Supplementary Information Appendix

The seasonality of non-polio enteroviruses in the United States: patterns and drivers

Margarita Pons-Salort, M. Steven Oberste, Mark A. Pallansch, Glen R. Abedi, Saki Takahashi, Bryan T. Grenfell, Nicholas C. Grassly

Contents

S1 Seasonal model

To further describe the seasonal pattern of incidence in each state and each year, we fitted a seasonal model with four harmonic terms to the proportion of cases reported each month, for each year of data with >50 cases reported. The seasonal model used was:

$$I_{ij}(t) = a_{ij} + b_{ij}\cos\left(\frac{2\pi t}{12}\right) + c_{ij}\sin\left(\frac{2\pi t}{12}\right) + d_{ij}\cos\left(\frac{4\pi t}{12}\right) + e_{ij}\sin\left(\frac{4\pi t}{12}\right)$$

where $I_{ij}(t)$ is the proportion of cases in state i and year j reported during month t, for $t \in 1, \ldots, 12$. The estimated coefficients for each state i and each year j were used to extract the amplitude of the annual, A_{ij}, and semi-annual, SA_{ij}, components, the timing of the peak, P_{ij} (measured as the phase of the annual component), and the relative contribution of the semi-annual component, R_{ij}, as follows:

$$A_{ij} = \sqrt{b^2 + c^2}$$

$$SA_{ij} = \sqrt{d^2 + e^2}$$

$$P_{ij} = \begin{cases} \arctan(c/b) & \text{if } b, c > 0 \\ \arctan(c/b) + \pi & \text{if } b, c < 0 \\ -\arctan(c/b) + \frac{\pi}{2} & \text{if } c > 0, b < 0 \\ -\arctan(c/b) + \frac{3\pi}{2} & \text{if } c < 0, b > 0 \end{cases}$$

$$R_{ij} = \frac{SA_{ij}}{A_{ij} + SA_{ij}}$$

S2 Testing for differences in the latitudinal gradients for the amplitude and the timing of the peak

To test for differences between enteroviruses and poliomyelitis in the latitudinal gradients of their seasonal characteristics, we also fitted single univariate regressions jointly to enteroviruses and poliomyelitis data adding an indicator variable for the type of case. The joint regression for a characteristic Y (amplitude or timing of the peak) was given by:

$$Y_i = (\alpha + \sigma I_{\text{polio}}) + (\beta + \delta I_{\text{polio}}) \times \text{latitude}_i$$

where I_{polio} takes the value of 1 for polio and 0 for enteroviruses. Therefore, the parameter α is the intercept for enteroviruses, and $\alpha + \sigma$ is the intercept for poliomyelitis. And similarly, β is the slope for enteroviruses, and $\beta + \delta$ is the slope for poliomyelitis. For both characteristics, we tested a model with common slope and different intercept ($\delta = 0$), and a model with different intercept and slope.

S3 Linear mixed-effects models for the intensity of transmission

Our aim was to describe the relationship between the intensity of enterovirus transmission and climatic and demographic factors in order to identify good predictors of the annual seasonality of cases.

S3.1 Data and data exploration

The intensity of enterovirus transmission was estimated for each month t and each state s over the whole period of the study (1983–2013), when "enough" cases were reported (Methods). $R_{s,t}$ is therefore the estimate of the case reproduction number at time t (in months) in state s. As a comparison, we also estimated the case reproduction number for poliomyelitis. The seasonal variation of those estimates in each state is shown in Figure S20, and the time-series of estimates for enterovirus are shown in Figure S21 and Figure S22.

Demographic variables included: population size (Figure S6), population density (Figure S7), annual number of live births (Figure S8), and crude annual birth rates (Figure S9). These variables were given as a value per state per year.

Climatic variables included: temperature (Figure S10), precipitation (Figure S11), dew point (Figure S12), potential evaporation (Figure S13), pressure (Figure S14), relative humidity (Figure S15), and specific humidity (Figure S16). These variables were monthly summaries (extracted from the NARR Monthly Means dataset) and therefore, were given as a value per state per month. A brief description of each variable, including its units, is given in Table S5.

S3.1.1 Collinearity of the explanatory variables

Collinearity between certain pairs of climatic or demographic variables was expected. To explore it, we did pairwise scatterplots and estimated coefficients of correlation for the demographic variables (Figure S17) and the climatic variables (Figure S18) separately. Population size and annual number of births were highly correlated. Among the climatic variables, temperature and dew point, as well as specific humidity and dew point, were highly correlated.

S3.1.2 Relationship between the intensity of transmission and climatic variables

As an exploratory approach to assess the relationship between the intensity of enterovirus transmission ($R_{s,t}$) and the climatic variables we used pairplots (Figure S23 and Figure S24). Based on those, we expected that temperature, dew point, potential evaporation and/or specific humidity had an important role in the statistical analysis.

S3.2 Mixed-effects model

S3.2.1 Model description

We modelled the intensity of enterovirus transmission $R_{s,t}$ with a linear mixed-effects model that included climatic and demographic variables as fixed effects and state as a random effect. The random effect across states was added because the data were nested at the state level, and therefore, we expected observations within a state to be more similar than observations between states, and to account for possible non-observed state-specific effects (i.e. variation across states that was not captured by the fixed effects).

Note that we used a linear model after a logarithmic transformation of the intensity of transmission, $\log_{10}(R_{s,t})$. In a preliminary analysis using the untransformed outcome variable, the residuals showed an increase in spread for larger fitted values, therefore suggesting that the hypothesis of homogeneity was violated. A $\log_{10}$-transformation of $R_{s,t}$ solved this issue.

The models were fitted to data using the R package "nlme" [1].

S3.2.2 Model selection: bottom-up strategy

Because some pairs of explanatory variables showed a high degree of correlation, and including correlated predictors in a model may lead to misleading effects, using a top-down strategy (starting with the full model and dropping a covariable at each step) to find the optimal fixed structure was not a good option. We therefore adopted a bottom-up strategy using AIC and BIC as selection criteria. We started with all possible models with one fixed-effect variable and chose the one with the highest AIC and BIC. At each step, we refitted all the possible models obtained by adding one fixed covariable, and chose the one that made the largest improvement on AIC and BIC. We iterated the process until AIC and/or BIC no longer improved, and all the covariables were significant (i.e. significantly different from 0 at the 5% level).

The models used at this stage were described by the following equation:

$$\log_{10}(R_{s,t}) = (\alpha + a_s) + \beta_1 X_{1,s,t} + \beta_2 X_{2,s,t} + \cdots + \beta_n X_{n,s,t} + \varepsilon_{s,t} \tag{1}$$

where $X_1, \ldots, X_n$ are continuous fixed-effects variables, α is the intercept and the β's are the coefficients of the fixed effects. a_s is the random intercept across states, which was assumed to follow a 0-mean normal distribution, $a_s \sim \text{Normal}(0, \sigma_a^2)$, and $\varepsilon_{s,t}$ are the residuals, which in a first instance were assumed to be independent, $\varepsilon_{s,t} \sim \text{Normal}(0, \sigma_\varepsilon^2)$.

The assumption of independence between consecutive observations could be violated, because we worked with time-series, and therefore we had to account for temporal auto-correlation by adding a residual temporal correlation structure. This was added once we already had the optimal fixed-effects structure, to avoid that the auto-correlated errors accounted for variability that should be accounted for in the fixed effects. We assumed a temporal auto-regressive structure on the residuals, as follows:

$$\varepsilon_{s,t} = \sum_{i=1}^{m} \rho_i \varepsilon_{s,t-i} + \eta_{s,t} \qquad \eta_{s,t} \sim \text{Normal}(0, \sigma_\eta^2) \tag{2}$$

and we tested the significance of increasing the order (m), until the AIC and/or BIC no longer improved. This auto-correlation structure was applied on each state time-series, and $\rho_1, \ldots, \rho_m$ were estimated for all the time-series together (i.e. all modelled time-series had the same estimated ρ's).

S3.3 Results of model selection

Finding the optimal fixed-effects structure. The AIC and BIC values for all the models tested with an increasing number of fixed-effects variables are given in Table S8. The best model with one fixed-effect variable was the dew point, followed by temperature and specific humidity. When adding a second variable, the best model included dew point and potential evaporation. These two variables showed a relatively high correlation (Figure S18), and therefore, we had to be cautious when using those two into the same model. The sign and the size effects (Table S11) did not abruptly change compared to their respective one–fixed-effect models (Table S9 and Table S10), suggesting that misleading effects were not occurring. However, the size of both effects slightly diminished, suggesting some amount of overlapped effects. When adding a third variable, the best model included dew point, potential evaporation and pressure.

Exploring the role of pressure. Pressure markedly varied geographically (as it is strongly inversely correlated to elevation), but in a given state, it did not show a strong seasonal pattern and remained fairly constant over the year (Figure S14). We therefore also considered pressure as a categorical variable by using an indicator variable representing pressure >900 hPa (Table S12). We chose that cutoff based on the data in Figure S14. The AIC and BIC values for the optimal model with pressure treated as a categorical variable were very close to those obtained with pressure treated as a continuous variable (Table S8).

Only observations from two states (the 7 from Colorado [CO] and the 79 from New Mexico [NM]) had values of pressure <900 hPa, meaning that including pressure to the model was only needed to explain those or some of those observations. We therefore tested whether the observations for one of those two states were leading this effect. We found that when removing the 7 observations for CO from the data, adding pressure only decreased the AIC by 4 points, and BIC did not decrease. When removing the 79 observations for NM, adding pressure decreased the AIC by 9 points, and the BIC by 3 points. Therefore, the improvement observed when adding pressure as a third variable was mostly led by the observations from CO, which could perhaps be outliers. However, because the AIC still slightly improved when those observations were removed, we decided to present the results for the whole dataset (i.e. without removing any observations).

Exploring the possible role of demographic variables. Because pressure dit now show a seasonal pattern, it could be that it was capturing the effect of some other non-climatic variable. We then tested whether any demographic variable captured that effect, but none did improve the AIC or BIC (Table S8). The optimal fixed-effects structure was therefore given by dew point, potential evaporation and pressure, that we treated as categorical (Table S12).

Adding a temporal auto-correlation structure. At this stage, the auto-correlation function (ACF) of the normalised residuals by state indicated that there was auto-correlation in the data (Figure S25). When adding the temporal auto-correlation structure defined by equation (2) to the model, AIC and BIC improved for an increasing number of terms up to 12 (AIC=−1988 and BIC=−1886, compared to AIC=−1657 and BIC=−1623 for the model without auto-correlated errors). The estimated values of the coefficients were: $\rho_1 = 0.24, \rho_2 = -0.10, \rho_3 = -0.04, \rho_4 = -0.21, \rho_5 = -0.05, \rho_6 = -0.08, \rho_7 = -0.06, \rho_8 = -0.10, \rho_9 = -0.04, \rho_{10} = -0.03, \rho_{11} = 0.04, \rho_{12} = 0.16$. The ACF of the residuals of the final model (Table S13) suggested that adding the auto-regressive structure removed any temporal auto-correlation in the residuals (Figure S26).

Final model. The estimated regression coefficients, standard errors and p-values of the final model are shown in Table S13. This model had a marginal and conditional R^2 (based on the definitions given in [2]) of $R^2_M = 0.47$ and $R^2_C = 0.55$ respectively, meaning that it explained a relatively large amount of the variance of $R_{s,t}$ and that the random effects captured only a small amount of it (the difference between R^2_C and R^2_M was small).

Best univariable model. Because the best model with only one fixed effect was dew point, it was likely that a large part of the variability in the intensity of transmission of the final model was captured by the effect of the dew point alone. The model with only dew point as fixed effects after accounting for auto-regressive errors had $R^2_M = 0.23$ and $R^2_C = 0.29$, indicating that dew point alone was an important predictor of the intensity of enterovirus transmission. The estimated regression coefficients, standard errors and p-values of this model are shown in Table S14.

S3.4 Model validation

Diagnostic plots. Visual exploration of the normalised residuals of the final model (obtained with restricted maximum likelihood [REML] estimation) suggested that they followed a normal distribution (Figure S27A). When the residuals were plotted against fitted values (Figure S27B), the spread looked regular, suggesting that the hypothesis of homogeneity was not violated, and when they were plotted against each explanatory variable (Figure S27C–E), they did not show any pattern that could suggest violation of the independence assumption.

Spatial auto-correlation. Spatial auto-correlation in spatial data is a common feature, and therefore, it was expected. Not surprisingly, the estimated random intercepts exhibited a latitudinal gradient (Figure S28), indicating the existence of an un-measured variable with a spatial structure that determined $R_{s,t}$ and that was not accounted for in the fixed effects. A bubbled plot of the residuals per state did not show a spatial pattern or clustering of positive or negative residuals (Figure S29), thus suggesting that the random effect across states removed any spatial auto-correlation in the residuals, thereby making the use of explicit spatial regression models unnecessary.

S3.5 Model selection: other considerations

As a test, we also performed model selection with a top-down strategy via hypothesis testing (using the likelihood ratio test) to find the best model. Although this implied starting from a full model that included highly correlated covariables, we obtained the same final model that we obtained with the bottom-up procedure, which supports the results of model selection.

S4 Supplementary Figures

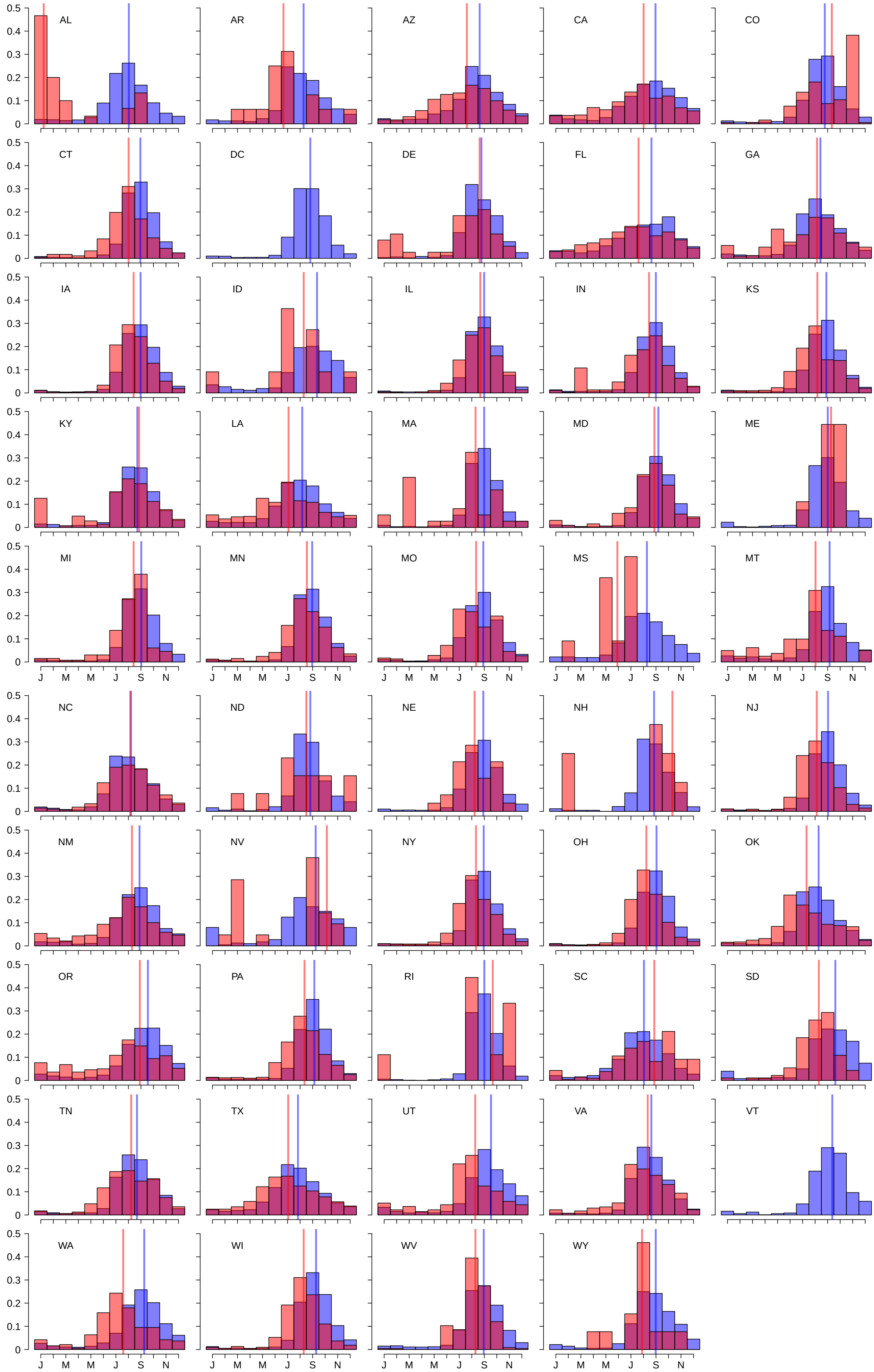


Figure S1: Distribution of non-polio enterovirus (red) and poliomyelitis (blue) cases per state throughout the year. The vertical lines indicate the estimated mean timing of cases.

Figure S2: Latitudinal gradients in the seasonal pattern of enterovirus (red) and poliomyelitis (blue) cases as in Figure 3 of the main text with the name of the states.

