

Supplementary material - The seasonality of cholera in sub-Saharan Africa: a statistical modelling study

Javier Perez-Saez, Justin Lessler, Elizabeth C. Lee, Francisco J. Luquero, Espoir Bwenge Malembaka, Flavio Finger, Jose Paulo Langa, Sebastian Yennan, Benjamin Zaitchik, Andrew S. Azman

Contents

S1 Cholera data and processing	S1
S1.1 Spatial matching and aggregation to GADM Units	S2
S1.2 Temporal aggregation	S2
S1.3 Selection of GADM level for analysis	S2
S2 Seasonality model description	S3
S2.1 Base model	S3
S2.2 Offset model	S5
S2.3 Mixture model	S5
S2.4 Priors	S6
S3 Model comparison	S6
S4 Posterior checks	S9
S4.1 Posterior variance shrinkage	S9
S4.2 Posterior retrodictive checks	S10
S5 Additional results	S11
S6 Seasonality grouping	S15
S7 Correlation with hydro-climatic variables	S17
S8 Alternative cholera occurrence definitions	S19

S1 Cholera data and processing

Cholera incidence data consisted of reports from various sources (ministries of health, WHO, Médecins Sans Frontières, ProMED, ReliefWeb, scientific literature and publicly available epidemiologic reports) with different spatial and temporal resolutions. We define as a location-period (LP) a set of spatial polygons corresponding to the areas where cholera was reported, and a temporal period covered by the incidence report (for instance a given first level administrative unit during a specific epidemiological week). LPs of cholera reports contained in the database were aggregated in space and time to obtain the analysis datasets (first and second administrative levels at the monthly temporal resolution) on which models were run.

S1.1 Spatial matching and aggregation to GADM Units

Our spatial reference for analysis is the GADM Global Administrative Areas dataset (Hijmans, Garcia, and Wieczorek 2010). We first match LPs to GADM units at the country, first and second administrative levels. An LP was defined to match a GADM unit if: 1) the centroids of the two were mutually contained within each other, 2) the LP's area within 90% and 110% the size of the GADM unit's area. For the set of LPs that were within a given GADM units but for which the area was smaller than the 90% threshold we compared the union of their areas to that of the GADM unit within each data source, and flagged whether the union of areas was below, within, or above the 90-110% range used for determining matching. This flag was then used in temporal aggregation.

S1.2 Temporal aggregation

Analysis were run at the monthly time scale, we thus aggregated all sub-monthly observations in each LP to the monthly resolution. These monthly aggregates were then used to define cholera excess occurrence by comparing monthly incidence to the mean monthly incidence estimates extracted from Lessler et al. (2018). Observations of length longer than 1 month were not modified and were compared to the mean incidence with the corresponding exposure period. In cases where the union of areas of matching LPs flagged during spatial aggregation we applied the following rule: 1) if the union of areas was below the 90% threshold we set cholera occurrence to 1 if the monthly incidence was larger than the mean monthly incidence, and dropped the observation if it was smaller; 2) if the union of areas was larger than the 110% threshold we set cholera occurrence to 0 if the monthly incidence was below the mean monthly estimate, and dropped the observation if it was larger.

S1.3 Selection of GADM level for analysis

We selected the analysis GADM level for each country based on data availability. In each country, we computed the yearly fraction of observed administrative 2 level units for observations of length 1 to 3 months. We select the second administrative level if the mean fraction across years was larger than 10% and the maximum larger than 40%. If these conditions were not met we selected the first administrative level.

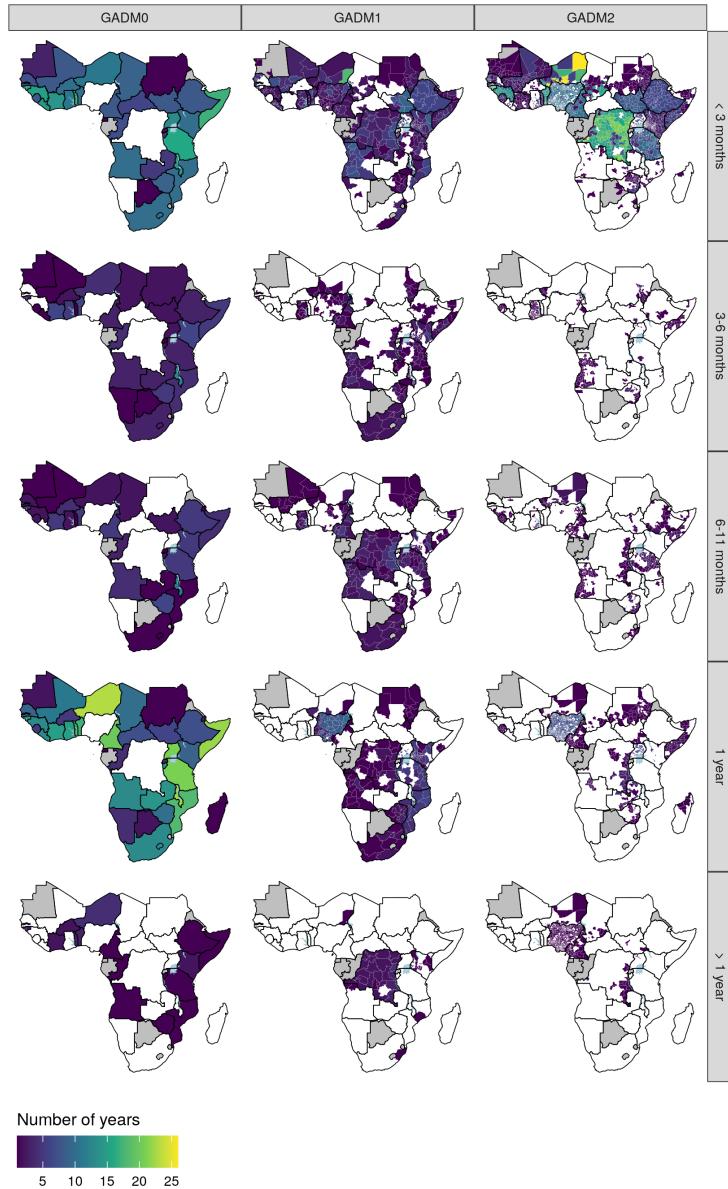


Figure S1: Map of cholera incidence data availability in sub-Saharan Africa. Availability is expressed in terms of the number of distinct years for which at least one observation is available, partitioned by GADM administrative level and observation length. Light gray indicate countries that were excluded due to lack of data.

S2 Seasonality model description

S2.1 Base model

Cholera occurrence is modeled as binary observation process at a monthly temporal resolution. The analysis is done on a per-country basis taking the administrative subdivisions as the spatial units of the disease process. The aim is to draw on information on all administrative levels by exploiting the nested structure of the administrative units, for instance districts within provinces within a country.

