

Supplementary Data for Bayesian investigation of SARS-CoV-2-related mortality in France

Louis Duchemin, Philippe Veber, Bastien Boussau

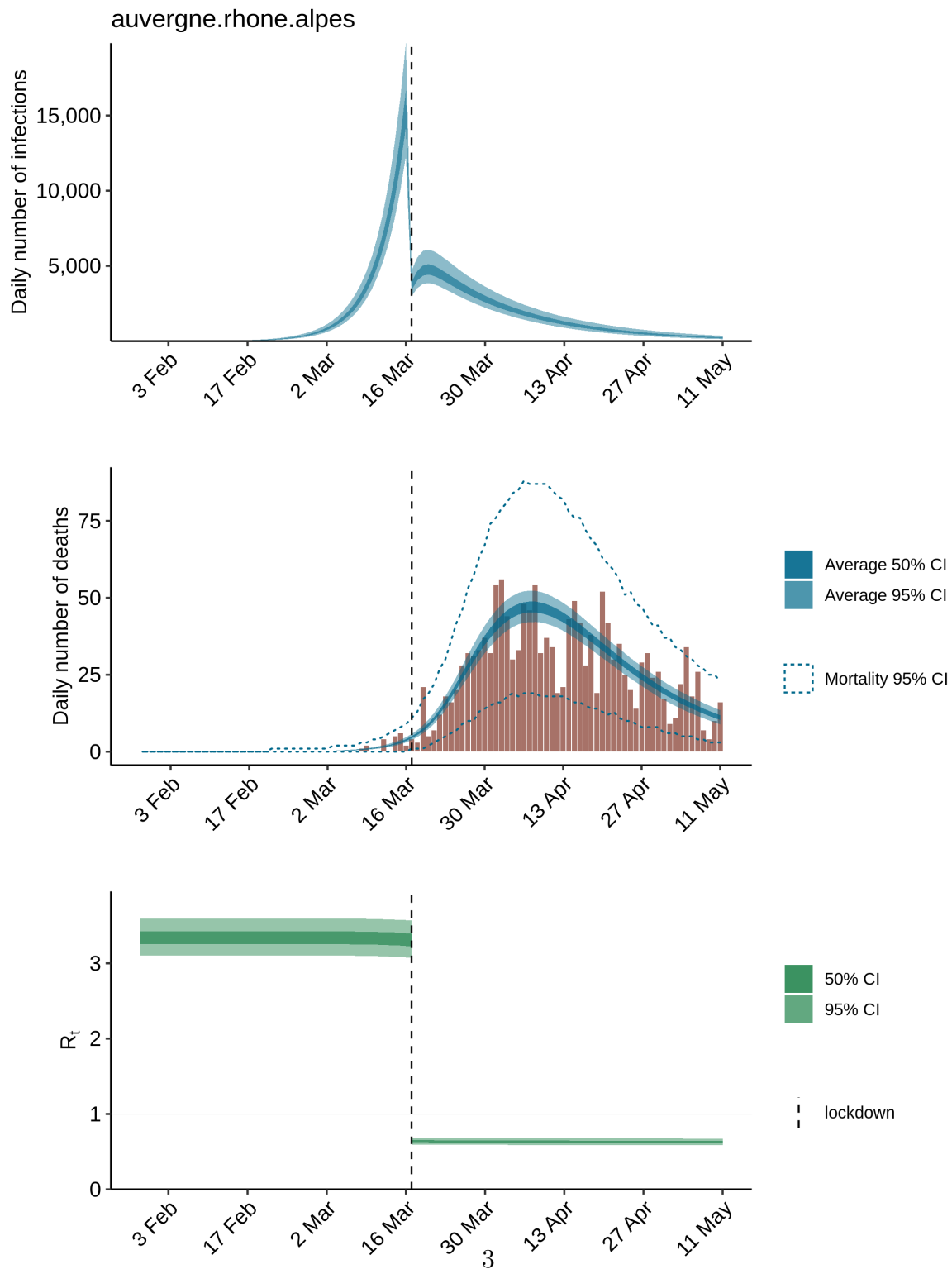
July 2, 2020

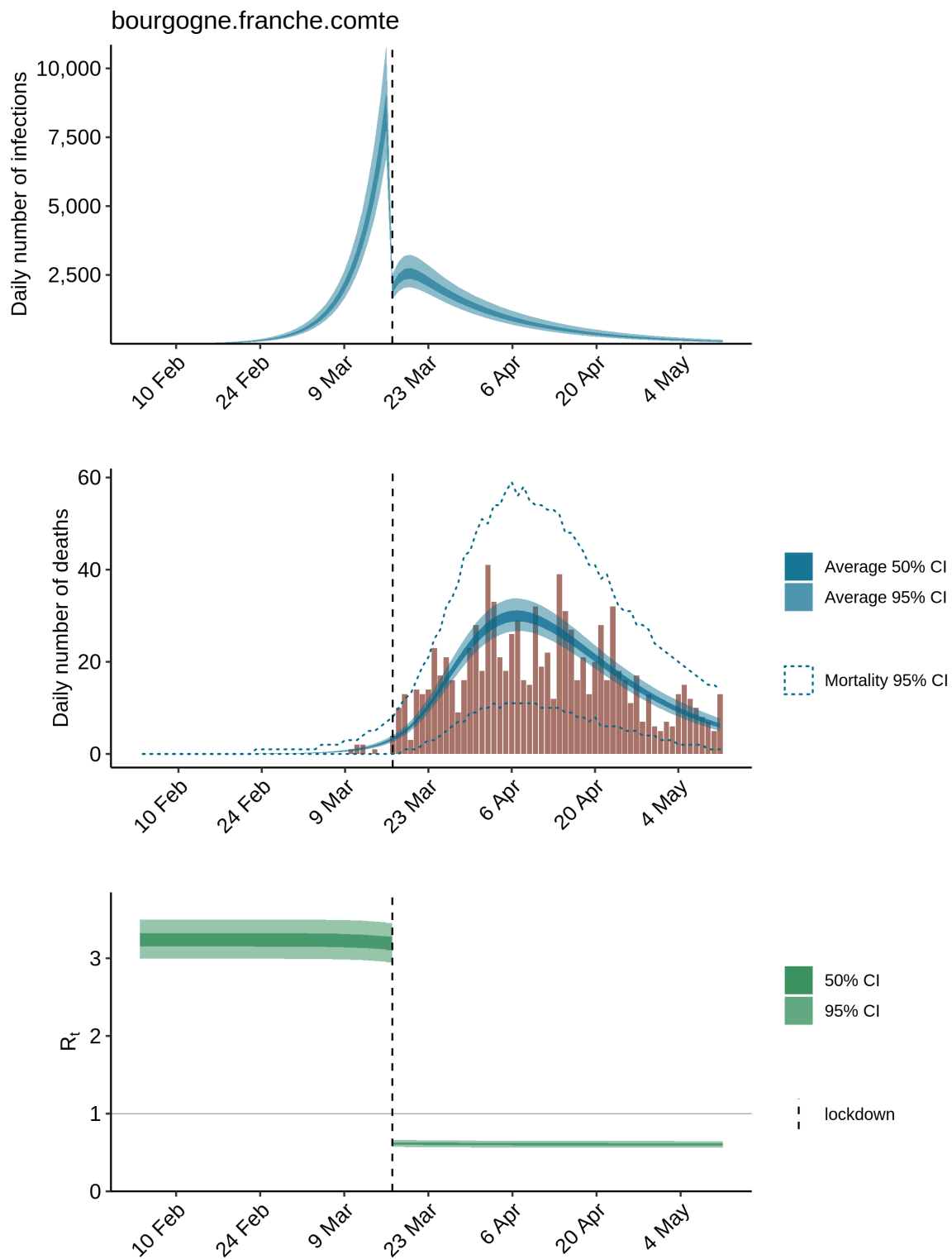
Contents

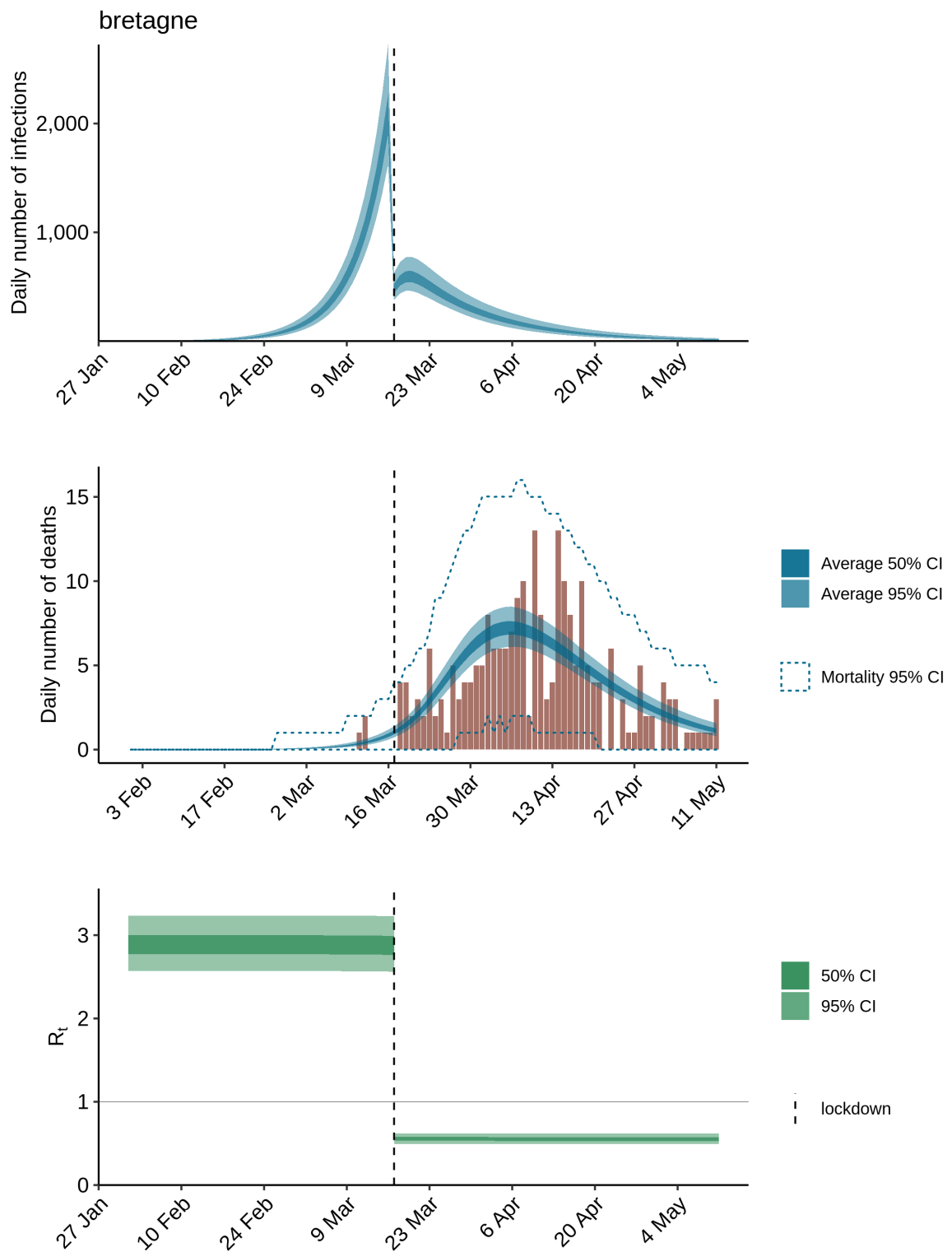
1	Region-wise estimates	3