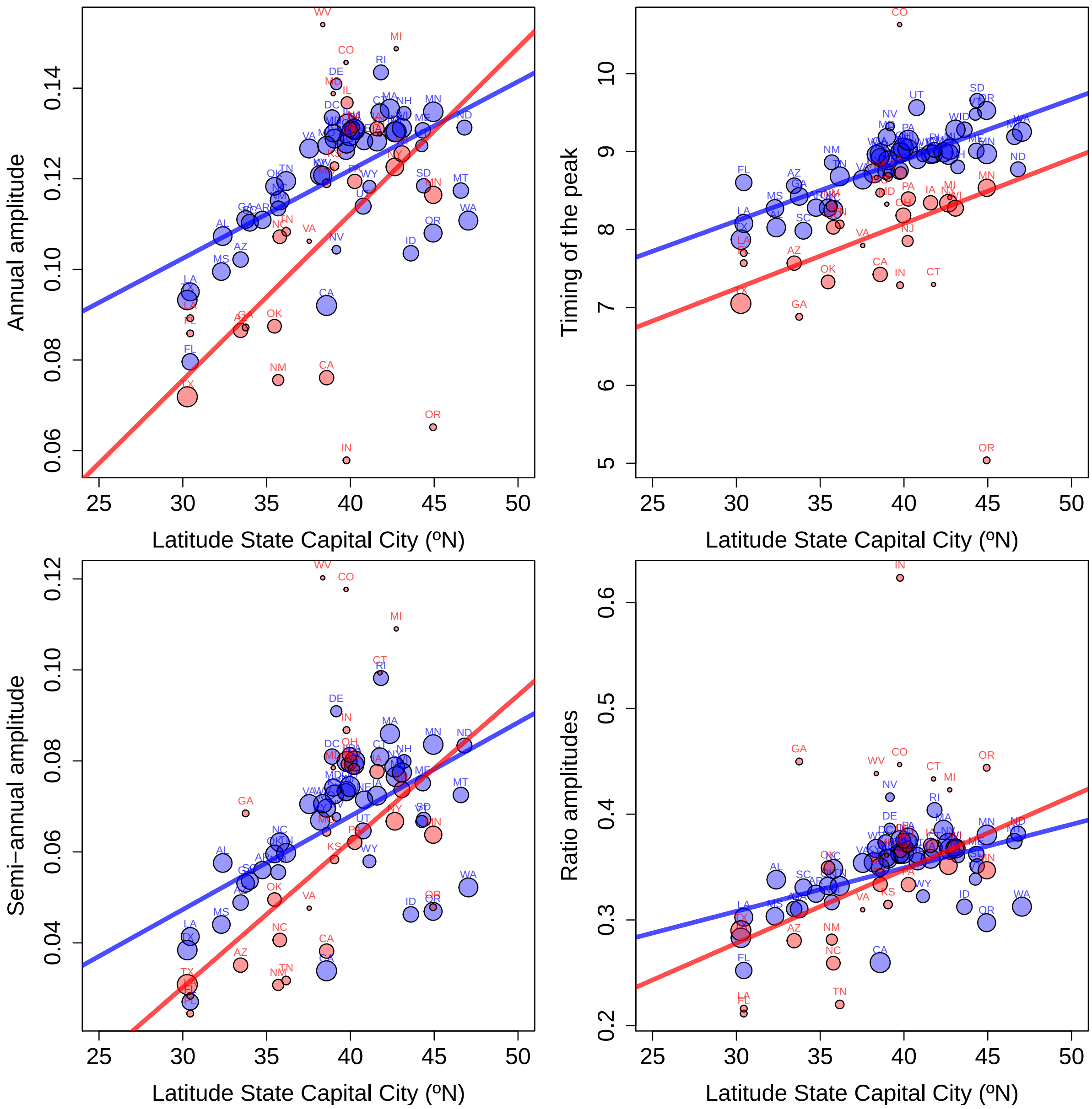


Figure S3: Sensitivity analysis for the latitudinal gradients. Latitudinal gradients obtained using the subset of states with data for both, enterovirus and poliomyelitis. The white circles indicate that those states have not been accounted for in the regression.

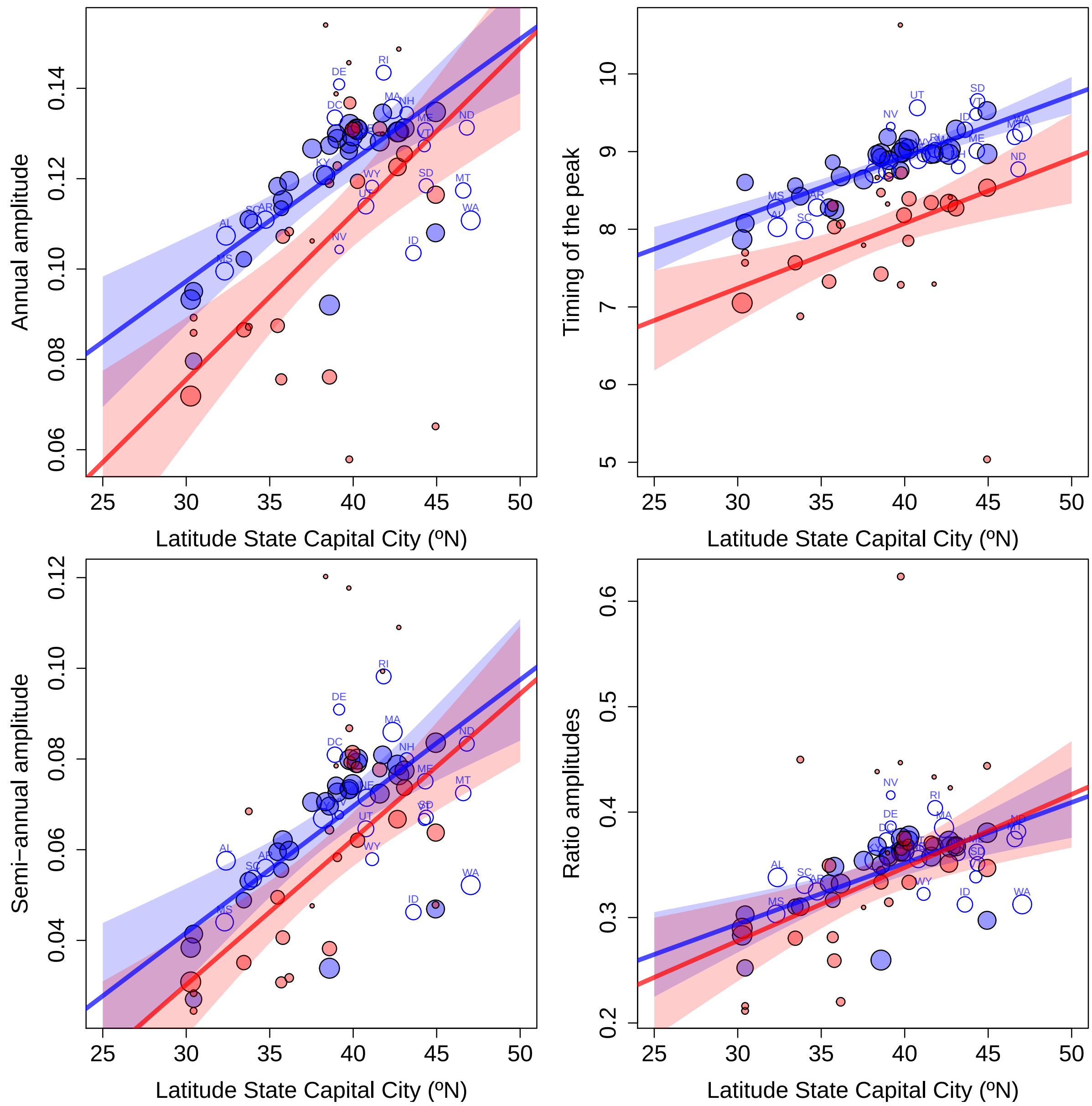


Figure S4: Sensitivity analysis for the latitudinal gradients in the seasonal pattern of enterovirus (red) and poliomyelitis (blue) cases. Latitudinal gradients obtained using the latitude of the state center of population, instead of the latitude of the state capital city.

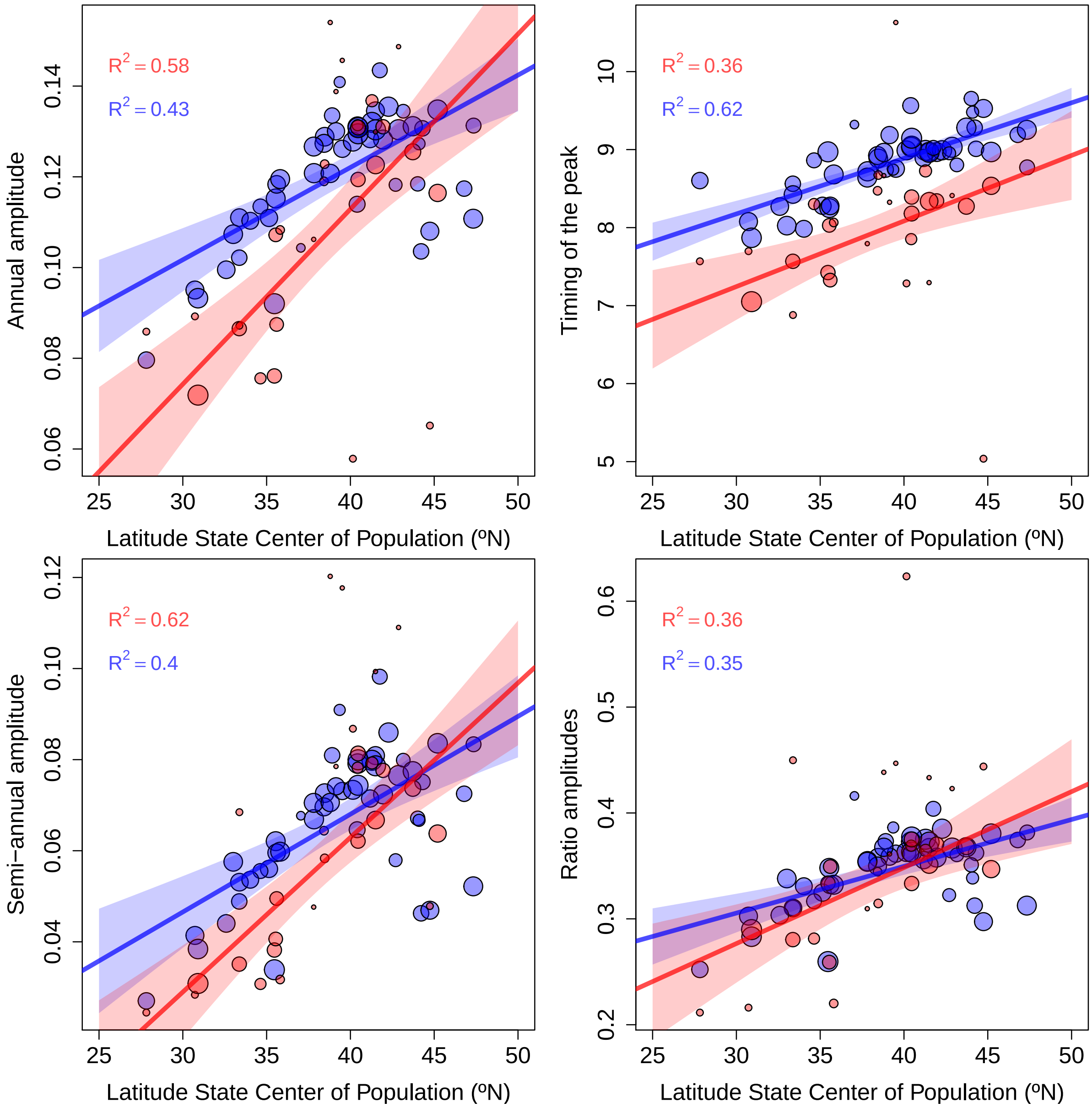


Figure S5: Sensitivity analysis for the latitudinal gradients in the mean timing of cases (top) and standard deviation on the timing of cases (bottom). Latitudinal gradients obtained using the latitude of the state center of population, instead of the latitude of the state capital city.

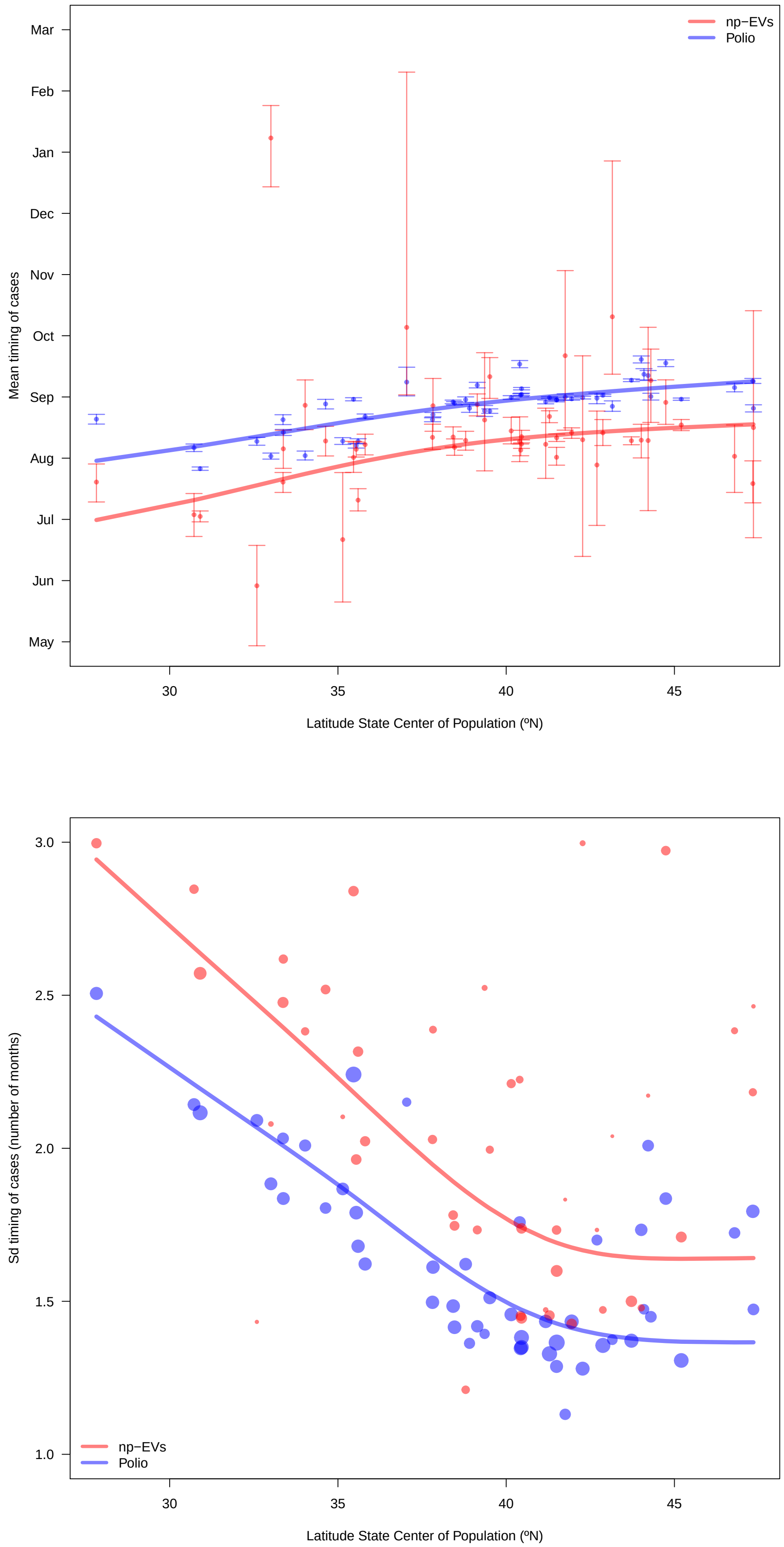


Figure S6: Population size (log10) by state from 1983 to 2013.