The data consists of observations of cholera occurrence per month. Assuming that the administrative level 2 is the lowest administrative level at which cholera is observed, the set of $n_{i,m,y}$ monthly observations for month m in administrative level 2 unit i of year y , $Y_{i,m,y}$, is modeled as resulting from a binomial distribution:

$$Y_{i,m,y} \sim \text{Binomial}(n_{i,m,y}, p_{i,m,y}),$$

where $p_{i,m,y}$ is the combined probability of there being cholera and observing it (the process and measurement models are combined). The monthly probabilities are assumed to follow the BYM model:

$$\text{logit}(p_{i,m,y}) = \beta_m + \eta_y + \theta_i + \phi_i,$$

where β_m is the monthly country-level cholera probability of month m , η_y is a year-specific random effect, and θ_i and ϕ_i are the non-spatial and spatial random effects. We assume that the β s follow a 1-dimensional cyclic random walk, i.e. $\beta_m - \beta_{m-1} \sim \mathcal{N}(0, \sigma_\beta)$ $\forall m \in [2, 12]$ and $\beta_1 - \beta_{12} \sim \mathcal{N}(0, \sigma_\beta)$ to impose the cyclic pattern. To allow the identifiability of the model we further impose $\sum_{m=1}^{12} \beta_m = 0$. We assume that the spatial random effects follow the ICAR model, and use the reparameterized version of the BYM model (Riebler et al. 2016):

$$\theta_i + \phi_i = \epsilon_i = \frac{1}{\sqrt{\tau}} \left(\sqrt{1-\rho} \theta_i^* + \sqrt{\frac{\rho}{s}} \phi_i \right),$$

where τ is the overall marginal precision (i.e. the total residual relative risk), $\theta^* \sim \mathcal{N}(0, 1)$, $\phi \sim \text{ICAR}$, s is a scaling factor that depends on the adjacency matrix of the neighborhood network to ensure $\text{Var}(\phi^*) \approx 1$, and $\rho \in [0, 1]$ indicates whether the marginal variance comes from spatial or non-spatial random effects.

At the upper administrative level the observation process at the admin unit level 1 j is given by the probabilities of cholera reporting in all areas at the level 2 contained in it, $i \sim j$, as:

$$\begin{aligned} Y_{j,m,y} &\sim \text{Binomial}(n_{j,m,y}, p_{j,m,y}) \\ p_{j,m,y} &= 1 - \prod_{i \sim j} 1 - p'_{i,m,y} \\ \text{logit}(p'_{i,m,y}) &= \text{logit}(p_{i,m,y}) + \xi_l, \end{aligned}$$

where $p'_{i,m,y}$ is the probability of cholera reported at the lower administrative level modified by a reporting modifier ξ_l at administrative level l . For instance, in the case models were run at administrative level 2, a different effect modifier was applied to the first administrative level and country-wide observations. The country-level monthly observations follow the same observation process with the product over all spatial subunits.

We also account for multi-month observations. The probability of observing cholera in admin unit i in year y for period $T_n = \{m_1, \dots, m_n\}$ composed of n months is given by:

$$p_{i,T_n,y} = 1 - \prod_{m=m_1}^{m_n} 1 - p_{i,m,y}.$$

Similarly multi-month observations at any upper admin level j is obtained by multiplying the individual probabilities:

$$p_{j,T_n,y} = 1 - \prod_{i \sim j} \prod_{m=m_1}^{m_n} 1 - p'_{i,m,y}.$$

S2.2 Offset model

The base model considers yearly random effects based on the calendar year, which may be unrealistic in some situations. For instance if cholera has a single peak in December/January, calendar-based random effects will apply only to the first “half” of the epidemic with the second half falling in the following calendar year.

To account for this fact we introduce an offset in the assignment of the yearly random effect based on the month in order to align the yearly random effects to “cholera” years by introducing a latent discrete variable representing the offset. The probability of occurrence becomes:

$$\text{logit}(p_{i,j,m}) = \beta_m + \eta_{f(yr,m,z)} + \epsilon_i,$$

where $f(yr, m, z)$ gives the year corresponding to the month accounting for the offset.

$$f(yr, m, z) = \begin{cases} yr & \text{if } m + z \leq 12 \\ yr + 1 & \text{if } m + z > 12 \end{cases}, \quad z \in [0, 11].$$

For the purpose of sampling z can be marginalized out of the likelihood as:

$$\begin{aligned} p(Y|\Theta) &= \sum_{z=0}^{11} p(Y|\Theta, z)p(z), \\ &= \sum_{z=0}^{12} \text{Uniform}(z|0, 11) \prod_{j=1}^{n_{obs}} \text{Binomial}(Y_j|n_j, p_{j,z}) \end{aligned}$$

where Θ is the vector of all parameters defining the occurrence probabilities $(\beta, \eta, \rho, \tau, \phi, \theta, \xi)$, and subscript j indicates the observation (combination of admin level, month and year).

And the log-likelihood as:

$$\log(p(Y|\Theta)) = \log \left\{ \sum_{z=0}^{11} \exp(l_z) \right\},$$

where l_z are the individual sum of marginal log-likelihoods and the log of the prior on z :

$$l_z = \log(p(Y|\Theta, z)p(z)) = \log(\text{Uniform}(z|0, 12)) + \sum_{j=1}^{n_{obs}} \log(\text{Binomial}(Y_j|n_j, p_{j,z})).$$

S2.3 Mixture model

Assuming the possible coexistence of two distinct groups of seasonalities within the same country, a latent class model for the probabilities of observing cholera can be written in terms of a mixture of two different vectors of seasonality coefficients β^{g_1, g_2} :

$$\text{logit}(p_{i,j,m}) = \lambda_i \beta_m^{g_1} + (1 - \lambda_i) \beta_m^{g_2} + \eta_{f(yr,m,z)} + \epsilon_i,$$

with $\lambda_i \in (0, 1)$ representing the mixture probabilities, with priors $\lambda \sim \text{beta}(0.5, 0.5)$. To mitigate label switching issues we impose the constraints $\beta_1^{g_2} > \beta_1^{g_1}$ and $\beta_8^{g_2} < \beta_8^{g_1}$. We allow for spatial autocorrelation in mixture coefficients by implementing an approach similar to that used for spatial random effects above:

$$\begin{aligned} \text{logit}(\lambda_i) &= \theta_{\lambda_i} + \phi_{\lambda_i}, \\ \theta_{\lambda_i} + \phi_{\lambda_i} &= \frac{1}{\sqrt{\tau_\lambda}} \left(\sqrt{1 - \rho_\lambda} \theta_{\lambda_i}^* + \sqrt{\frac{\rho_\lambda}{s}} \phi_{\lambda_i} \right), \end{aligned}$$

with parameter definitions as for the spatial random effects.

S2.4 Priors

We use the following priors:

$$\begin{aligned}\xi_l &\sim \mathcal{N}(0, 1), \\ \tau_\beta &\sim \mathcal{N}(0, 2.5), \\ \eta_t &\sim \mathcal{N}(0, \sigma_\eta), \\ \sigma_\eta &\sim \mathcal{N}(0, 2.5),\end{aligned}$$

where $\tau_\beta = 1/\sigma_\beta^2$ is the precision of the cyclic random walk of seasonality coefficients. Priors for the spatial random effects were described above. Penalized-complexity (PC) priors are chosen for τ and ϕ following (Riebler et al. 2016): $\pi(\tau) = \frac{\theta}{2} \tau^{-3/2} \exp(-\theta\tau^{-1/2})$, with $\theta = -\log(\alpha_\tau)/U_\tau$ where $P\left(\frac{1}{\sqrt{\tau}} > U_\tau\right) = \alpha_\tau$ with α_τ, U_τ set to represent our belief in the importance of marginal variance. Similarly a PC prior for ρ can be formulated in terms of the probability $P(\rho < U_\rho) = \alpha_\rho$, which does not have a closed-form expression but for which tabulated values can be computed. We set $U_\tau = 1$ and $\alpha_\tau = 0.75$, and $U_\rho = .5$ and $\alpha_\rho = 2/3$ following Riebler et al. (2016). The same priors were chosen for spatially autocorrelated mixture probabilities for the mixture models.

