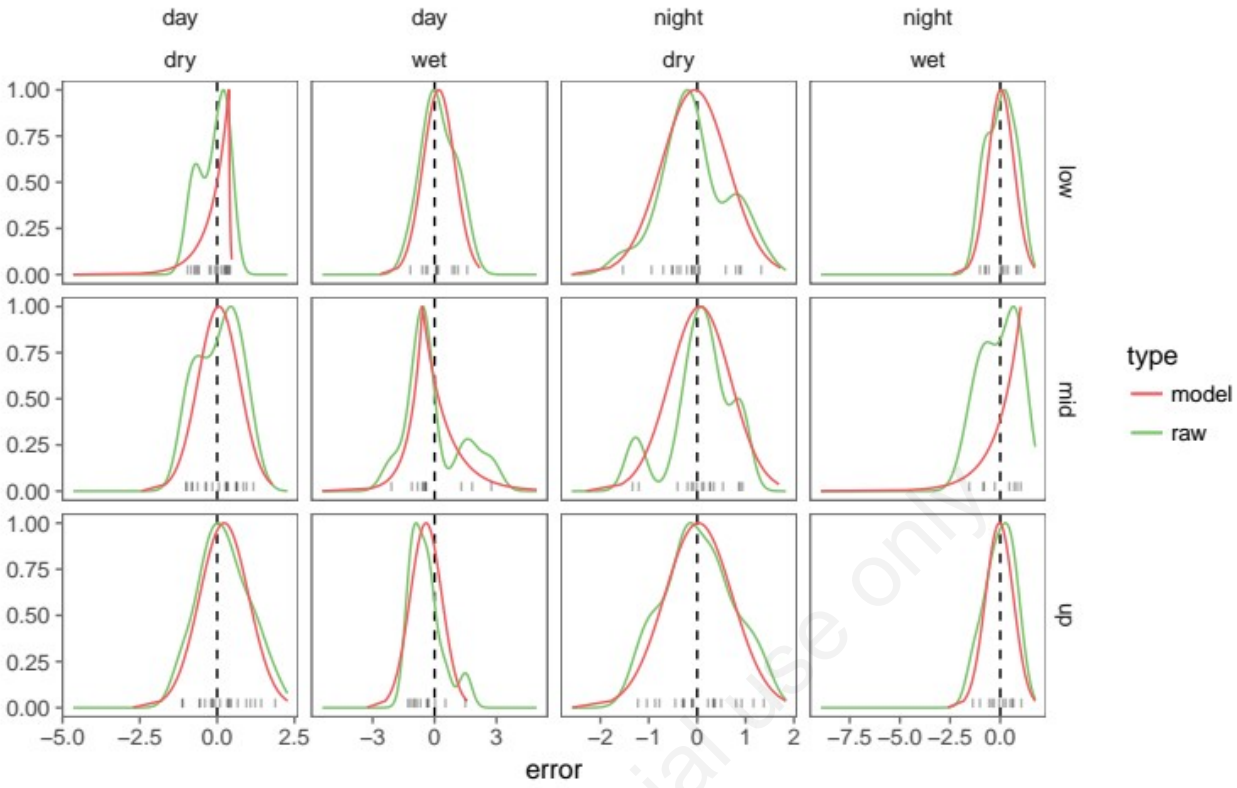


Figure A5. Error models of temperature regression estimates. The frequency distribution of the errors of the multivariate linear regression model (black ticks and green density lines) are shown against the best-fitting error model (red), by latitude zone (low = Panamasso, mid = Lioulgou, high = Tougou) and season (wet *vs* dry).

References

Babak O, Deutsch CV, 2009. Statistical approach to inverse distance interpolation. Stoch Environ Res Risk Assess 23:543-53.

Cheema M, Wim GMB, 2012. Local calibration of remotely sensed rainfall from the TRMM satellite for different periods and spatial scales in the Indus Basin. Int J Remote Sens 33:2603-27.

Dembélé M, Sander JZ, 2016. Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa. Int J Remote Sens 37:3995-4014.

Forsythe WC, Rykiel Jr. EJ, Stahl RS, Wu H, Schoolfield RM, 1995. A model comparison for daylength as a function of latitude and day of year. Ecol Model 80:87-95.

Garske T, Ferguson NM, Ghani AC, 2013. Estimating air temperature and its influence on malaria transmission across Africa. PloS One 8:e56487.

Hong Y, Hsu KL, Moradkhani H, Sorooshian S, 2006. Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. Water Resour Res 42:W08421.

Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG, 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens Environ 83:195-213.

McMillan H, Jackson B, Clark M, Kavetski D, Woods R, 2011. Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. J Hydrol 400:83-94.

Moradkhani H, Sorooshian S, 2009. General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. In: Sorooshian S, Hsu KL, Coppola E, Tomassetti B, Verdecchia M, Visconti G, eds. Hydrological modelling and the water cycle. Water Science and Technology Library, vol 63. Springer, Berlin, Heidelberg, pp 1-24.

Perez-Saez J, Mande T, Ceperley N, Bertuzzo E, Mari L, Gatto M, Rinaldo A, 2016. Hydrology and density feedback control the ecology of intermediate hosts of schistosomiasis across habitats in seasonal climates. Proc Natl Acad Sci USA 113:6427-32.

Rodell M, Houser PR, Jambor UEA, Gottschalck J, Mitchell K, Meng CJ, K Arsenault, Cosogrove B, Radakovich J, Bosilovich M, Entin J, Walker J, Lohmann D, Toll D, 2004. The global land data assimilation system. Bull Am Meteorol Soc 85:381-94.

Tian Y, Huffman GJ, Adler RF, Tang L, Sapiano M, Maggioni V, Wu H, 2013. Modeling errors in daily precipitation measurements: Additive or multiplicative? Geophys Res Lett 40:2060-5.

Wan Z, 2015. MYD11A2 MODIS/Aqua Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V006. NASA EOSDIS Land Processes DAAC. Available from: https://cmr.earthdata.nasa.gov/search/concepts/C194001210-LPDAAC_ECS.html

Weiss DJ, Bhatt S, Mappin B, Van Boeckel TP, Smith DL, Hay SI, Gething PW, 2014. Air temperature suitability for Plasmodium falciparum malaria transmission in Africa 2000-2012: a high-resolution spatiotemporal prediction. Malar J 13:171.

Xie P, Arkin PA, 1997. Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. Bull Am Meteorol Soc 78:2539-58.

Yu K, Zhang J, 2005. A three-parameter asymmetric Laplace distribution and its extension. Commun Stat-Theor M 34:1867-79.