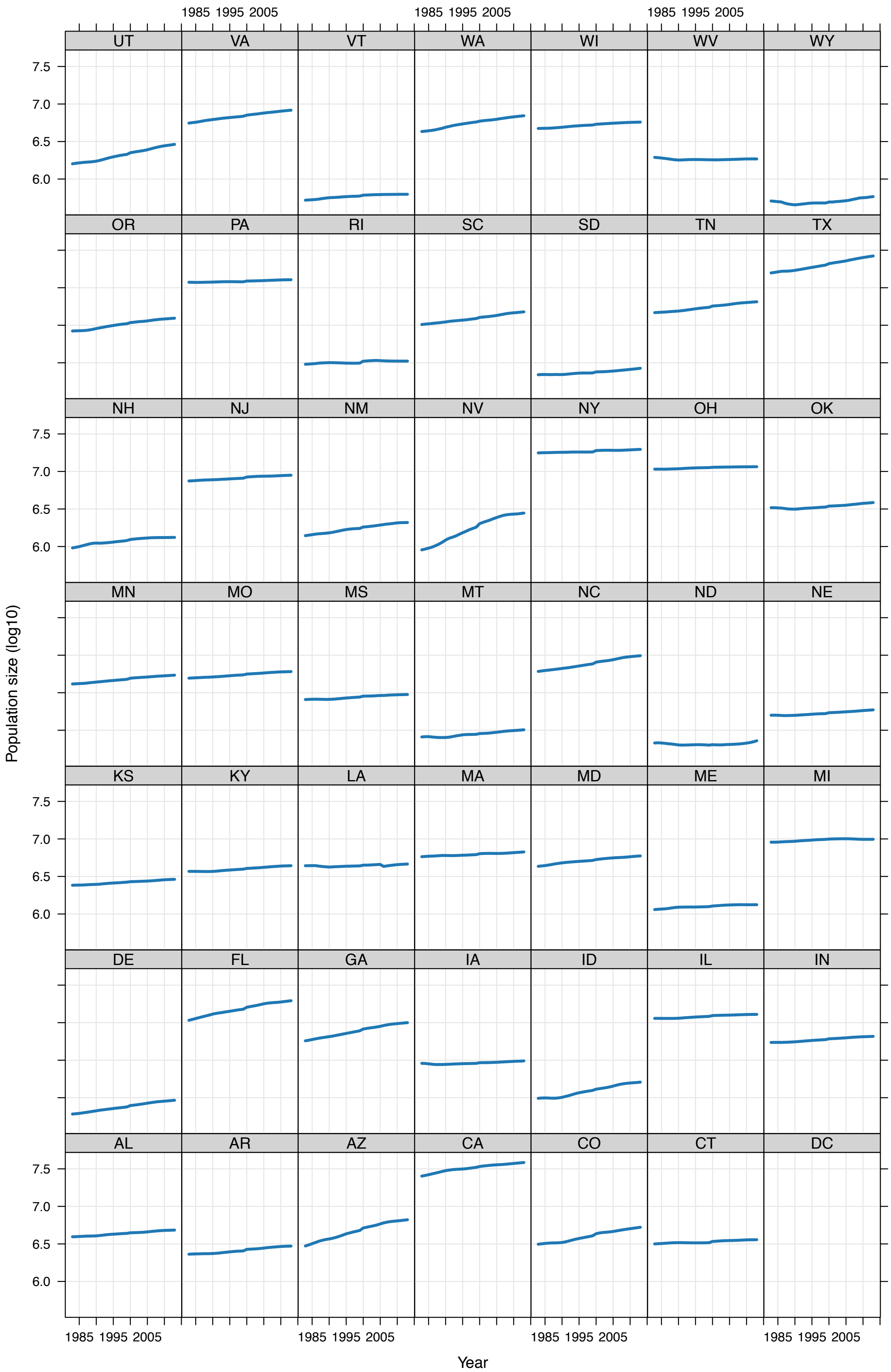


Figure S7: Population density (log10) by state from 1983 to 2013.

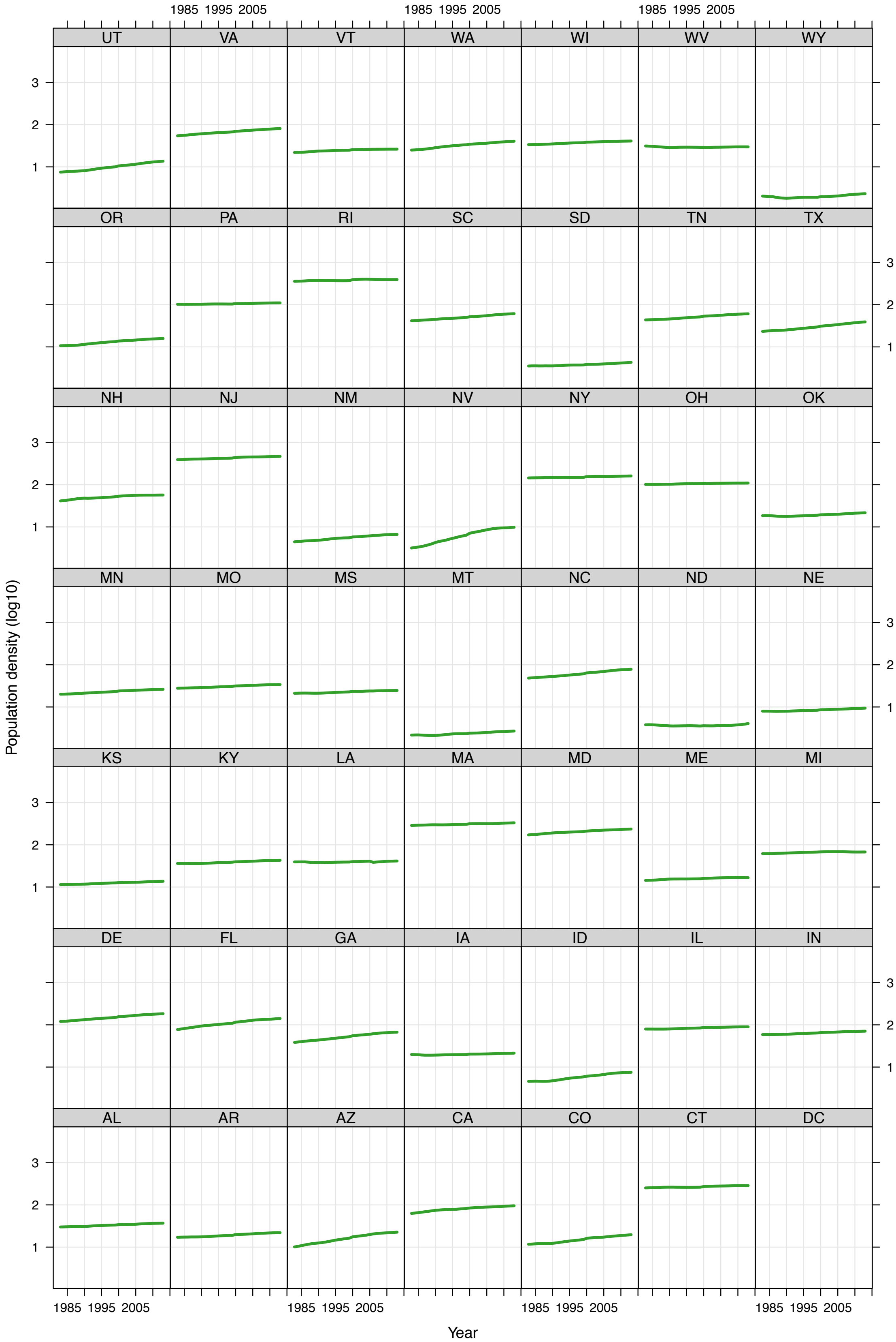


Figure S8: Number of live births (log10) by state from 1983 to 2013.

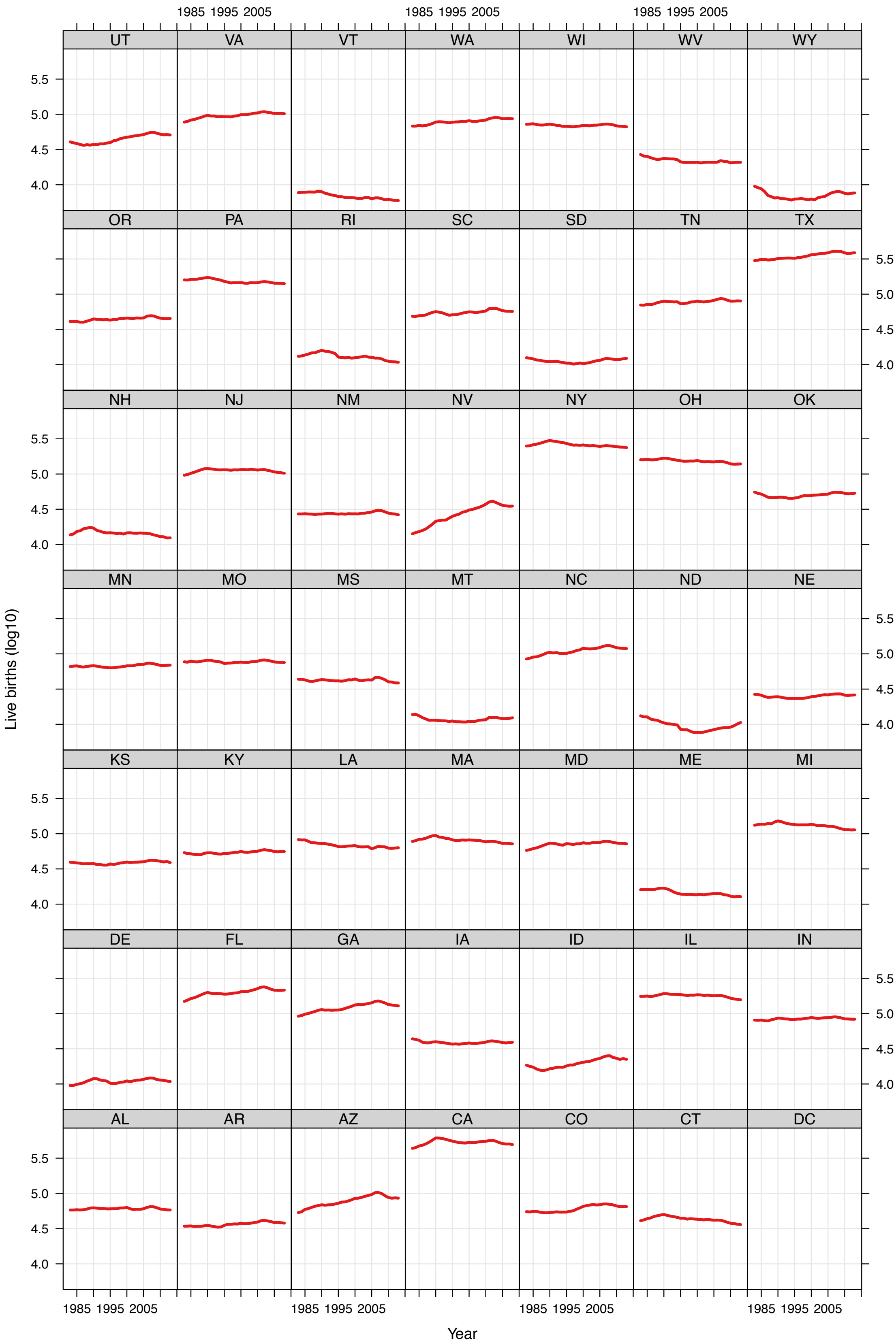


Figure S9: Crude birth rates (per 1000) by state from 1983 to 2013.

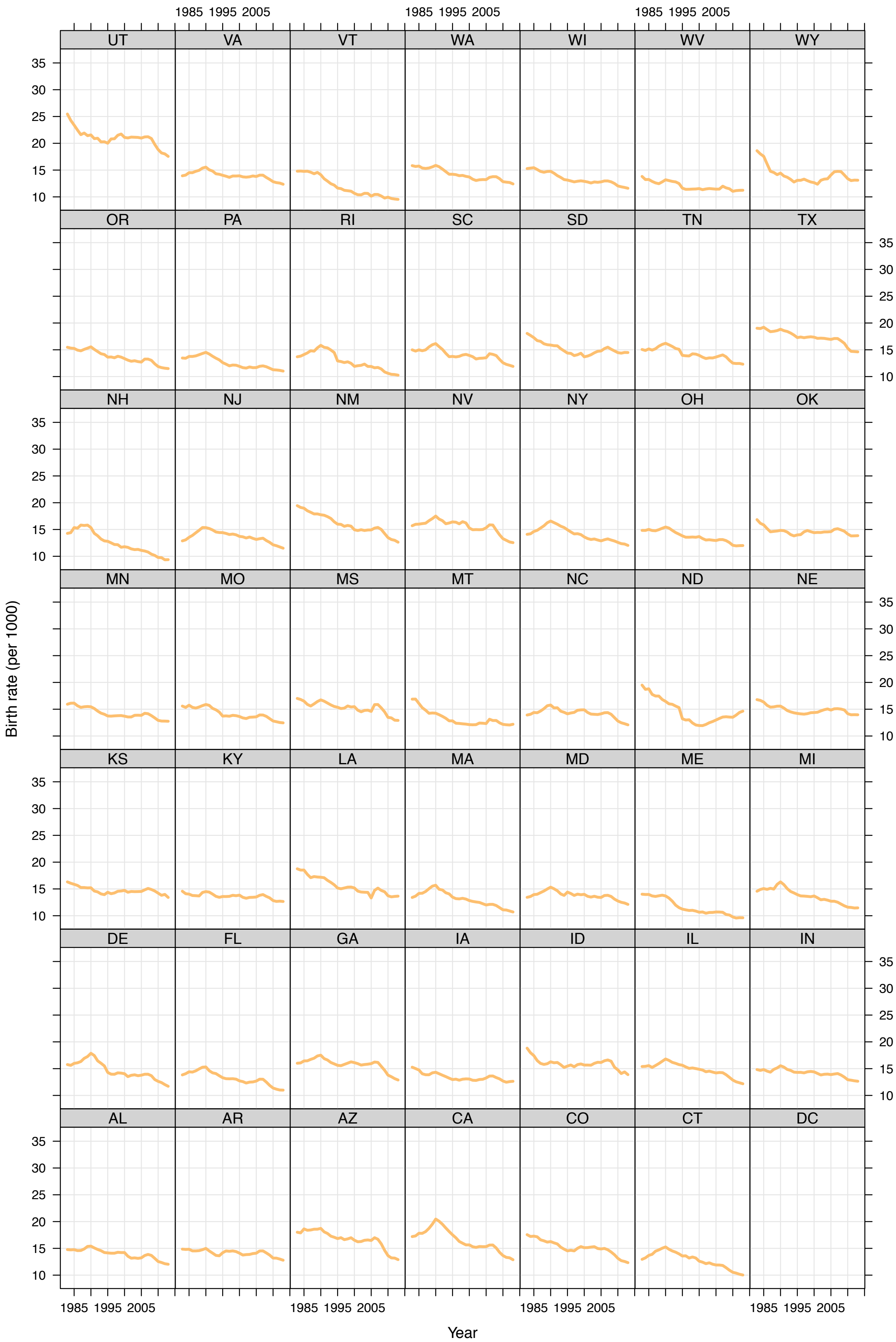


Figure S10: Annual variation of monthly estimates of air temperature at 2m (°C) for 1983–2013.

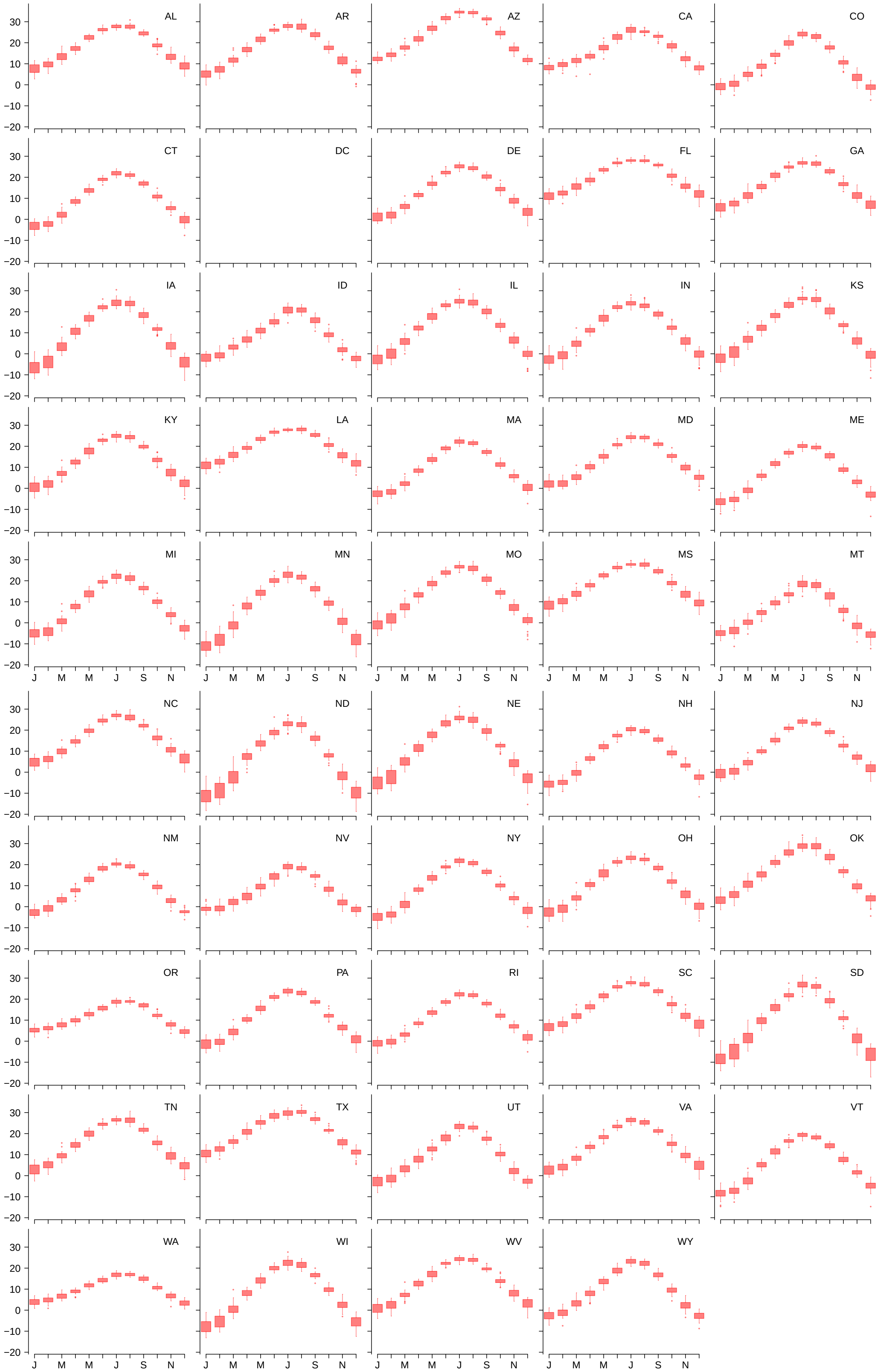


Figure S11: Annual variation of monthly estimates of accumulated precipitable water at surface (kg/m^2) for 1983–2013.