S3 Model comparison

Table S1: Model fit results and model selection. ELPD diff: difference in estimated log predictive density with respect to best-fitting model. SE: standard error of the ELPD difference. LOO-IC: Leave-one-out cross-validation estimate Information Criterion. Rank: model rank. Weight: Evidence weight in favor of the model.

model	ELPD diff	SE	LOO-IC	rank	weight
Angola					
offset	-	-	1582.3	1	1.00
mixture	-88.8	17.3	1760.0	2	0.00
base	-92.1	15.7	1766.5	3	0.00
null	-110.3	17.5	1802.9	4	0.00
mixture_offset	-905.2	161.4	3392.7	5	0.00
Benin					
mixture	-	-	1042.8	1	0.94
base	-2.7	2.9	1048.2	2	0.06
null	-31.9	8.6	1106.5	3	0.00
offset	-90.3	11.1	1223.3	4	0.00
mixture_offset	-1725.3	446.6	4493.3	5	0.00
Burkina Faso					
base	-	-	29.0	1	0.61
mixture	-0.6	0.2	30.1	2	0.34
null	-2.6	1.1	34.1	3	0.05
offset	-13.4	3.7	55.7	4	0.00
mixture_offset	-297.6	114.6	624.1	5	0.00
Burundi					
null	-	-	180.2	1	0.41
base	-0.2	1.6	180.5	2	0.35
mixture	-0.6	1.4	181.3	3	0.23
offset	-20.8	3.7	221.7	4	0.00
mixture_offset	-245.0	76.3	670.1	5	0.00
Côte d'Ivoire					
offset	-	-	132.8	1	1.00
null	-9.4	4.8	151.7	2	0.00
mixture	-9.6	4.9	151.9	3	0.00
base	-10.6	5.1	154.0	4	0.00
mixture_offset	-68.1	17.9	269.0	5	0.00
Cameroon					
mixture	-	-	1804.7	1	1.00
mixture_offset	-9.8	11.0	1824.2	2	0.00
base	-64.8	10.3	1934.3	3	0.00
null	-75.7	12.8	1956.0	4	0.00
offset	-252.3	18.9	2309.2	5	0.00
Central African Republic					
mixture	-	-	75.2	1	0.32
base	-0.3	0.3	75.8	2	0.24
offset	-0.3	3.1	75.9	3	0.23
null	-0.4	1.1	76.0	4	0.21
mixture_offset	-464.0	94.4	1003.2	5	0.00

(Continued on Next Page...)

Table S1: (*continued*)

model	ELPD diff	SE	LOO-IC	rank	weight
Democratic Republic of the Congo					
mixture	-	-	21240.7	1	0.99
mixture_offset	-4.9	37.0	21250.5	2	0.01
base	-132.7	20.2	21506.2	3	0.00
null	-184.5	24.6	21609.6	4	0.00
offset	-379.7	39.2	22000.1	5	0.00
Djibouti					
null	-	-	23.4	1	0.49
offset	-1.0	1.0	25.3	2	0.18
mixture	-1.1	0.5	25.5	3	0.17
base	-1.1	0.7	25.6	4	0.16
mixture_offset	-21.3	13.4	65.9	5	0.00
Ethiopia					
mixture	-	-	2499.4	1	1.00
base	-46.3	9.6	2592.0	2	0.00
null	-103.8	15.7	2706.9	3	0.00
offset	-132.1	15.9	2763.7	4	0.00
mixture_offset	-1164.4	161.0	4828.3	5	0.00
Ghana					
null	-	-	357.4	1	0.97
mixture	-4.0	2.4	365.5	2	0.02
base	-4.2	2.0	365.9	3	0.01
offset	-42.7	9.3	442.9	4	0.00
mixture_offset	-491.2	87.0	1339.8	5	0.00
Guinea					
mixture	-	-	709.4	1	1.00
base	-17.9	5.4	745.1	2	0.00
mixture_offset	-18.4	5.9	746.2	3	0.00
offset	-69.9	9.9	849.2	4	0.00
null	-79.1	13.8	867.6	5	0.00
Guinea-Bissau					
mixture	-	-	719.1	1	0.91
base	-2.4	2.8	723.8	2	0.09
null	-95.9	17.6	910.8	3	0.00
offset	-253.5	35.1	1226.0	4	0.00
mixture_offset	-647.2	125.8	2013.5	5	0.00
Kenya					
offset	-	-	1805.2	1	1.00
mixture	-41.7	13.8	1888.7	2	0.00
null	-82.6	13.7	1970.4	3	0.00
base	-87.3	12.9	1979.9	4	0.00
mixture_offset	-1191.5	150.0	4188.2	5	0.00
Liberia					
null	-	-	376.1	1	0.82
mixture	-1.8	3.1	379.8	2	0.13
base	-2.8	3.4	381.7	3	0.05
mixture_offset	-6.1	5.2	388.4	4	0.00
offset	-66.3	8.0	508.7	5	0.00
Madagascar					
mixture	-	-	0.7	1	0.21
base	-0.0	-	0.7	2	0.21
null	-0.0	-	0.8	3	0.20
offset	-0.1	0.1	0.9	4	0.19
mixture_offset	-0.1	0.1	0.9	5	0.19
Malawi					
offset	-	-	1247.6	1	1.00
mixture	-12.6	12.4	1272.8	2	0.00
base	-29.6	11.3	1306.9	3	0.00
null	-59.3	13.6	1366.3	4	0.00
mixture_offset	-1149.6	99.4	3546.8	5	0.00
Mali					
mixture	-	-	347.7	1	0.95
base	-2.9	2.1	353.5	2	0.05
null	-21.7	7.3	391.1	3	0.00
offset	-150.5	10.6	648.7	4	0.00
Mozambique					
mixture	-	-	240.3	1	0.93
base	-2.7	1.0	245.6	2	0.07
null	-7.7	3.3	255.7	3	0.00
offset	-47.0	7.2	334.2	4	0.00
mixture_offset	-732.8	201.2	1705.8	5	0.00
Namibia					
mixture	-	-	20.3	1	0.37
base	-0.1	0.4	20.5	2	0.33
null	-0.2	0.2	20.8	3	0.30
offset	-13.7	3.8	47.7	4	0.00
mixture_offset	-689.2	620.7	1398.7	5	0.00
Niger					
mixture	-	-	2366.0	1	1.00
base	-31.6	7.1	2429.3	2	0.00

(Continued on Next Page...)