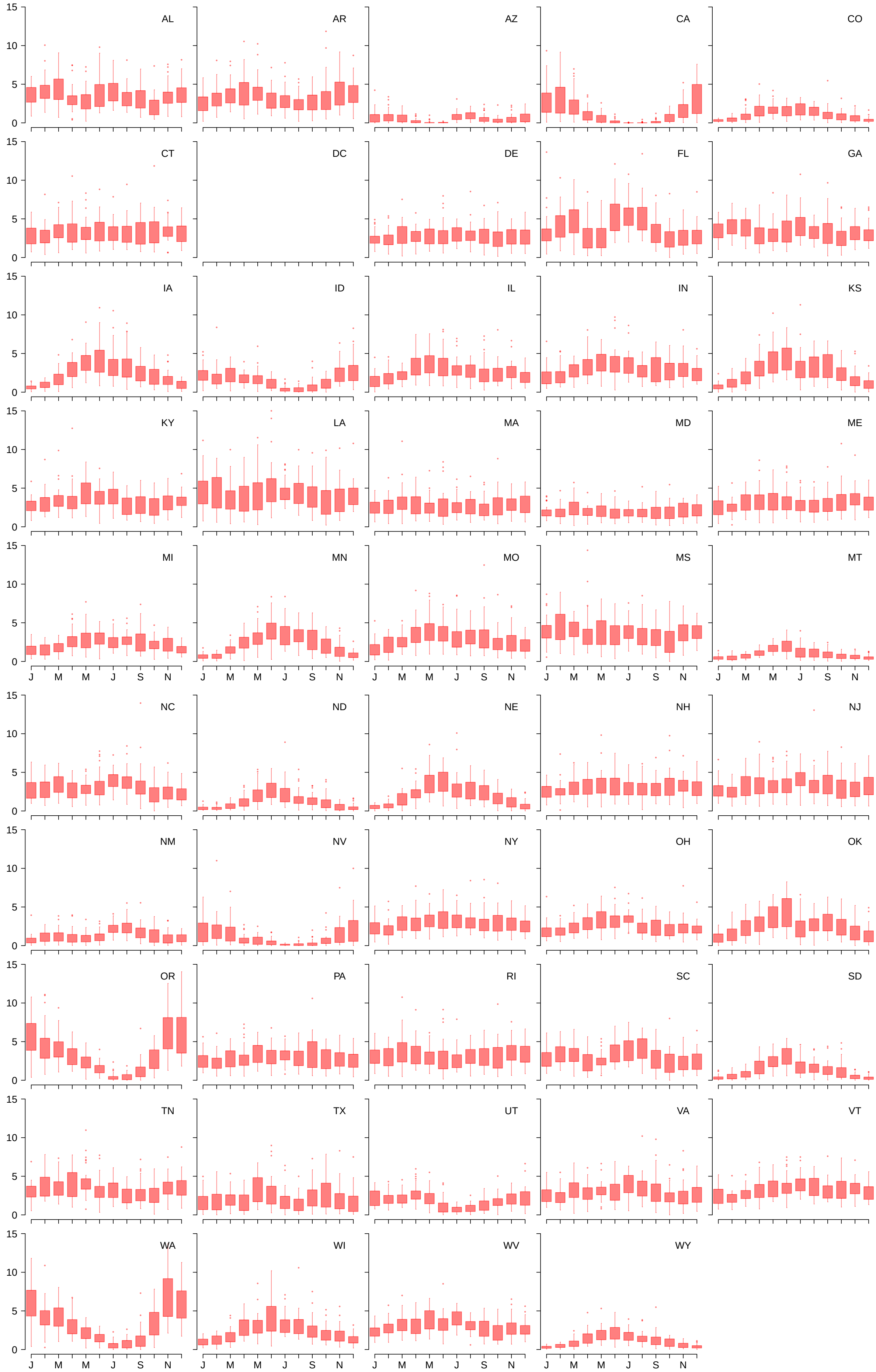


Figure S12: Annual variation of monthly estimates of dew point temperature (°C) for 1983–2013.

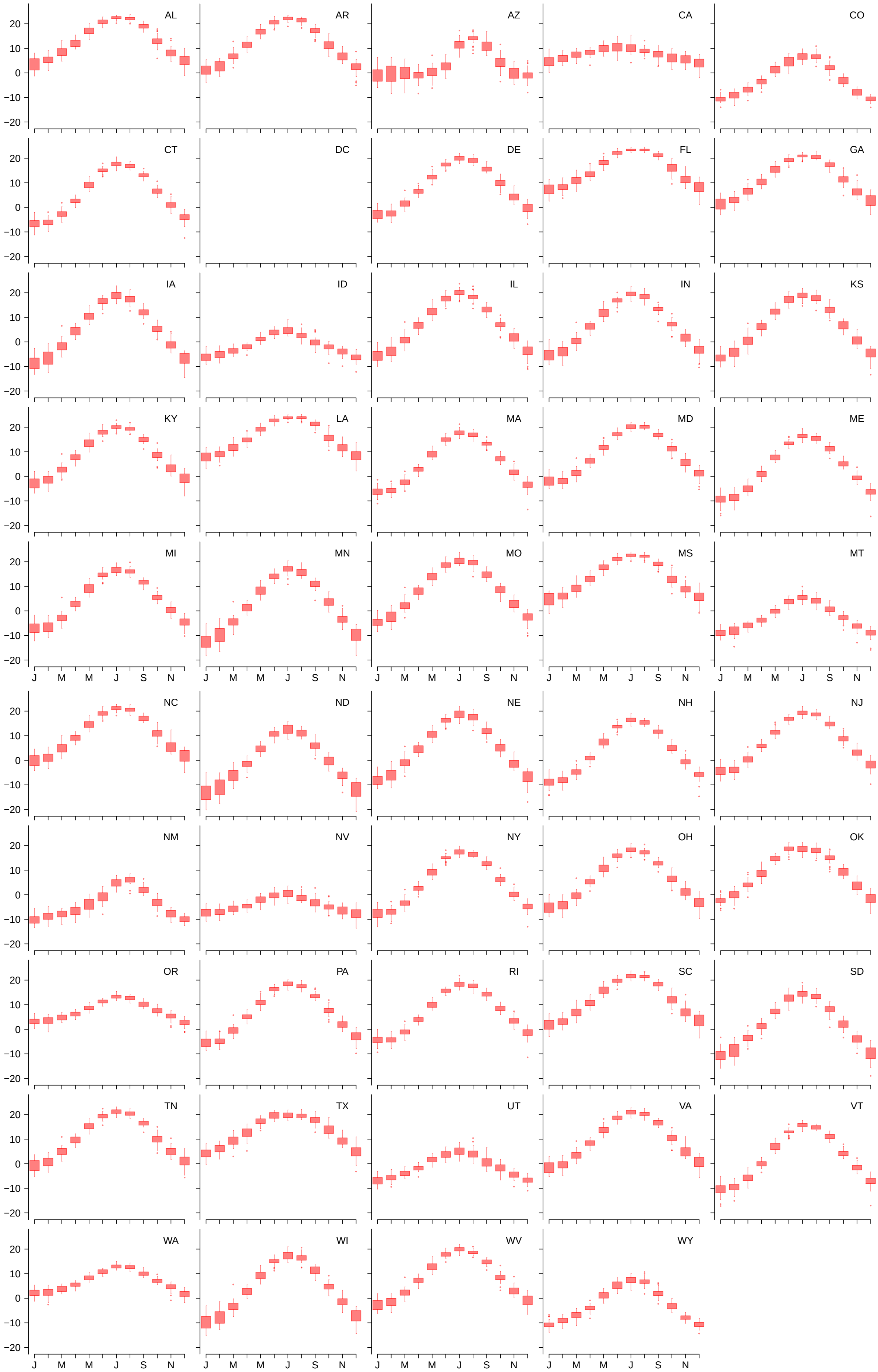


Figure S13: Annual variation of monthly estimates of potential evaporation (kg/m^2) for 1983–2013.

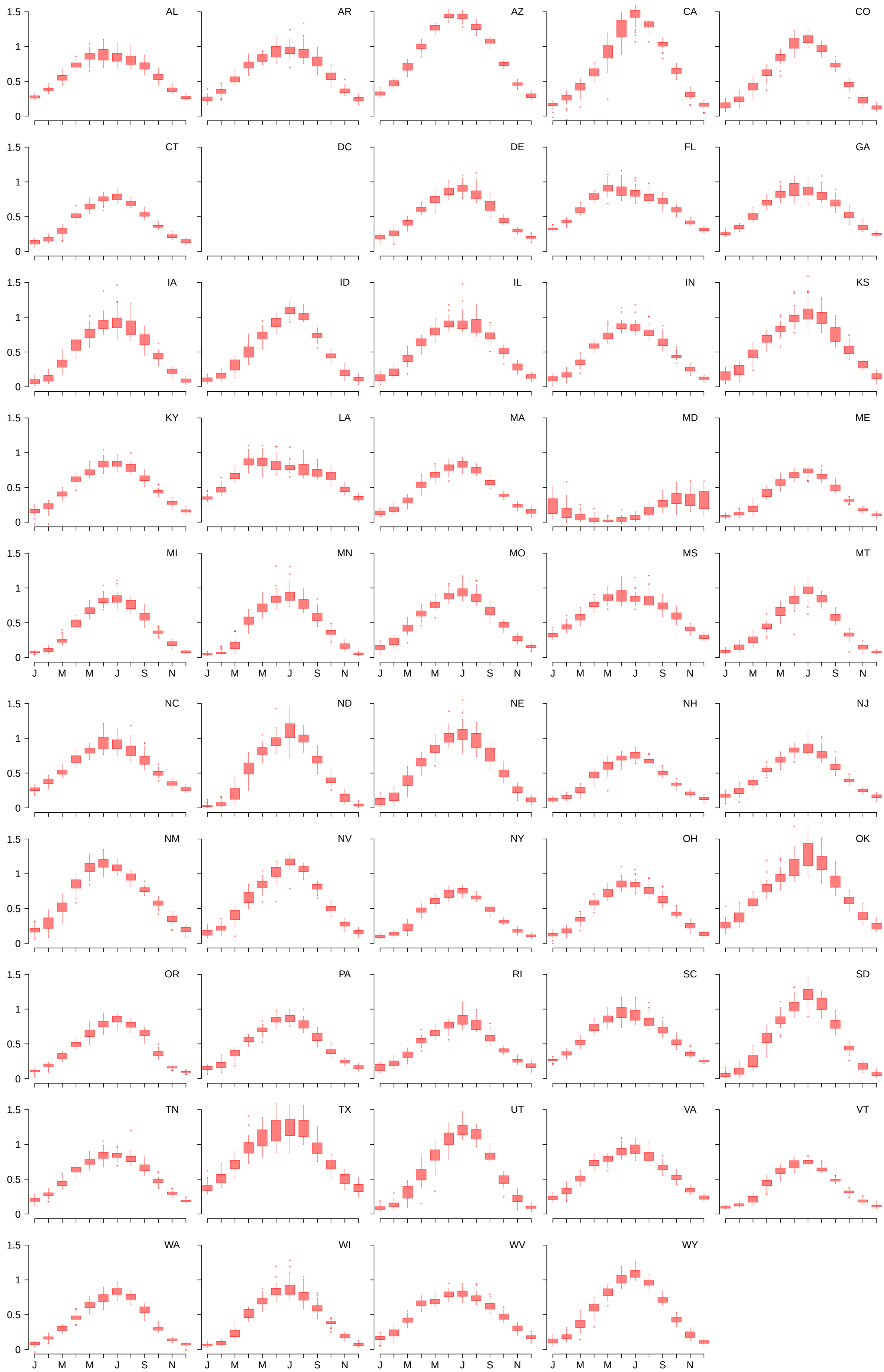


Figure S14: Annual variation of monthly estimates of vapor pressure (hPa) for 1983–2013.

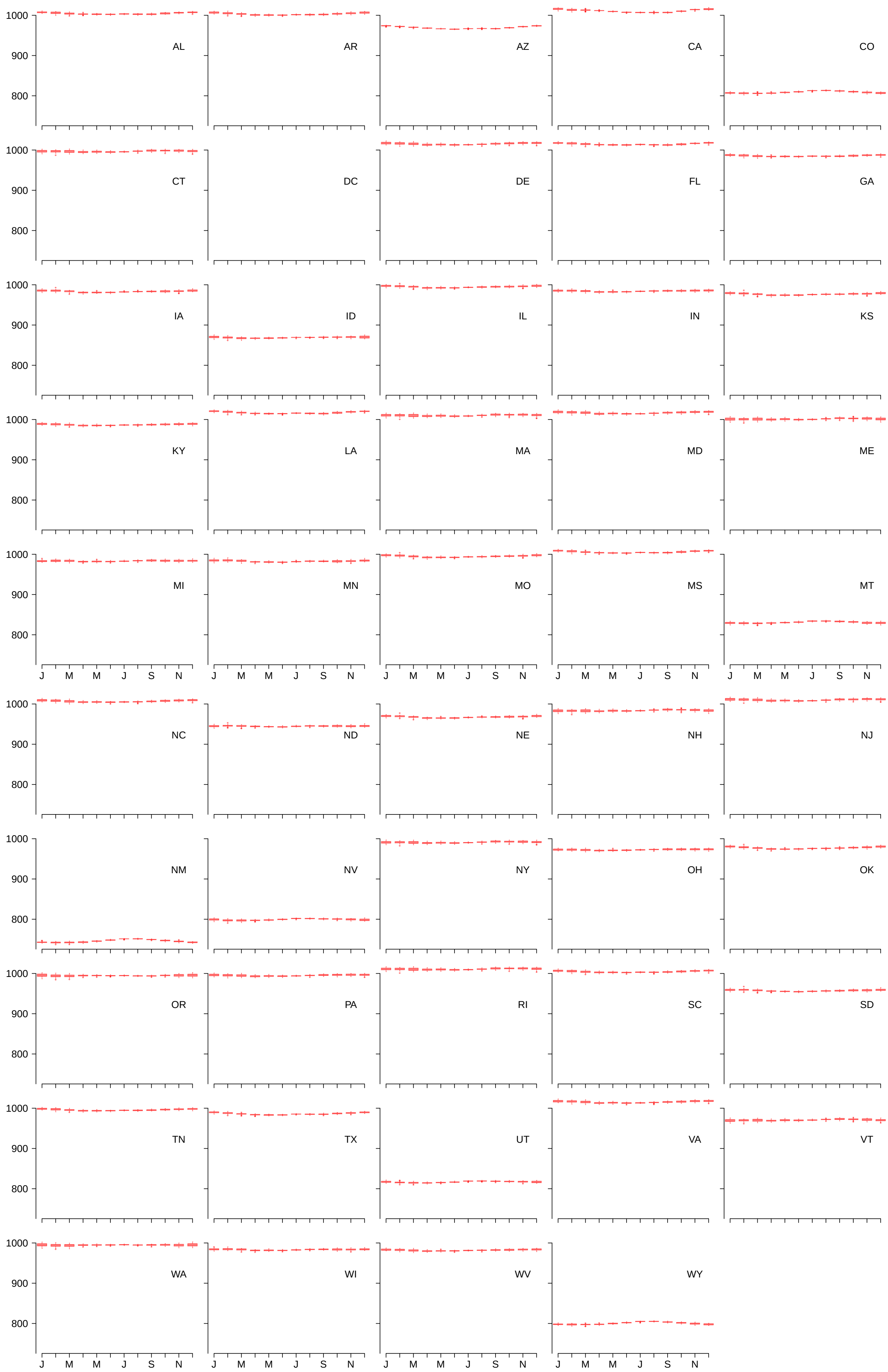


Figure S15: Annual variation of monthly estimates of relative humidity at 2m (%) for 1983–2013.

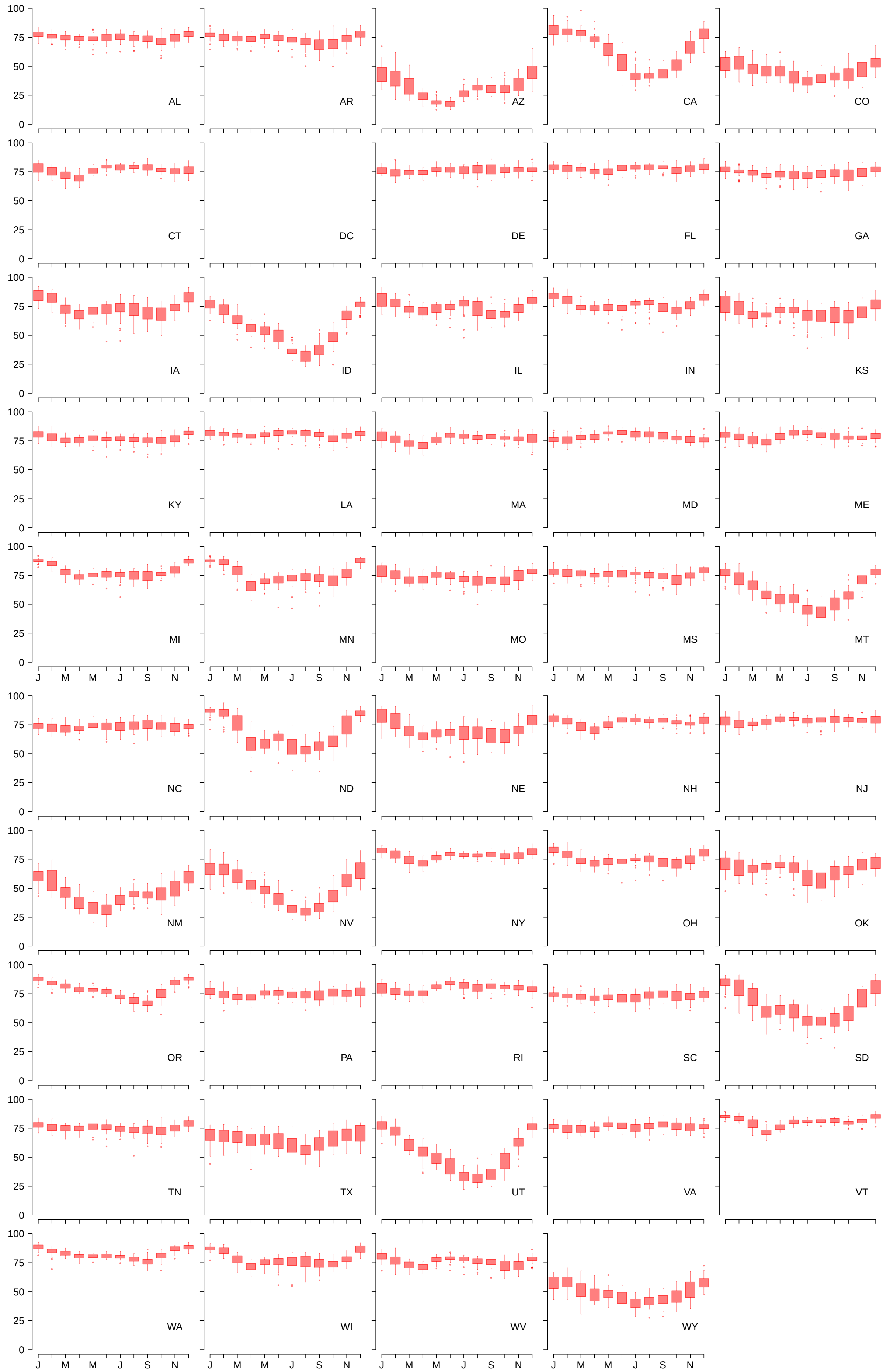


Figure S16: Annual variation of monthly estimates of specific humidity at 2m (g/kg) for 1983–2013.

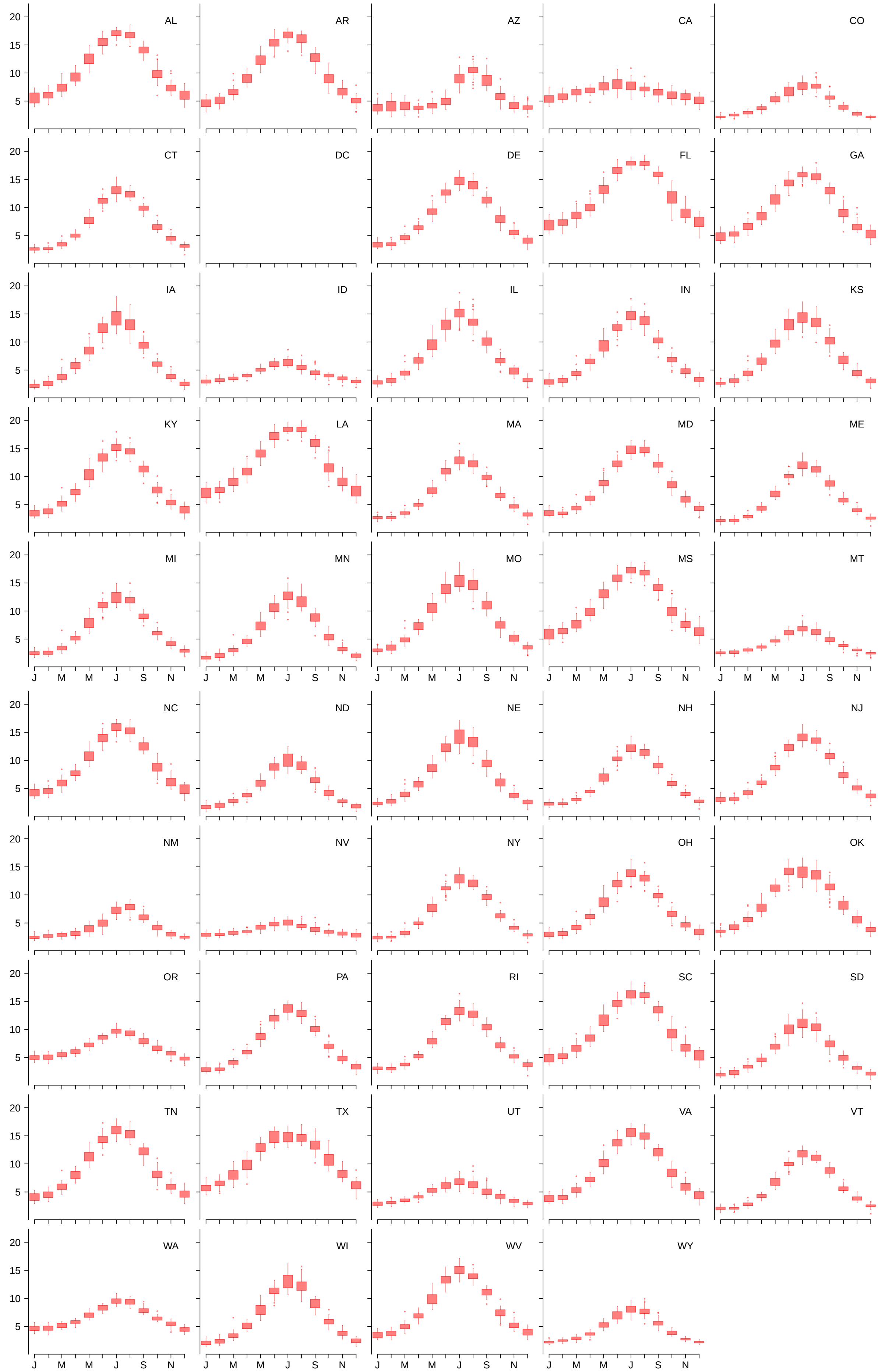


Figure S17: Collinearity between demographic variables.

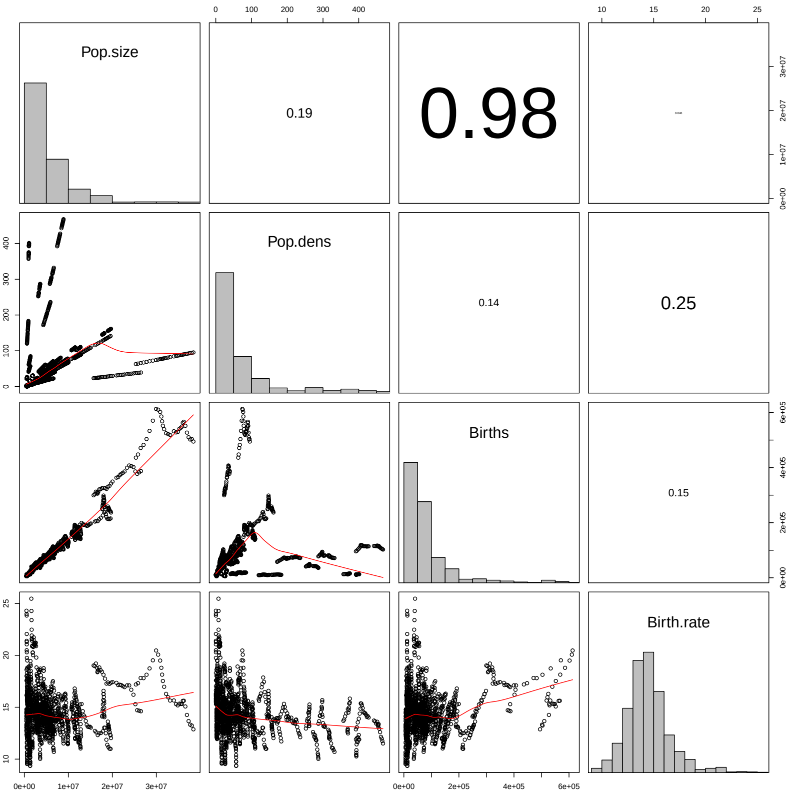


Figure S18: Collinearity between climatic variables.

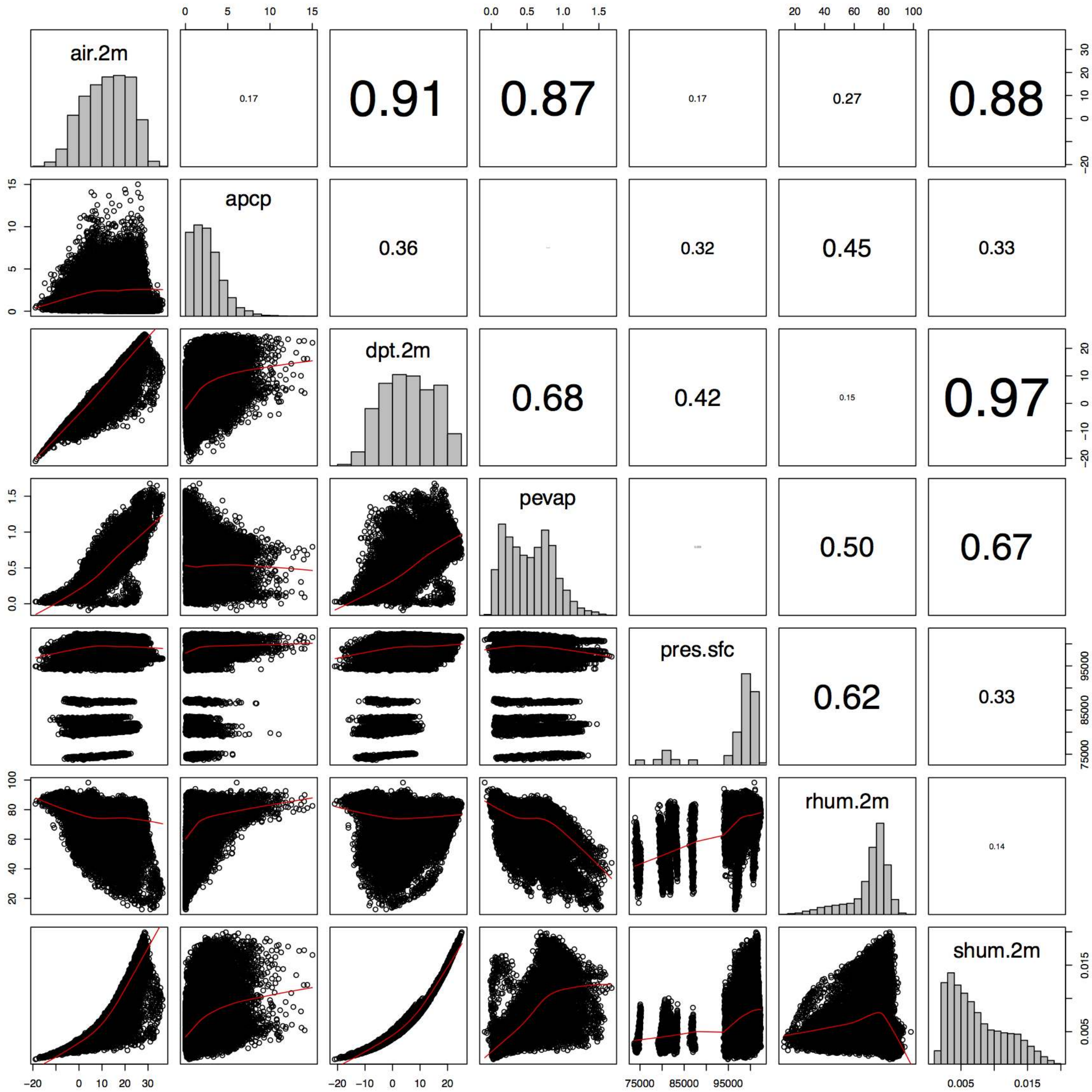


Figure S19: Linear models for the timing of the peak (top) and the amplitude (bottom) of enterovirus cases for the best single climatic (left) and demographic (right) predictors.

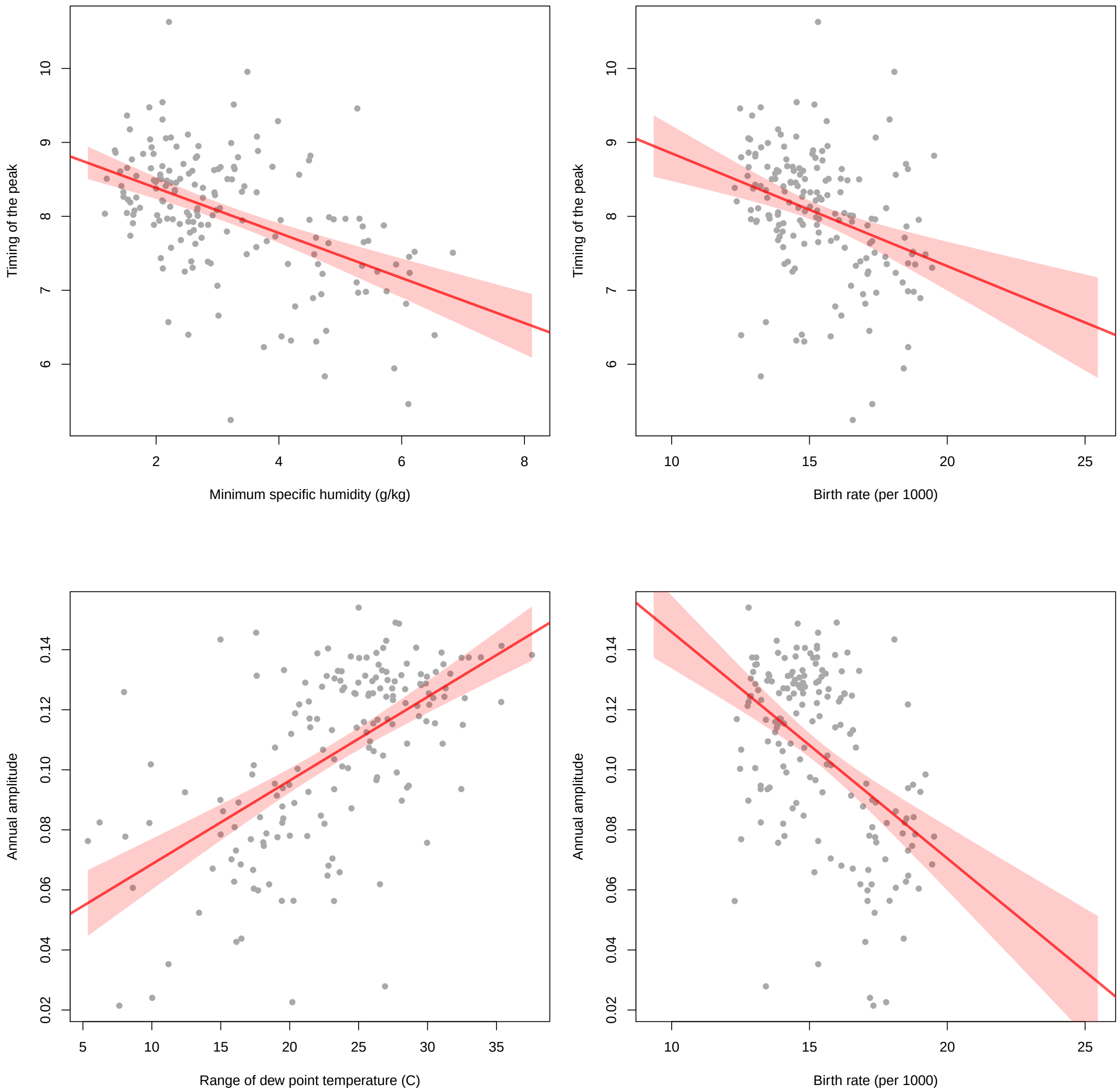


Figure S20: Annual variation of the monthly estimates of the case reproduction number ($\log_{10}(R_{s,t})$) for enterovirus (red) and poliomyelitis (blue) cases by state for the periods 1983–2013 (enterovirus) and 1931–1954 (poliomyelitis). The gray horizontal line indicates $R_{s,t} = 1$.