Table S1: (*continued*)

model	ELPD diff	SE	LOO-IC	rank	weight
mixture_offset	-118.4	22.9	2602.8	3	0.00
null	-133.7	19.2	2633.4	4	0.00
offset	-150.2	20.5	2666.5	5	0.00
Nigeria					
mixture_offset	-	-	15298.2	1	1.00
mixture	-81.9	45.7	15462.1	2	0.00
base	-244.4	44.4	15787.1	3	0.00
offset	-537.2	54.7	16372.6	4	0.00
null	-541.0	56.0	16380.2	5	0.00
Senegal					
null	-	-	179.2	1	0.85
mixture	-2.3	1.5	183.8	2	0.08
base	-2.5	1.7	184.2	3	0.07
offset	-56.5	4.7	292.2	4	0.00
mixture_offset	-173.6	32.4	526.3	5	0.00
Sierra Leone					
mixture	-	-	265.6	1	0.99
base	-5.1	2.8	275.7	2	0.01
offset	-7.8	6.6	281.1	3	0.00
null	-14.5	7.5	294.7	4	0.00
mixture_offset	-839.1	163.3	1943.8	5	0.00
Somalia					
mixture_offset	-	-	1336.6	1	1.00
mixture	-22.5	9.7	1381.6	2	0.00
base	-25.8	9.5	1388.3	3	0.00
null	-65.9	13.9	1468.4	4	0.00
offset	-137.2	13.3	1611.1	5	0.00
South Africa					
base	-	-	190.0	1	0.63
mixture	-0.5	1.0	191.1	2	0.37
null	-7.8	3.4	205.6	3	0.00
offset	-8.1	4.5	206.3	4	0.00
mixture_offset	-285.8	133.7	761.6	5	0.00
South Sudan					
mixture	-	-	1112.4	1	1.00
base	-38.6	8.9	1189.6	2	0.00
offset	-41.4	16.7	1195.2	3	0.00
null	-59.8	15.4	1231.9	4	0.00
mixture_offset	-4847.1	559.5	10806.6	5	0.00
Sudan					
mixture	-	-	783.0	1	1.00
base	-42.5	11.5	868.1	2	0.00
offset	-128.5	16.0	1040.1	3	0.00
null	-230.8	19.5	1244.6	4	0.00
mixture_offset	-878.0	272.1	2539.0	5	0.00
Tanzania					
mixture_offset	-	-	3172.1	1	1.00
mixture	-143.9	27.5	3459.9	2	0.00
offset	-147.6	25.4	3467.4	3	0.00
base	-150.8	27.2	3473.7	4	0.00
null	-169.5	29.5	3511.2	5	0.00
Tchad					
mixture	-	-	480.4	1	1.00
base	-7.3	3.4	495.1	2	0.00
mixture_offset	-40.4	7.1	561.1	3	0.00
null	-69.5	11.7	619.3	4	0.00
offset	-99.6	11.2	679.6	5	0.00
Togo					
null	-	-	375.6	1	1.00
mixture	-5.4	2.3	386.4	2	0.00
base	-9.0	2.3	393.7	3	0.00
offset	-57.6	9.0	490.7	4	0.00
mixture_offset	-580.1	159.3	1535.8	5	0.00
Uganda					
mixture	-	-	2000.5	1	1.00
base	-18.1	7.3	2036.6	2	0.00
null	-21.9	10.4	2044.3	3	0.00
offset	-62.8	19.3	2126.2	4	0.00
mixture_offset	-1683.6	292.2	5367.7	5	0.00
Zambia					
offset	-	-	142.5	1	0.99
mixture	-5.5	3.5	153.5	2	0.00
base	-6.0	3.8	154.5	3	0.00
null	-8.3	4.1	159.0	4	0.00
mixture_offset	-301.2	99.6	744.9	5	0.00
Zimbabwe					
mixture	-	-	751.5	1	0.91
base	-2.4	1.7	756.2	2	0.09
null	-10.4	8.7	772.2	3	0.00
offset	-22.6	13.1	796.7	4	0.00

(Continued on Next Page...)

Table S1: (continued)

model	ELPD diff	SE	LOO-IC	rank	weight
mixture_offset	-6795.0	1108.9	14341.6	5	0.00

ture_offset & -6795.0 & 1108.9 & 14341.6 & 5 & 0.00 * \end{longtable}

S4 Posterior checks

S4.1 Posterior variance shrinkage

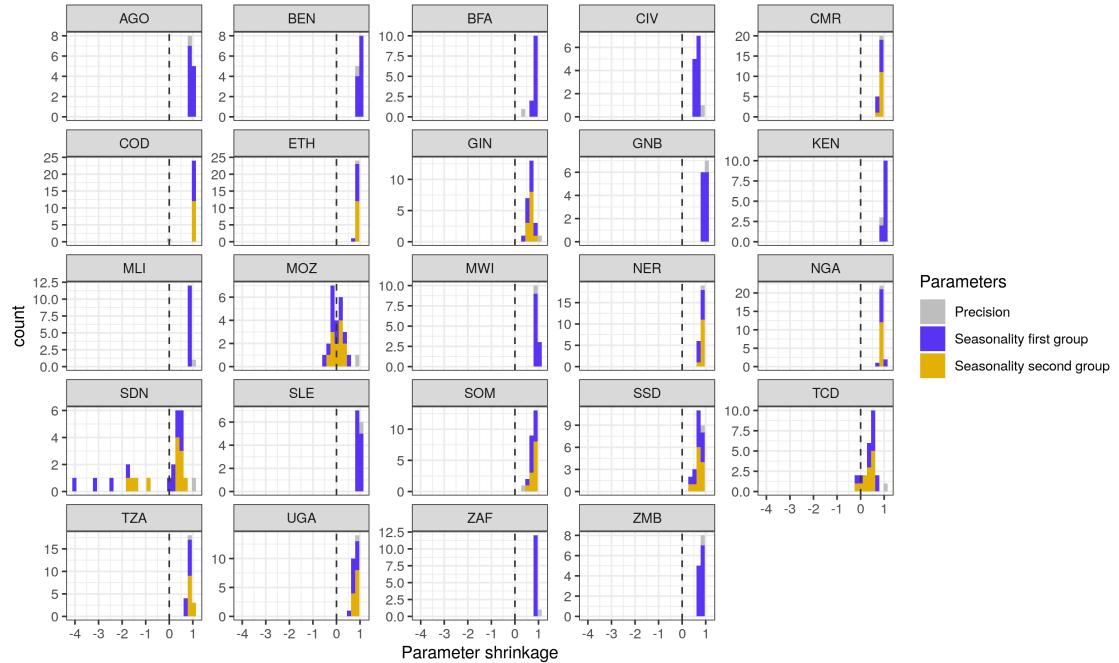


Figure S2: Impact of priors. For each country the impact of priors on seasonality-related parameters was evaluated in terms of the shrinkage of variance from the prior to the posterior, computed as $1 - \text{var}(\text{posterior})/\text{var}(\text{prior})$.