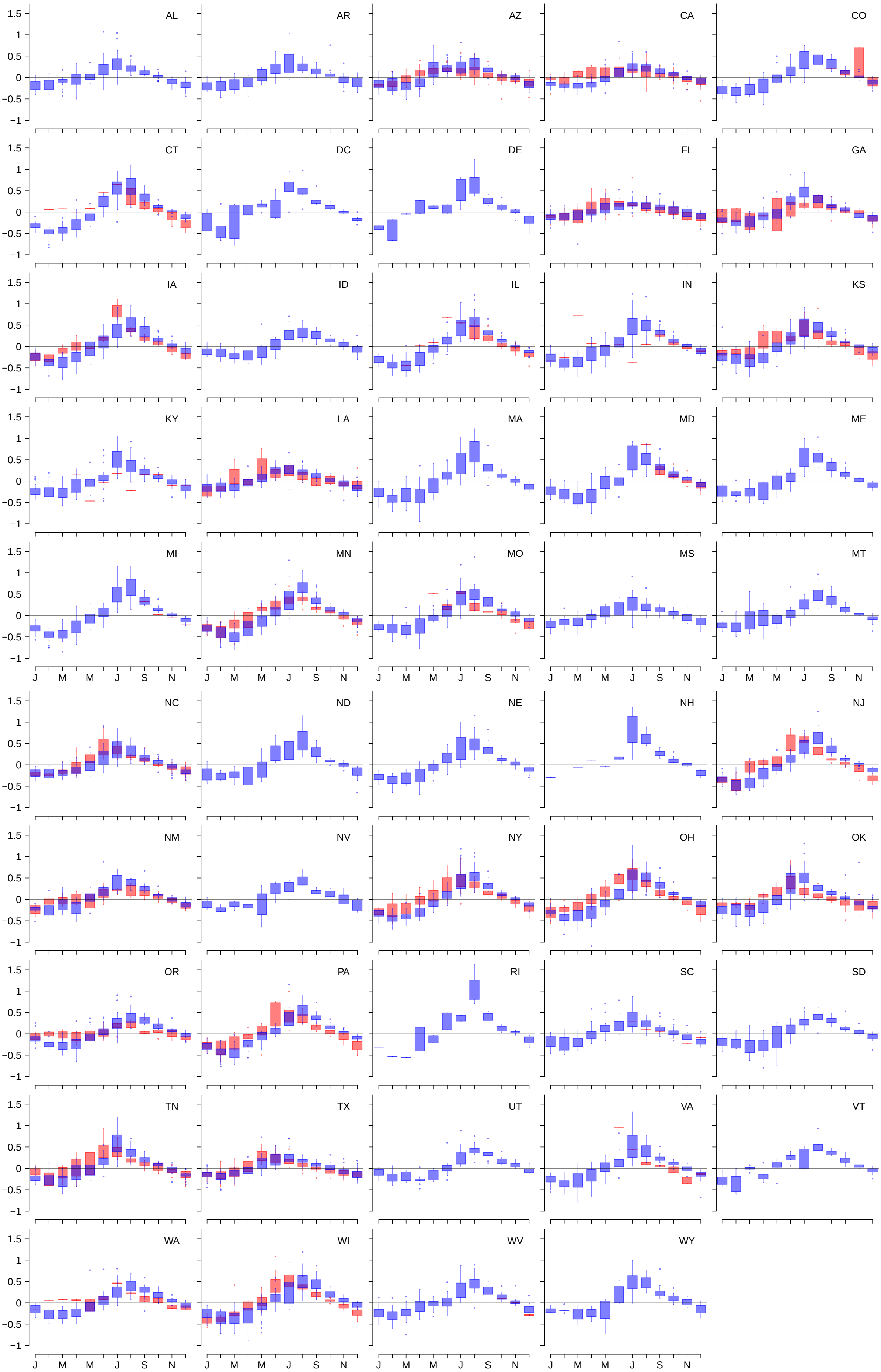


Figure S21: Time-series of the estimates of the case reproduction number ($\log_{10}(R_{s,t})$) for enterovirus by state.

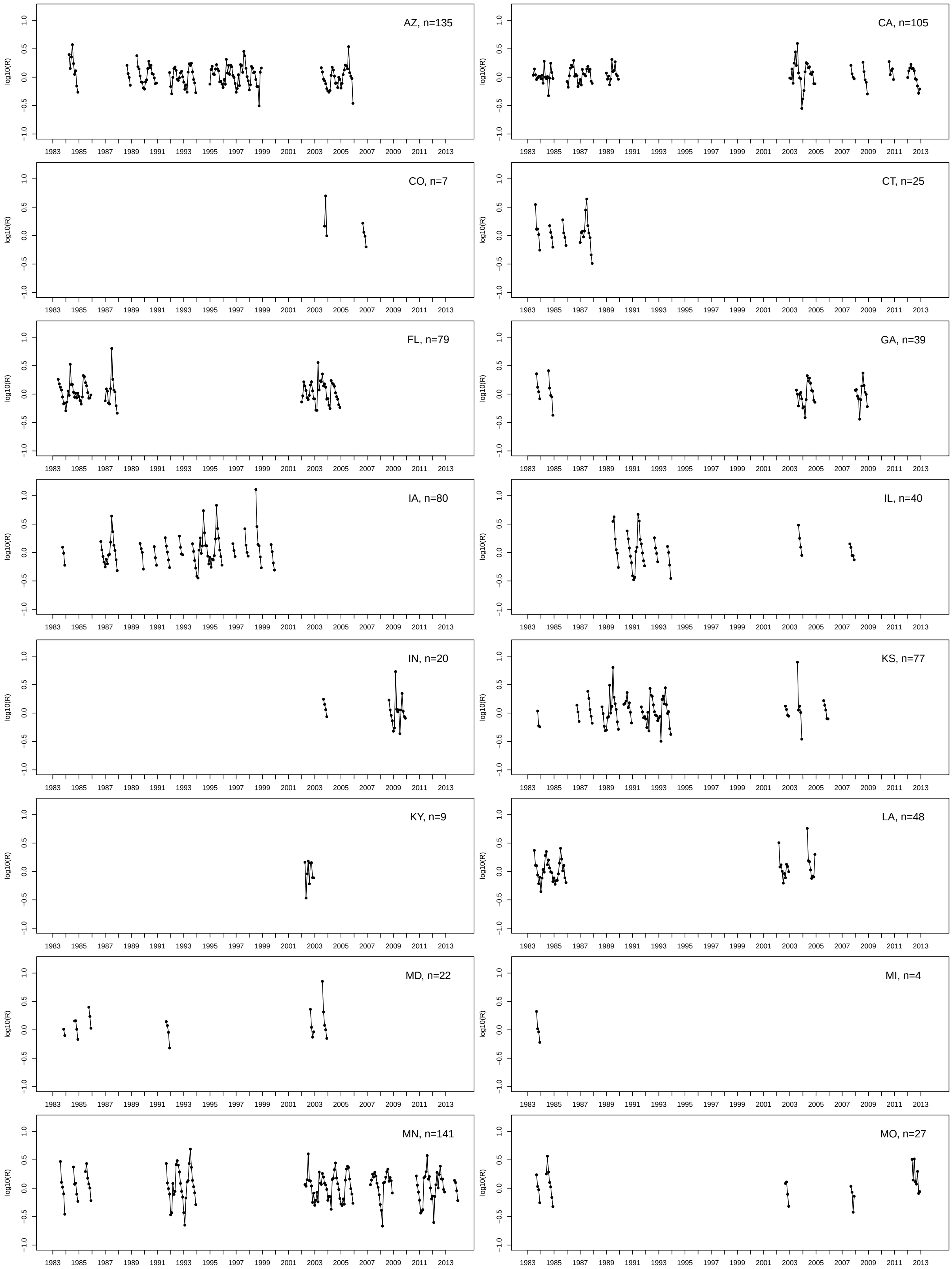


Figure S22: Time-series of the estimates of the case reproduction number ($\log_{10}(R_{s,t})$) for enterovirus by state (cont.).

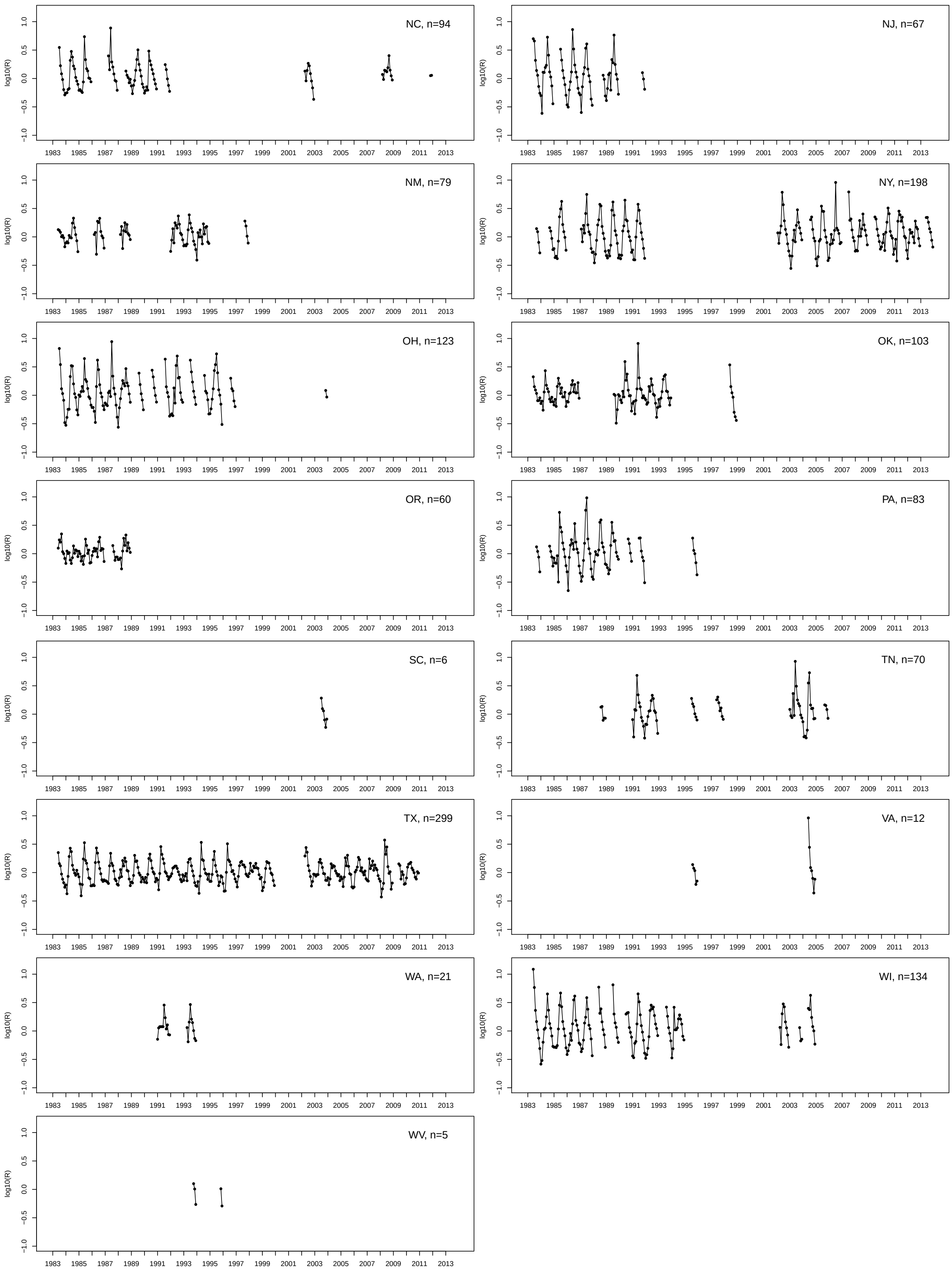


Figure S23: Pairplots of the estimates of the case reproduction number of enterovirus ($\log_{10}(R_{s,t})$) against monthly estimates of the climatic variables. Each color corresponds to a state. The lines are cubic splines per state. The dashed horizontal line indicates $R_{s,t} = 1$. On the left (right), data for the 8 (18) states with the highest number of point estimates of $R_{s,t}$.

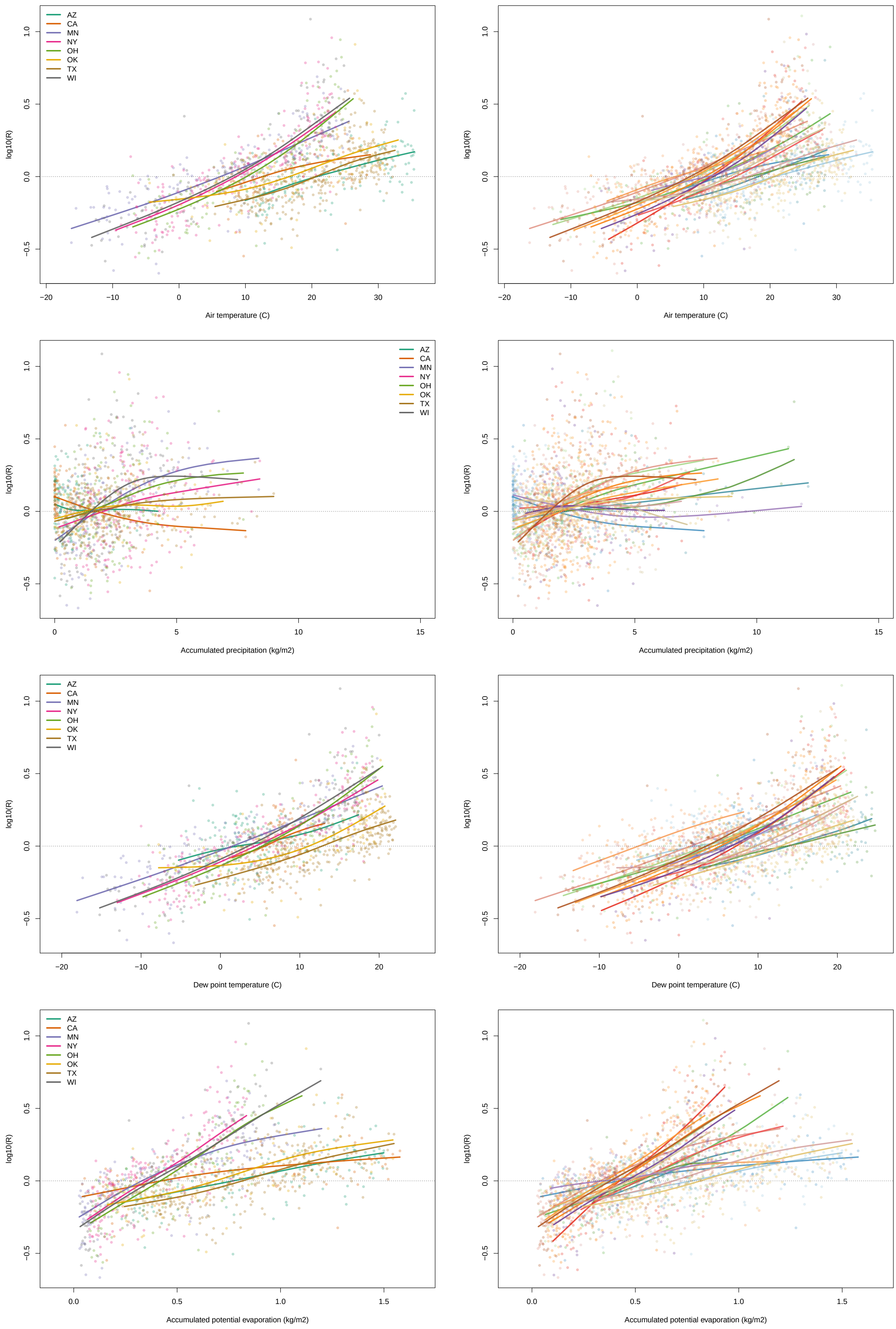


Figure S24: Pairplots of the estimates of the case reproduction number of enterovirus ($\log_{10}(R_{s,t})$) against monthly estimates of the climatic variables (cont. from Figure S23). Each color corresponds to a state. The lines are cubic splines per state. The dashed horizontal line indicates $R_{s,t} = 1$. On the left (right), data for the 8 (18) states with the highest number of point estimates of $R_{s,t}$.

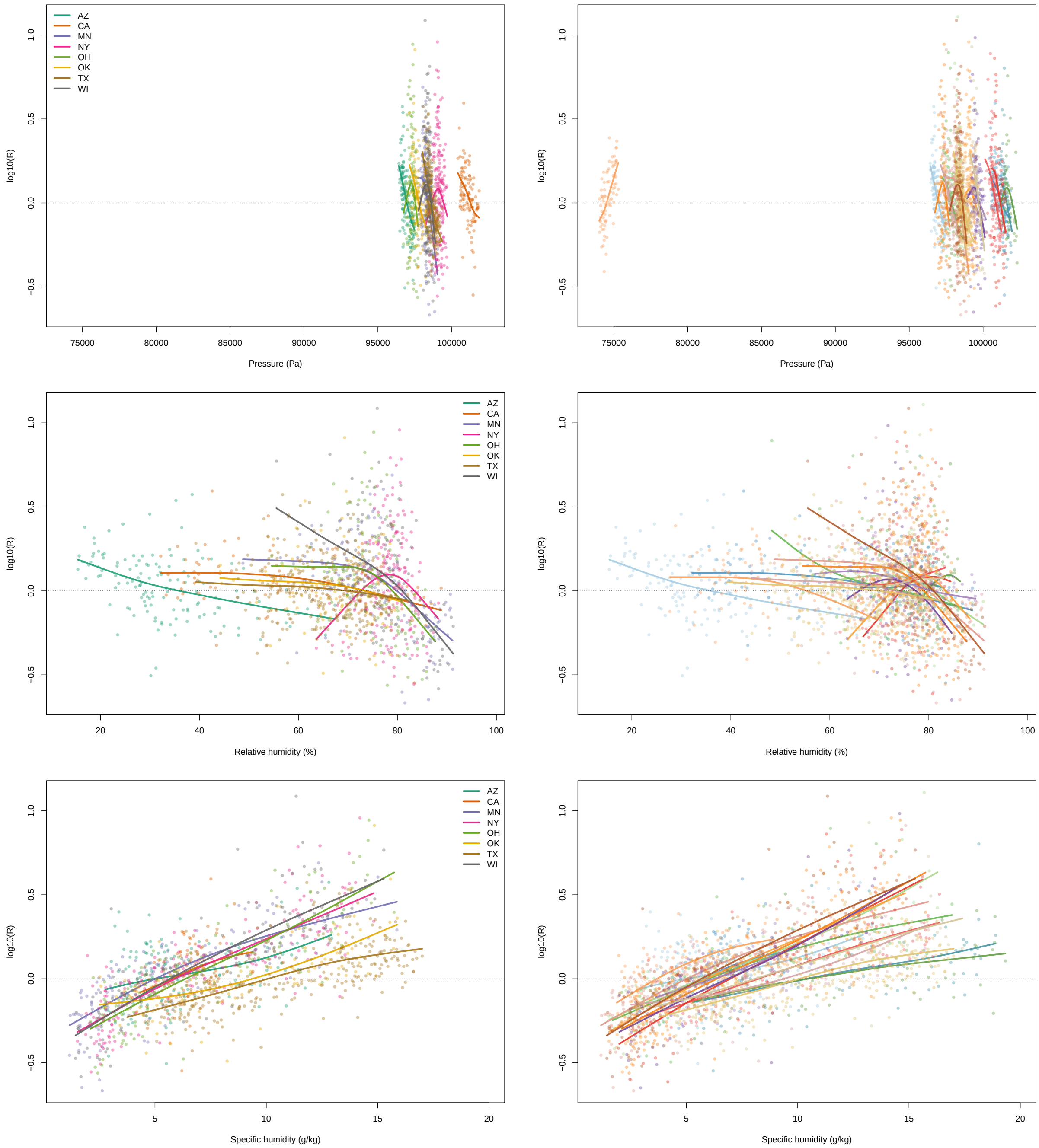


Figure S25: ACF of the normalised residuals for the optimal model without auto-correlated errors estimated with REML.