S4.2 Posterior retrodictive checks

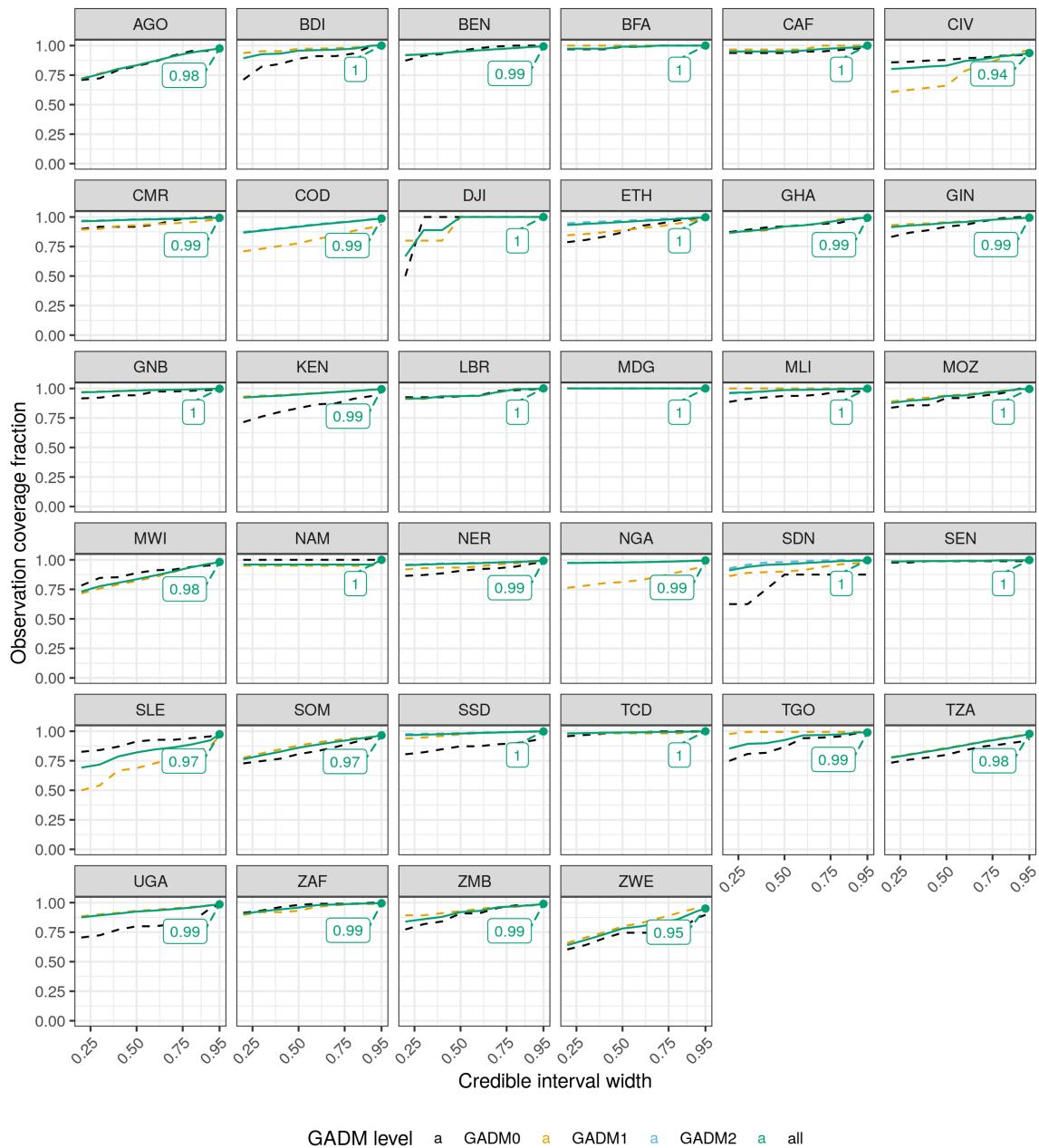


Figure S3: Model evaluation. For each country the performance of the selected model is given as the fraction of observations covered by posterior retrodictions as a function of the credible interval width. The coverage fraction is given for all observations as well as partitioned by administrative level.

S5 Additional results

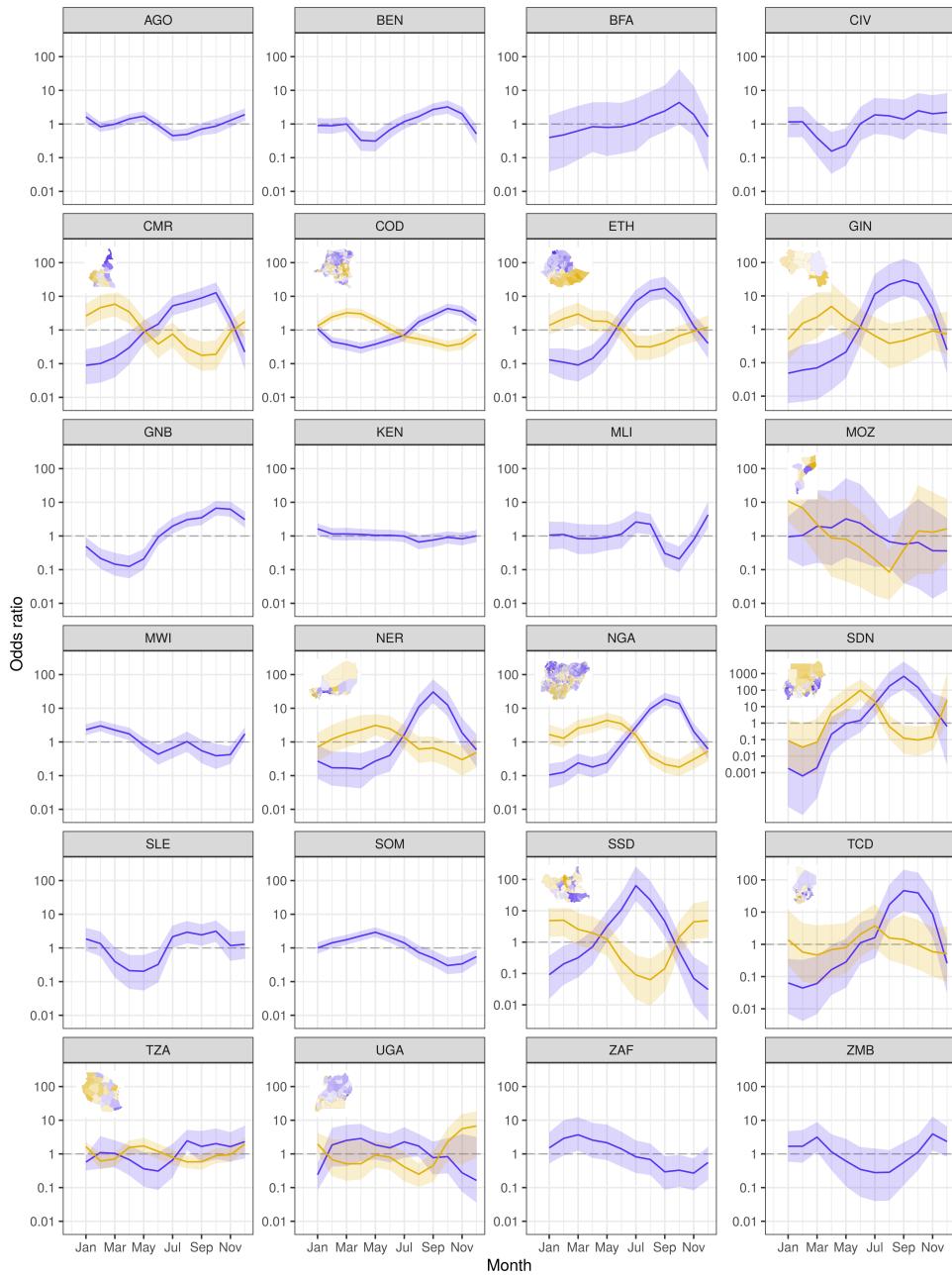


Figure S4: Country-level cholera seasonality time series. Excess cholera seasonality is shown in for the final selection of models in terms of the mean (line) and 95% CrI odds ratios.

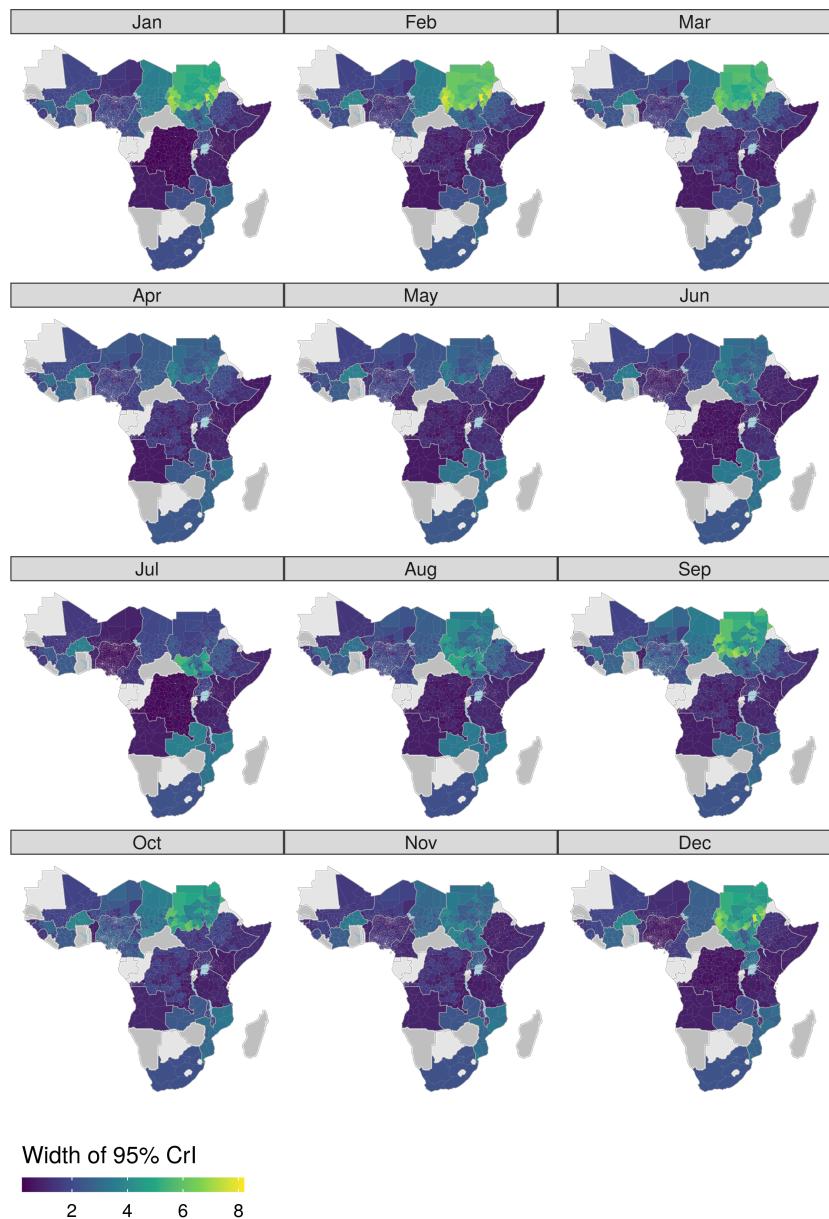


Figure S5: Uncertainty in seasonality coefficients in terms of the width of the 95 % CrI on the logit scale.

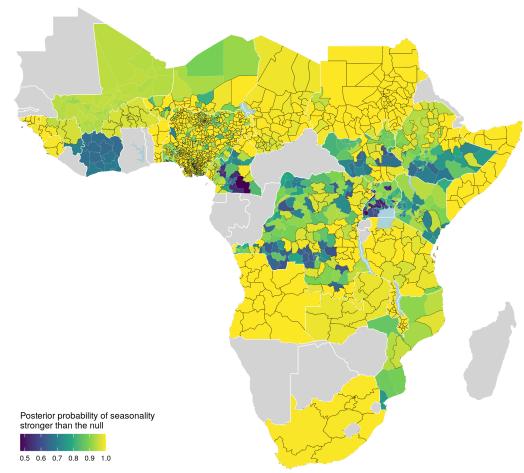


Figure S6: Uncertainty in seasonality index in terms of the posterior probability that the index is larger than the null hypothesis of equal occurrence probability during the year. Admininistrative units for which the posterior probability is above 0.9 are outlined in black.

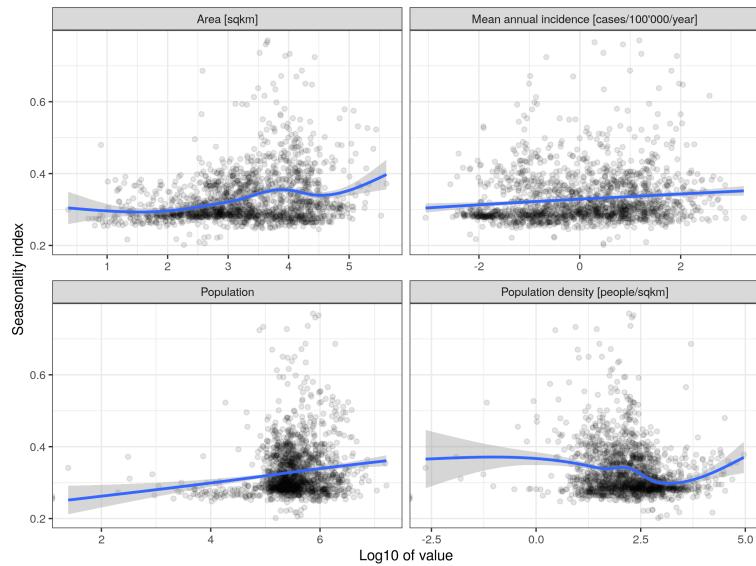


Figure S7: Seasonality strength and administrative unit characteristics. Each point corresponds to one administrative unit, lines correspond to GAM estimates of the relationship between the characteristics and the seasonality indices.

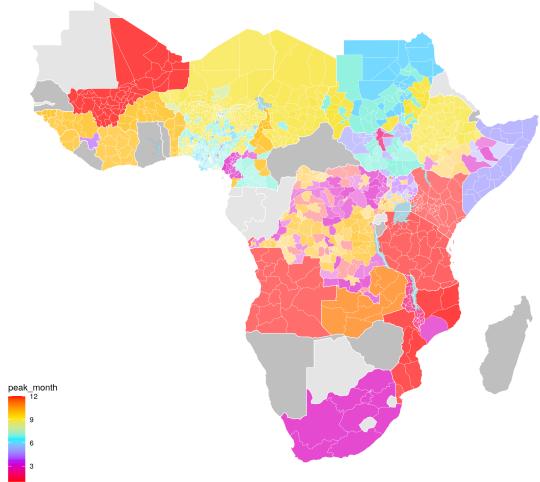


Figure S8: Seasonality peak map. Map of the month with largest estimated excess cholera occurrence odds ratio in countries for which a model with seasonality was selected.