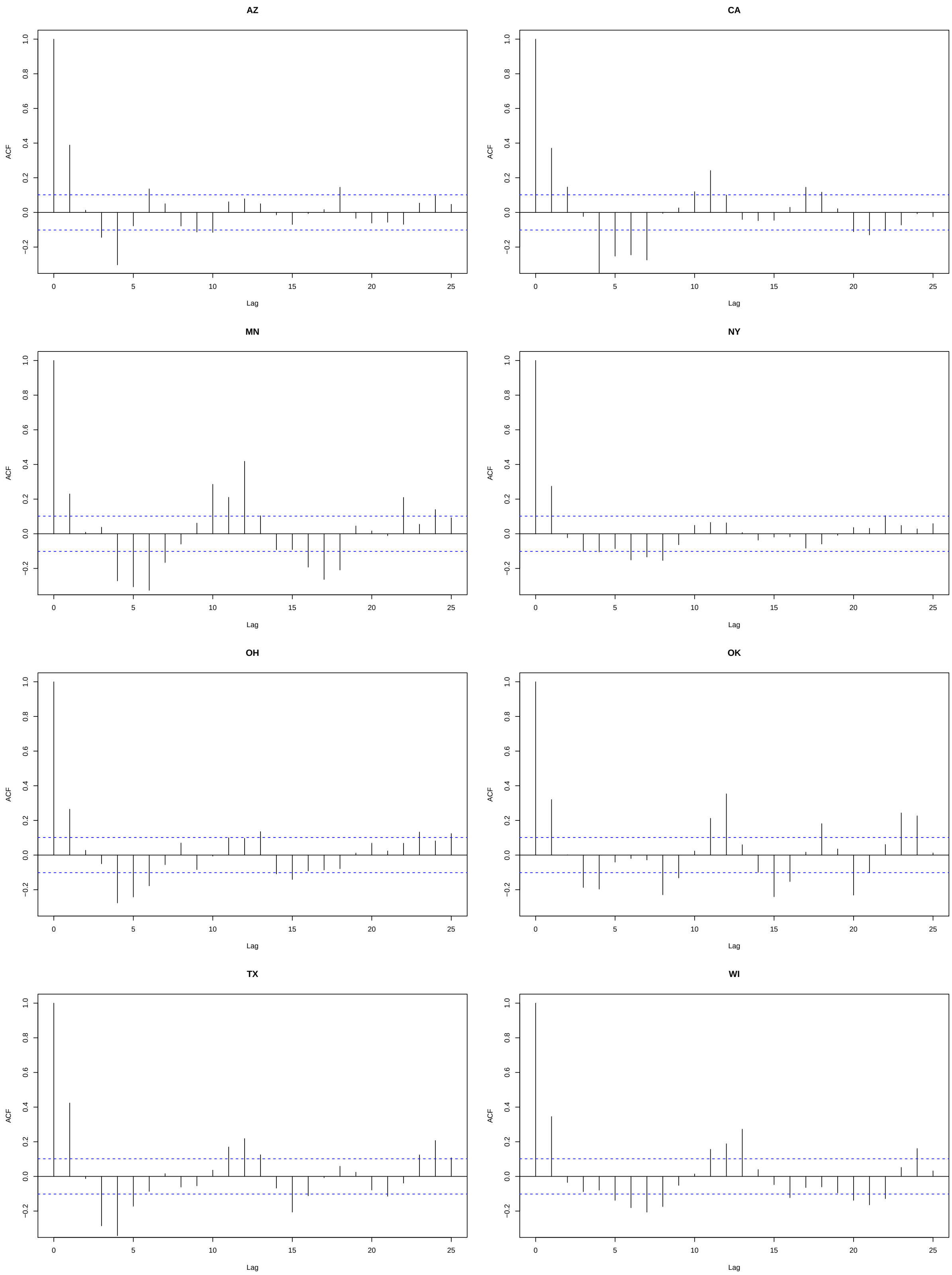


Figure S26: ACF of the normalised residuals for the optimal model with auto-correlated errors estimated with REML.

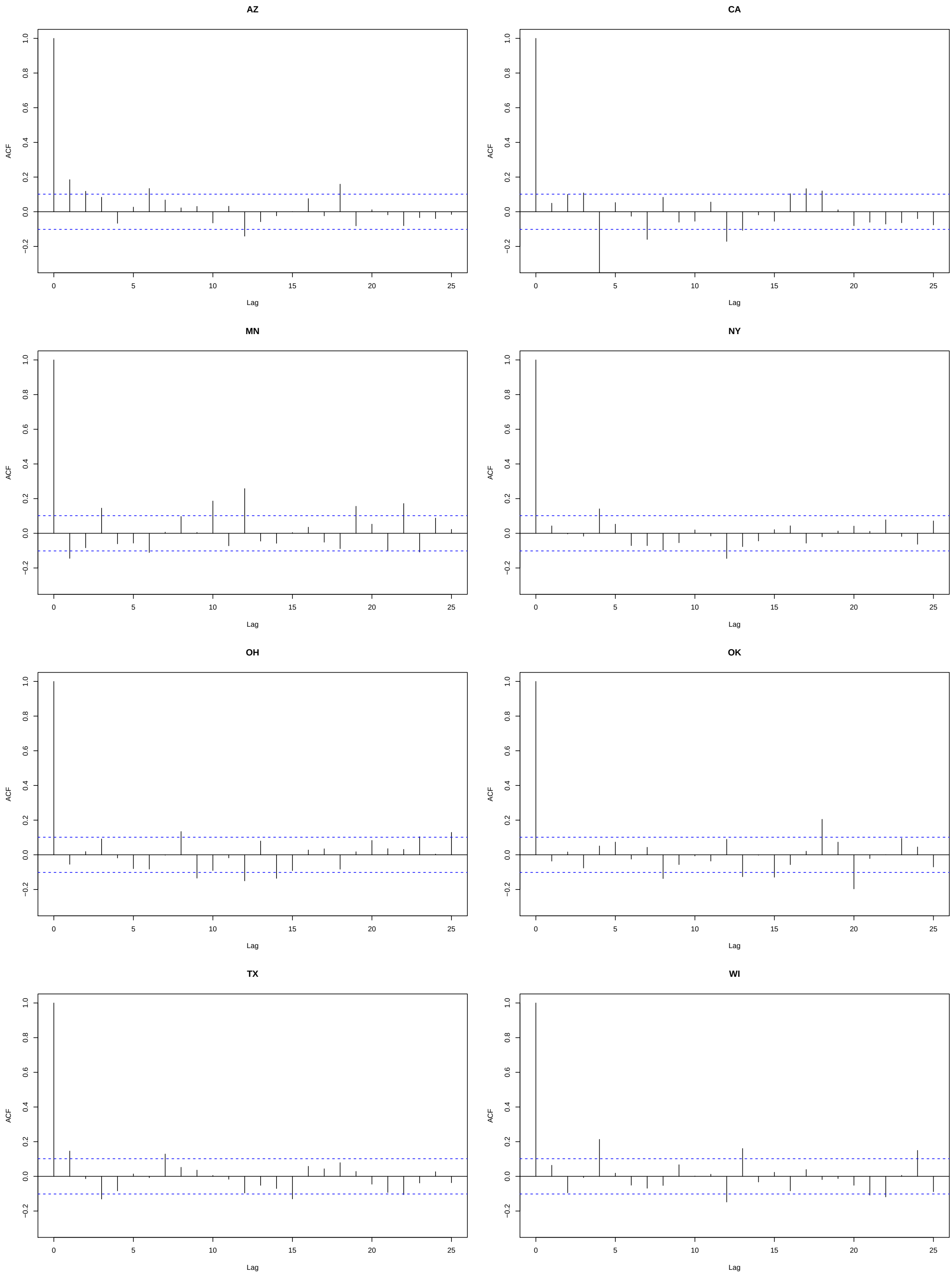


Figure S27: Diagnostic plots for validation of the final model estimated with REML.

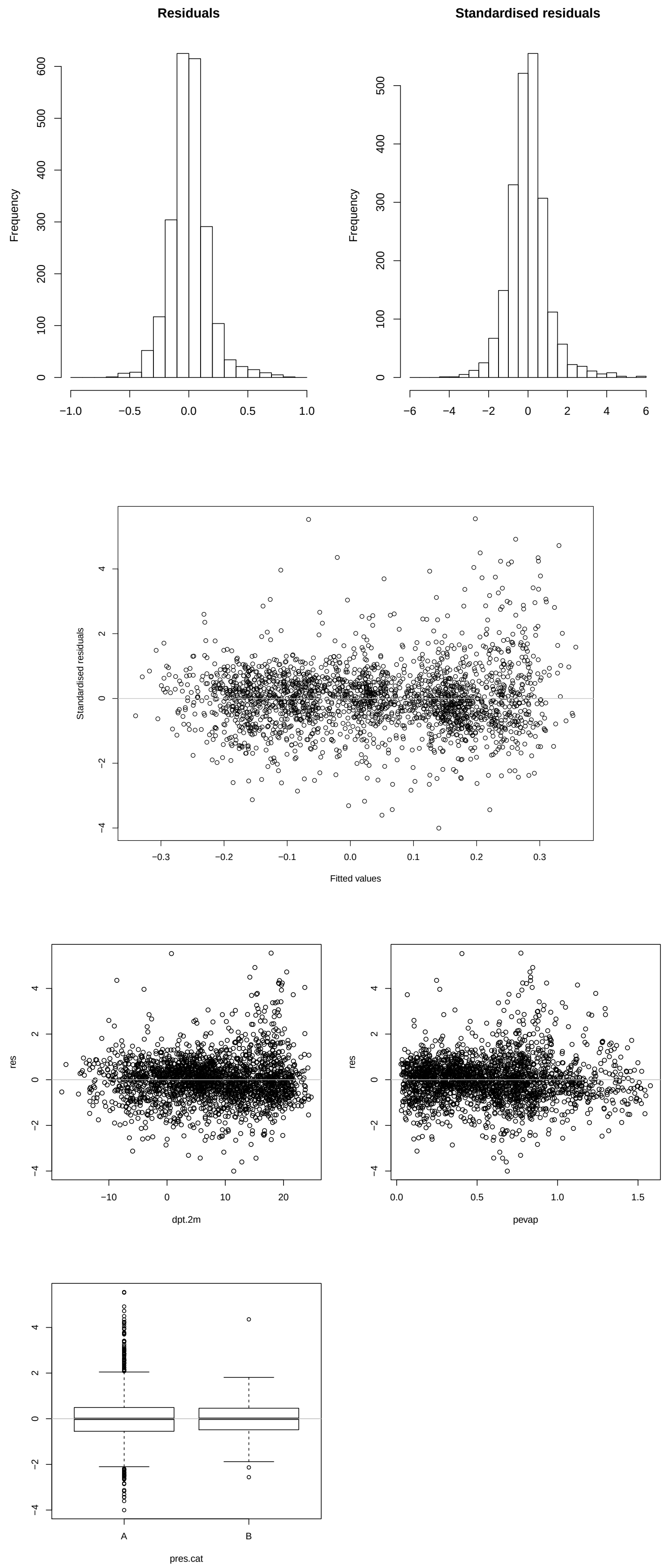
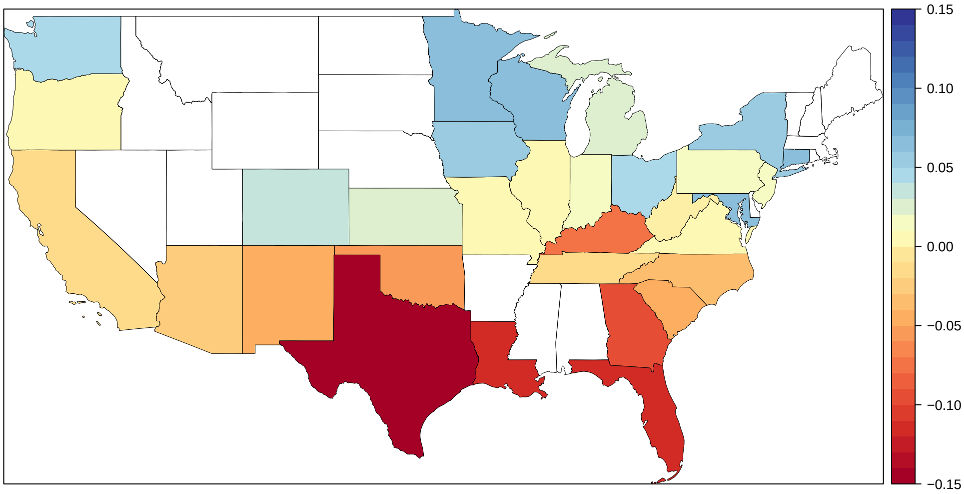


Figure S28: Estimated random intercept across states. (Top) Map of the US showing the estimated random intercept. White states are those with no data. (Bottom) Estimated random intercept across states against latitude of their state capital. Each point corresponds to a state. The line is a cubic spline.



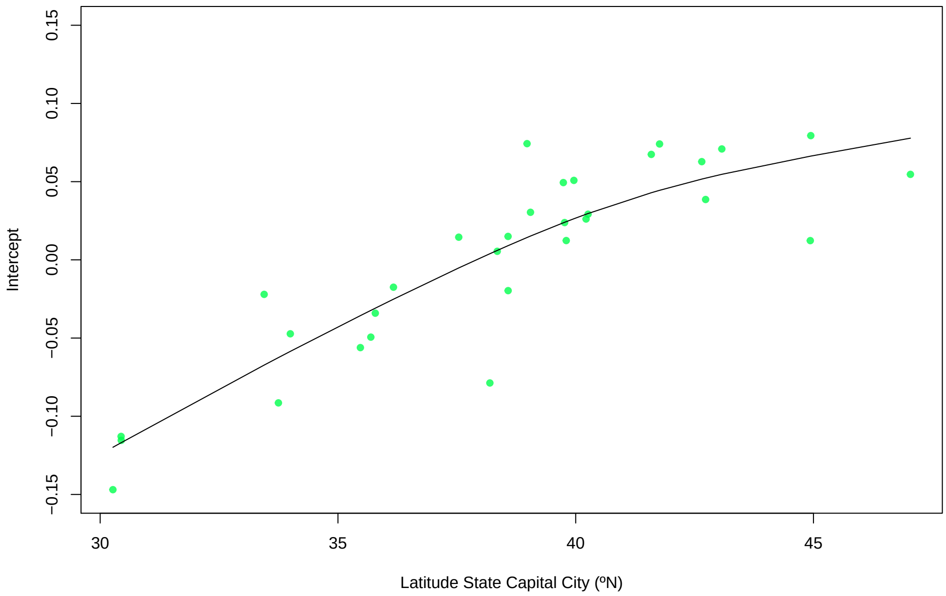


Figure S29: Bubbled plot of the normalised residuals for the final model. The size of the circles represents the absolute value of the residual, whereas the colour represents the sign (blue for negative and red for positive). The circles are plotted close to the coordinates of the state capital city.