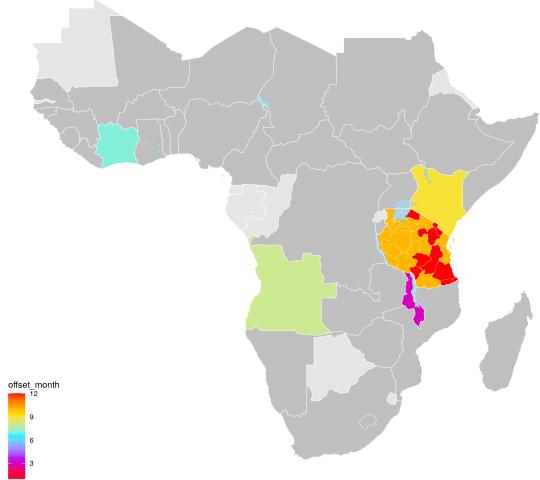


Figure S9: Offset map. Map of the month determining the start of the 'cholera year' in countries for which a model with offset was selected. Countries with two colors correspond to models with both mixture and offset.

S6 Seasonality grouping

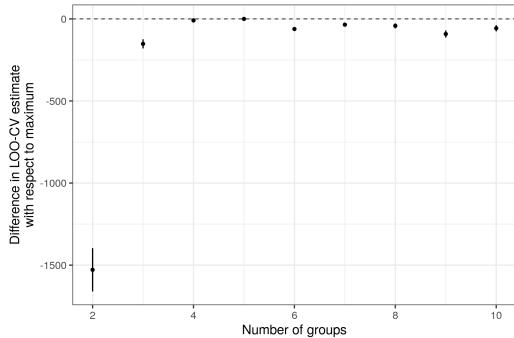


Figure S10: Model comparison of number of groupings. Model comparison was done in terms of the estimated log-predictive density (elpd). Pairwise differences in elpd were computed for all model combinations, and results are shown in terms of the mean (points) and 2 standard errors (lines). The model with highest elpd was the model with 5 groups.

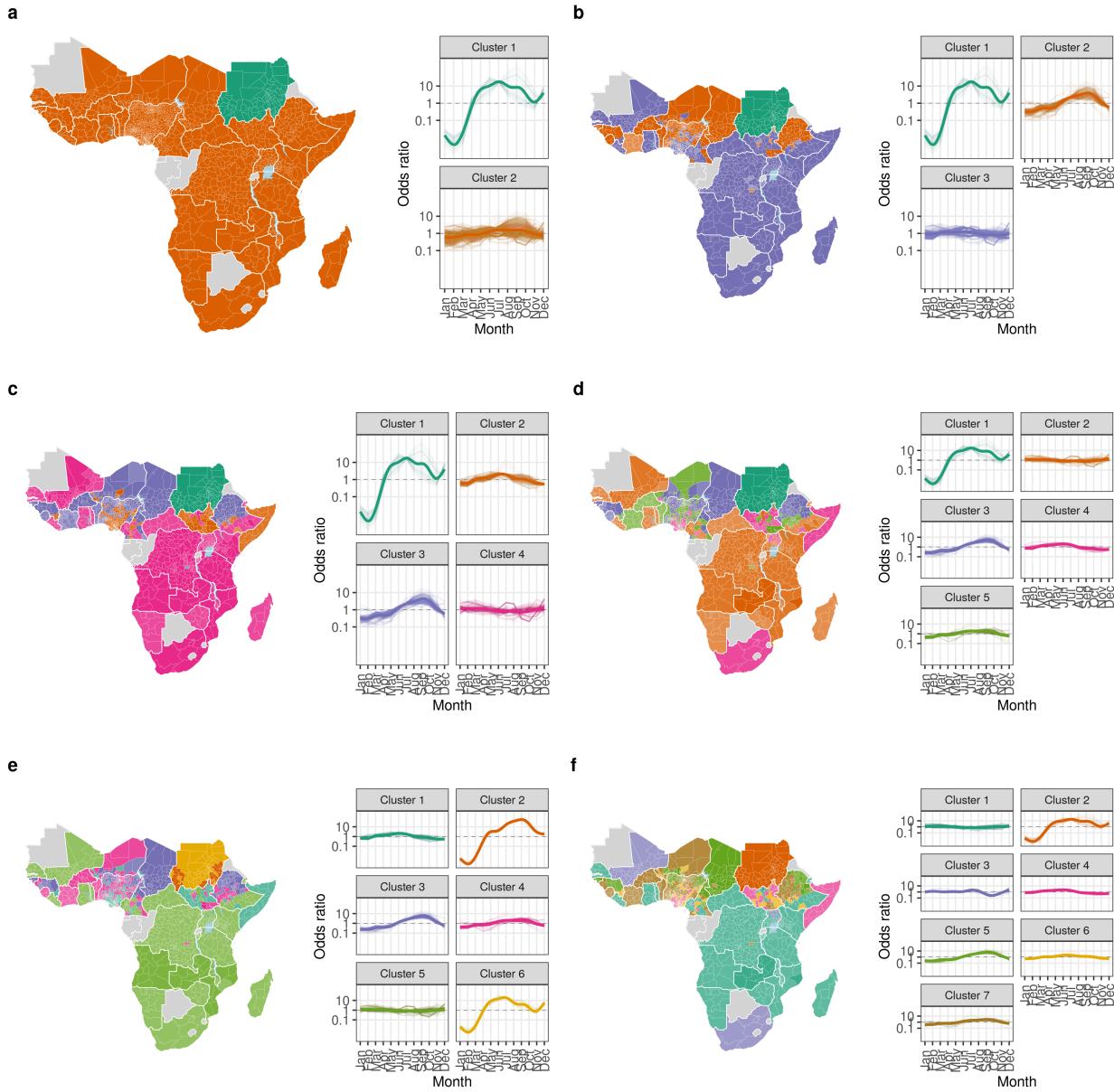


Figure S11: Clustering results for two to seven clusters.

S7 Correlation with hydro-climatic variables

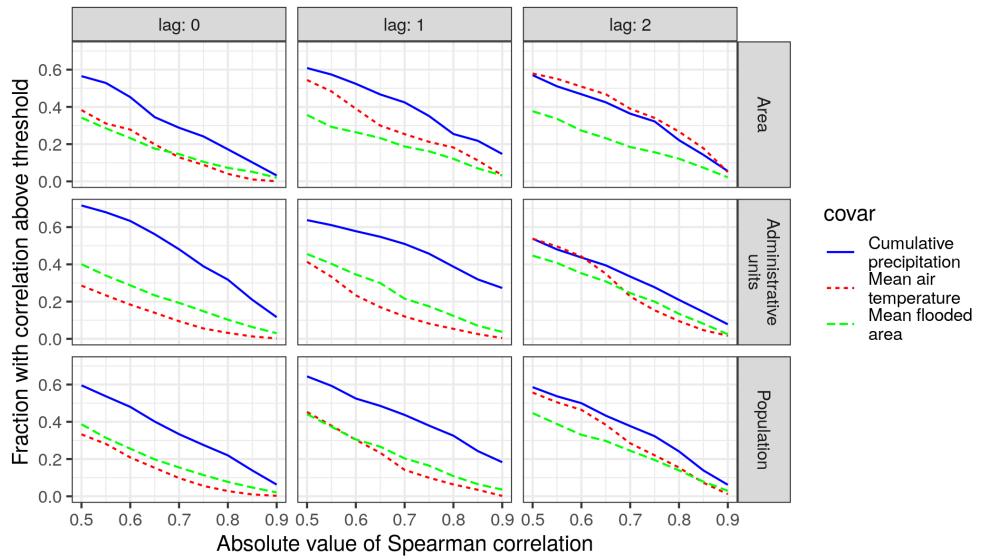


Figure S12: Importance of correlations with hydro-climatic variables. Importance is given in terms of the fraction of area, administrative units, and population with the absolute value of Spearman correlation above thresholds between 0.5 and 0.9.