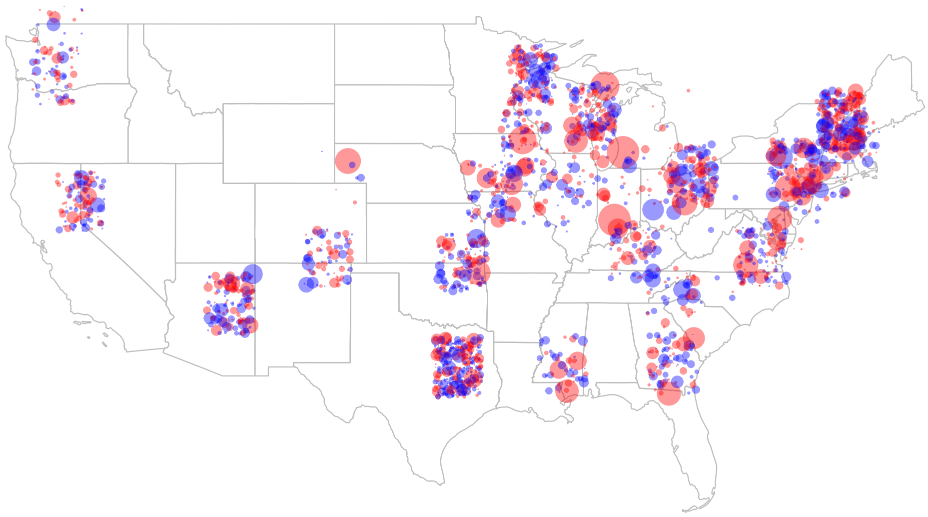
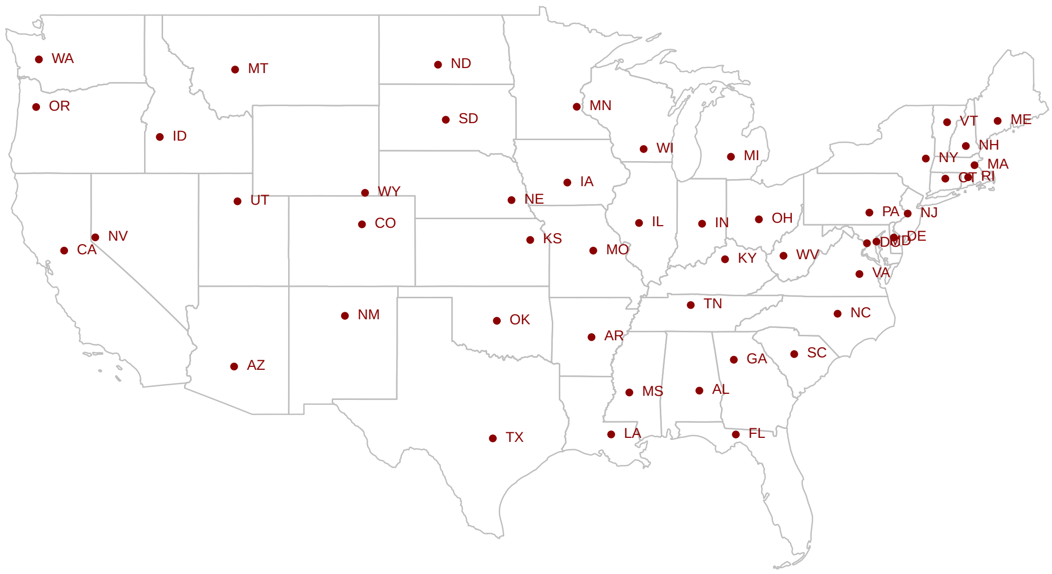


Figure S30: Map of the contiguous United States with the abbreviation of the names of the states. The points indicate the geographical situation of the state capital cities.



S5 Supplementary Tables

Table S1: Latitudinal regressions for the timing of the peak jointly for enterovirus and poliomyelitis cases.

	Common slope, different intercepts			Different slopes, different intercepts		
Parameter	Estimate	Std. Error	p-value	Estimate	Std. Error	p-value
α	5.11944	0.31026	$< 2e\text{-}16$	4.993347	0.666589	1.27e-10
σ	0.89150	0.08941	2.5e-15	1.051448	0.752395	0.167
β	0.07429	0.00782	1.9e-14	0.077584	0.017272	2.59e-05
δ	–	–	–	-0.004155	0.019404	0.831
AIC	61.3286			63.28025		
Adjusted R^2	0.73			0.73		

Table S2: Latitudinal regressions for the annual amplitude jointly for enterovirus and poliomyelitis cases.

	Common slope, different intercepts			Different slopes, different intercepts		
Parameter	Estimate	Std. Error	p-value	Estimate	Std. Error	p-value
α	0.0176478	0.0132589	0.18727	-0.0344587	0.0276468	0.2166
σ	0.0122823	0.0038210	0.00194	0.0783753	0.0312056	0.0142
β	0.0023070	0.0003342	1.49e-09	0.0036673	0.0007163	2.4e-06
δ	–	–	–	-0.0017168	0.0008048	0.0363
AIC	-424.1932			-426.8498		
Adjusted R^2	0.44			0.47		

Table S3: Latitudinal regressions for the timing of the peak jointly for enterovirus and poliomyelitis cases, accounting only for states with data for both.

	Common slope, different intercepts			Different slopes, different intercepts		
Parameter	Estimate	Std. Error	p-value	Estimate	Std. Error	p-value
α	4.997912	0.391197	$< 2\text{e-}16$	4.9933466	0.6903723	2.10e-09
σ	0.936243	0.095950	1.97e-13	0.9429517	0.8377121	0.265
β	0.077464	0.009977	2.67e-10	0.0775836	0.0178881	6.66e-05
δ	–	–	–	-0.0001745	0.0216462	0.994
AIC	53.35974			55.35967		
Adjusted R^2	0.75			0.74		

Table S4: Latitudinal regressions for the annual amplitude jointly for enterovirus and poliomyelitis cases, accounting only for states with data for both.

	Common slope, different intercepts			Different slopes, different intercepts		
Parameter	Estimate	Std. Error	p-value	Estimate	Std. Error	p-value
α	-0.0086698	0.0157277	0.58378	-0.0344587	0.0274112	0.214
σ	0.0134144	0.0038576	0.00102	0.0513108	0.0332613	0.129
β	0.0029940	0.0004011	8.09e-10	0.0036673	0.0007102	3.88e-06
δ	–	–	–	-0.0009858	0.0008595	0.257
AIC	-306.5844			-305.9837		
Adjusted R^2	0.55			0.56		

Table S5: Description of the climatic variables obtained from the NARR Monthly Means dataset. The units indicated here are those used in the mixed-effects models.

Variable	NARR	Unit	Description
Temperature	air.2m	$^\circ$C	Monthly mean air temperature at 2m
Precipitation	apcp	kg/m^2	Monthly accumulated total precipitation at surface
Dew point	dpt.2m	$^\circ$C	Monthly mean dew point temperature at 2m
Potential evaporation	pevap	kg/m^2	Monthly accumulated potential evaporation at surface
Pressure	pres.sfc	Pa	Monthly mean pressure at surface
Relative humidity	rhum.2m	%	Monthly mean relative humidity at 2m
Specific humidity	shum.2m	g/kg	Monthly mean specific humidity at 2m

Table S6: Estimated coefficients of the univariable linear models for the amplitude. The variables are ordered by increasing AIC.

Variable	Estimate	Std. Error	p-value	AIC	R^2
Dew point Range	2.785E-03	2.928E-04	<0.001	-785	0.35
Specific humidity Min	-1.218E-02	1.383E-03	<0.001	-776	0.32
Temperature Min	-2.119E-03	2.486E-04	<0.001	-773	0.30
Latitude	3.365E-03	4.042E-04	<0.001	-770	0.29
Temperature Range	2.868E-03	3.561E-04	<0.001	-767	0.28
Dew point Min	-2.638E-03	3.279E-04	<0.001	-767	0.28
Potential evaporation Max	-5.988E-02	7.627E-03	<0.001	-765	0.27
Potential evaporation Median	-8.884E-02	1.159E-02	<0.001	-763	0.26
Temperature Median	-2.997E-03	3.925E-04	<0.001	-762	0.26
Longitude	1.123E-03	1.528E-04	<0.001	-759	0.24
Birth rates	-7.537E-03	1.084E-03	<0.001	-754	0.22
Potential evaporation Min	-1.231E-01	1.850E-02	<0.001	-751	0.21
Specific humidity Range	4.980E-03	7.981E-04	<0.001	-747	0.19
Relative humidity Min	8.393E-04	1.395E-04	<0.001	-745	0.18
Specific humidity Median	-5.770E-03	1.029E-03	<0.001	-741	0.16
Potential evaporation Range	-5.582E-02	1.003E-02	<0.001	-740	0.16
Relative humidity Range	-1.014E-03	2.094E-04	<0.001	-734	0.12
Temperature Max	-2.693E-03	5.567E-04	<0.001	-734	0.12
Dew point Median	-2.159E-03	4.812E-04	<0.001	-731	0.11
Relative humidity Median	8.029E-04	1.825E-04	<0.001	-730	0.10
Population density (log10)	2.225E-02	5.333E-03	<0.001	-728	0.09
Relative humidity Max	8.776E-04	2.307E-04	<0.001	-726	0.08
Births (log10)	-2.179E-02	6.242E-03	0.001	-723	0.07
Precipitation Min	1.413E-02	4.205E-03	0.001	-723	0.06
Population size (log10)	-1.719E-02	6.531E-03	0.009	-718	0.04
Precipitation Median	6.240E-03	2.506E-03	0.014	-718	0.04
Dew point Max	1.331E-03	6.166E-04	0.032	-716	0.03
Pressure Range	-2.267E-05	1.150E-05	0.050	-715	0.02
Precipitation Range	-2.315E-03	1.256E-03	0.067	-715	0.02
Specific humidity Max	1.632E-03	8.966E-04	0.070	-715	0.02
Pressure Median	7.933E-07	5.002E-07	0.115	-714	0.01
Pressure Min	7.839E-07	5.002E-07	0.119	-714	0.01
Pressure Max	7.535E-07	5.046E-07	0.137	-714	0.01
Elevation	-8.242E-06	6.144E-06	0.182	-713	0.01
Precipitation Max	-1.025E-03	1.216E-03	0.401	-712	0.00

Table S7: Estimated coefficients of the univariable linear models for the timing of the peak. The variables are ordered by increasing AIC.

Variable	Estimate	Std. Error	p-value	AIC	R^2
Latitude	9.143E-02	1.197E-02	<0.001	375	0.26
Specific humidity Min	-3.046E-01	4.228E-02	<0.001	380	0.24
Dew point Min	-6.698E-02	9.940E-03	<0.001	385	0.21
Temperature Median	-7.728E-02	1.169E-02	<0.001	387	0.21
Temperature Min	-5.040E-02	7.649E-03	<0.001	387	0.21
Potential evaporation Min	-3.526E+00	5.357E-01	<0.001	387	0.21
Specific humidity Median	-1.784E-01	2.918E-02	<0.001	392	0.18
Temperature Max	-9.518E-02	1.567E-02	<0.001	392	0.18
Potential evaporation Median	-2.078E+00	3.543E-01	<0.001	394	0.17
Dew point Median	-7.761E-02	1.337E-02	<0.001	395	0.17
Potential evaporation Max	-1.181E+00	2.432E-01	<0.001	404	0.12
Temperature Range	5.491E-02	1.149E-02	<0.001	404	0.12
Birth rates	-1.525E-01	3.352E-02	<0.001	406	0.11
Dew point Range	4.614E-02	1.017E-02	<0.001	406	0.11
Births (log10)	-6.145E-01	1.825E-01	0.001	415	0.06
Population size (log10)	-5.255E-01	1.900E-01	0.006	418	0.04
Specific humidity Max	-6.955E-02	2.619E-02	0.009	419	0.04
Relative humidity Max	1.788E-02	6.802E-03	0.009	419	0.04
Potential evaporation Range	-8.072E-01	3.167E-01	0.012	419	0.04
Dew point Max	-4.572E-02	1.802E-02	0.012	420	0.04
Longitude	1.084E-02	5.169E-03	0.038	422	0.03
Relative humidity Min	8.964E-03	4.412E-03	0.044	422	0.02
Elevation	3.101E-04	1.765E-04	0.081	423	0.02
Pressure Max	-2.547E-05	1.452E-05	0.081	423	0.02
Pressure Range	-5.761E-04	3.388E-04	0.091	423	0.02
Pressure Median	-2.443E-05	1.441E-05	0.092	423	0.02
Pressure Min	-2.399E-05	1.441E-05	0.098	423	0.02
Relative humidity Median	8.500E-03	5.525E-03	0.126	424	0.01
Precipitation Range	-4.969E-02	3.808E-02	0.194	424	0.01
Precipitation Min	1.387E-01	1.255E-01	0.271	425	0.01
Precipitation Max	-3.341E-02	3.634E-02	0.359	425	0.01
Specific humidity Range	2.292E-02	2.648E-02	0.388	425	0.00
Population density (log10)	1.032E-01	1.623E-01	0.526	426	0.00
Relative humidity Range	-3.376E-03	6.561E-03	0.608	426	0.00
Precipitation Median	1.235E-02	7.494E-02	0.869	426	0.00

Table S8: Mixed-effects model selection using a bottom-up strategy.

Model	AIC	BIC
One fixed-effect variable		
Temperature (T)	-1545	-1522
Precipitation (R)	-52	-30
Dew point (DP)	**-1572**	**-1549**
Potential evaporation (PE)	-1092	-1069
Pressure (PR)	-18	41
Relative humidity (RH)	-20	3
Specific humidity (SH)	-1481	-1458
Two fixed-effect variables		
DP + T	-1611	-1582
DP + R	-1576	-1548
DP + PE	**-1646**	**-1618**
DP + PR	-1594	-1566
DP + RH	-1603	-1574
DP + SH	-1576	-1548
Three fixed-effect variables		
DP + PE + T	-1647	-1612
DP + PE + R	-1645	-1611
DP + PE + PR	**-1657**	**-1622**
DP + PE + PR (cat.)	**-1657**	**-1623**
DP + PE + RH	-1647	-1613
DP + PE + SH	-1648	-1614
DP + PE + Pop. size	-1647	-1613
DP + PE + Pop. density	-1645	-1611
DP + PE + Births	-1649	-1615
DP + PE + Birth rates	-1645	-1611

Table S9: Estimated parameters (using ML) of the model with only dew point as fixed effects before accounting for temporal auto-correlated errors.

Parameters	Mean (se)	p-value
Fixed effects		
Intercept, α	-0.130 (0.0215)	<0.001
Coeff. DP, β_{DP}	0.023 (0.0005)	<0.001
Random effects		
Var. random intercept, σ_a	0.114	
Var. residuals, σ_ε	0.166	

Table S10: Estimated parameters (using ML) of the model with only potential evaporation as fixed effects before accounting for temporal auto-correlated errors.

Parameters	Mean (se)	p-value
Fixed effects		
Intercept, α	-0.229 (0.0166)	<0.001
Coeff. PE, β_{PE}	0.506 (0.0129)	<0.001
Random effects		
Var. random intercept, σ_a	0.077	
Var. residuals, σ_ε	0.186	

Table S11: Estimated parameters (using ML) of the model with only dew point and potential evaporation as fixed effects before accounting for temporal auto-correlated errors.

Parameters	Mean (se)	p-value
Fixed effects		
Intercept, α	-0.178 (0.0200)	<0.001
Coeff. DP, β_{DP}	0.018 (0.0007)	<0.001
Coeff. PE, β_{PE}	0.159 (0.0180)	<0.001
Random effects		
Var. random intercept, σ_a	0.101	
Var. residuals, σ_ε	0.163	

Table S12: Estimated parameters (using ML) of the model with only dew point, potential evaporation and pressure (categorical) as fixed effects before accounting for temporal auto-correlated errors.

Parameters	Mean (se)	p-value
Fixed effects		
Intercept, α	-0.192 (0.0174)	<0.001
Coeff. DP, β_{DP}	0.018 (0.0007)	<0.001
Coeff. PE, β_{PE}	0.154 (0.0180)	<0.001
Coeff. PR (cat.), β_{PR}	0.262 (0.0678)	<0.001
Random effects		
Var. random intercept, σ_a	0.082	
Var. residuals, σ_ε	0.163	

Table S13: Estimated parameters (using REML) of the final model (i.e. optimal fixed-effects structure and accounting for temporal auto-correlated errors).

Parameters	Mean (se)	p-value
Fixed effects		
Intercept, α	-0.177 (0.0176)	<0.001
Coeff. DP, β_{DP}	0.014 (0.0009)	<0.001
Coeff. PE, β_{PE}	0.189 (0.0242)	<0.001
Coeff. PR (cat.), β_{PR}	0.179 (0.0613)	0.007
Random effects		
Var. random intercept, σ_a	0.069	
Var. residuals, σ_η	0.168	
Coeffs. AR errors		
ρ_1	0.24	
ρ_2	-0.11	
ρ_3	-0.04	
ρ_4	-0.21	
ρ_5	-0.05	
ρ_6	-0.08	
ρ_7	-0.06	
ρ_8	-0.10	
ρ_9	-0.04	
ρ_{10}	-0.03	
ρ_{11}	0.05	
ρ_{12}	0.16	

Table S14: Estimated parameters (using REML) of the best model with only one fixed-effect variable.

Parameters	Mean (se)	p-value
Fixed effects		
Intercept, α	-0.055 (0.0131)	<0.001
Coeff. DP, β_{DP}	0.012 (0.0009)	<0.001
Random effects		
Var. random intercept, σ_a	0.054	
Var. residuals, σ_η	0.191	
Coeffs. AR errors		
ρ_1	0.27	
ρ_2	-0.10	
ρ_3	-0.04	
ρ_4	-0.22	
ρ_5	-0.06	
ρ_6	-0.11	
ρ_7	-0.08	
ρ_8	-0.12	
ρ_9	-0.05	
ρ_{10}	-0.03	
ρ_{11}	0.06	
ρ_{12}	0.20	

References

[1] Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models* (2016). URL http://CRAN.R-project.org/package=nlme. R package version 3.1-125.

[2] Nakagawa, S. & Schielzeth, H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**, 133–142 (2013).