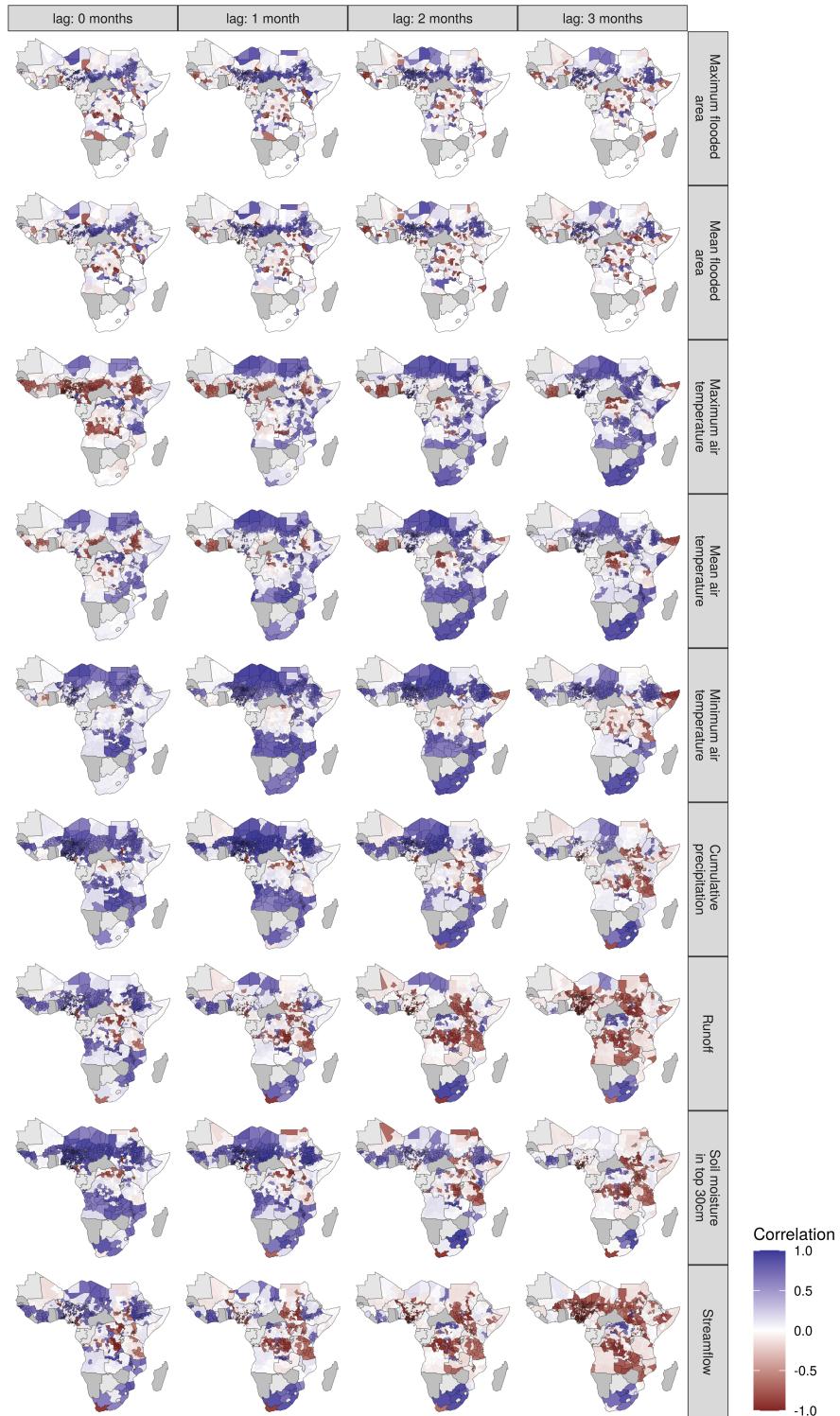


Figure S13: Correlation between excess cholera seasonality and hydro-climatic variables. Correlation is expressed in terms of Spearman's coefficient, with areas with significant correlations outlined in black.

S8 Alternative cholera occurrence definitions

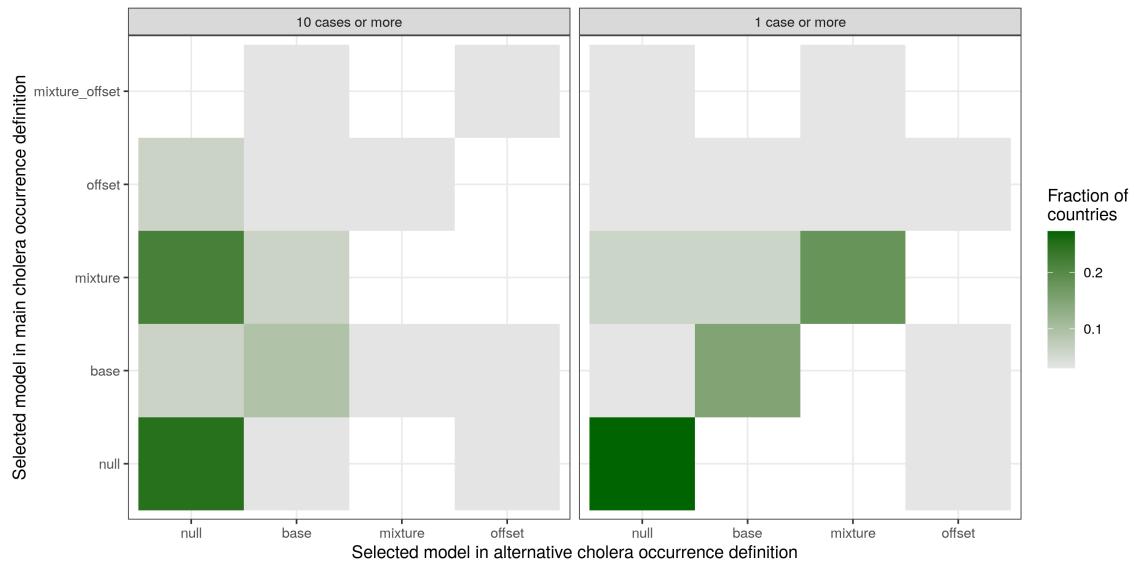


Figure S14: Comparison of model selection for alternative definitions of cholera occurrence

References

- Hijmans, Robert, Nell Garcia, and John Wieczorek (2010). “GADM: database of global administrative areas”. In: *Version 3.6 (released May 6, 2018).[Online]* Retrieved March 12, 2019 from.
- Lessler, Justin et al. (2018). “Mapping the burden of cholera in sub-Saharan Africa and implications for control: an analysis of data across geographical scales”. In: *The Lancet* 391.10133, pp. 1908–1915.
- Riebler, Andrea et al. (2016). “An intuitive Bayesian spatial model for disease mapping that accounts for scaling”. In: *Statistical methods in medical research* 25.4, pp. 1145–1165.