2	Heterogeneity between regions	16
2.1	Posterior decoding	16
2.2	Results	18
3	Posterior distributions of α parameters	22
4	Region-wise simulations for week-ends effect size	24
5	Region-wise simulations for elections effect size	31

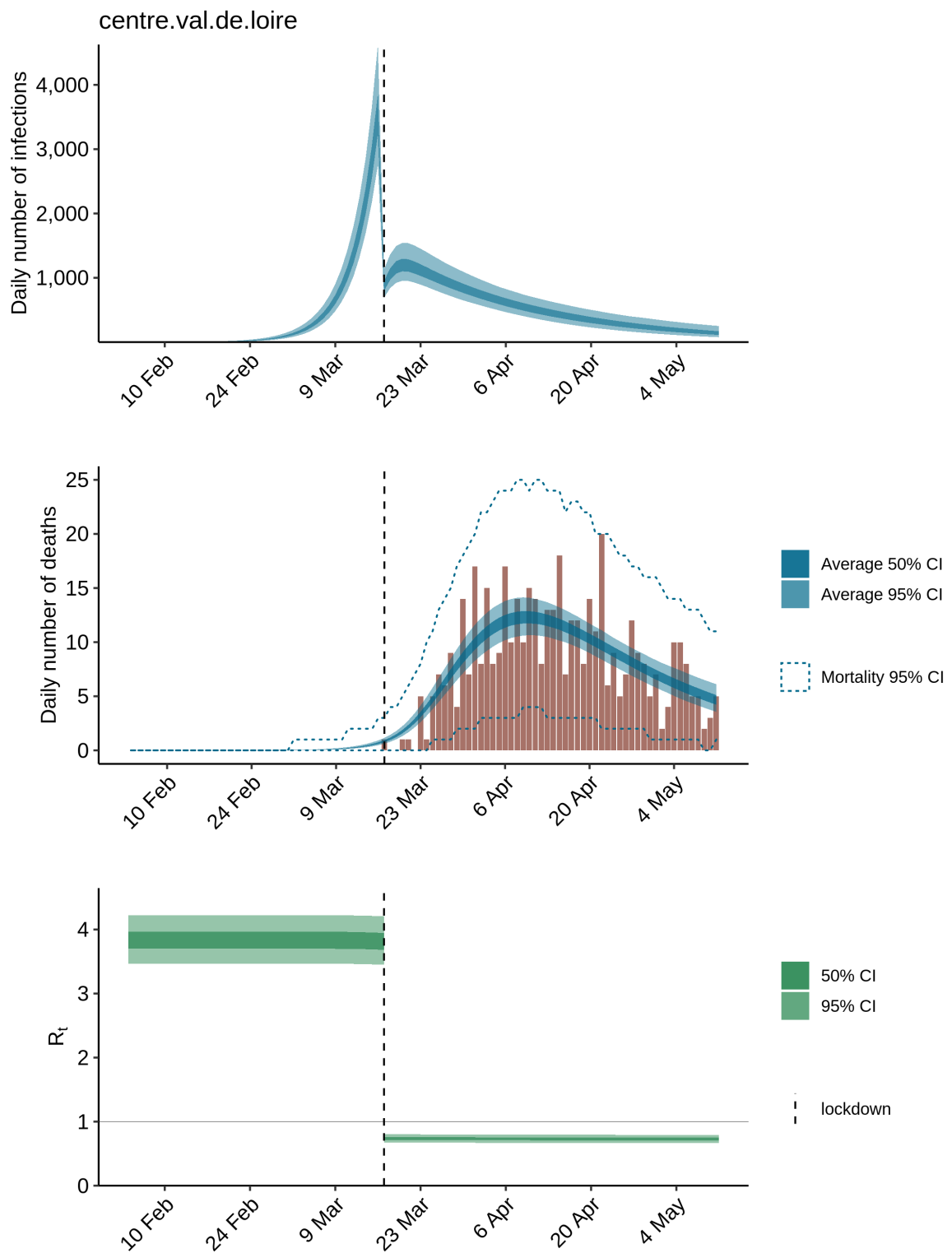
1 Region-wise estimates

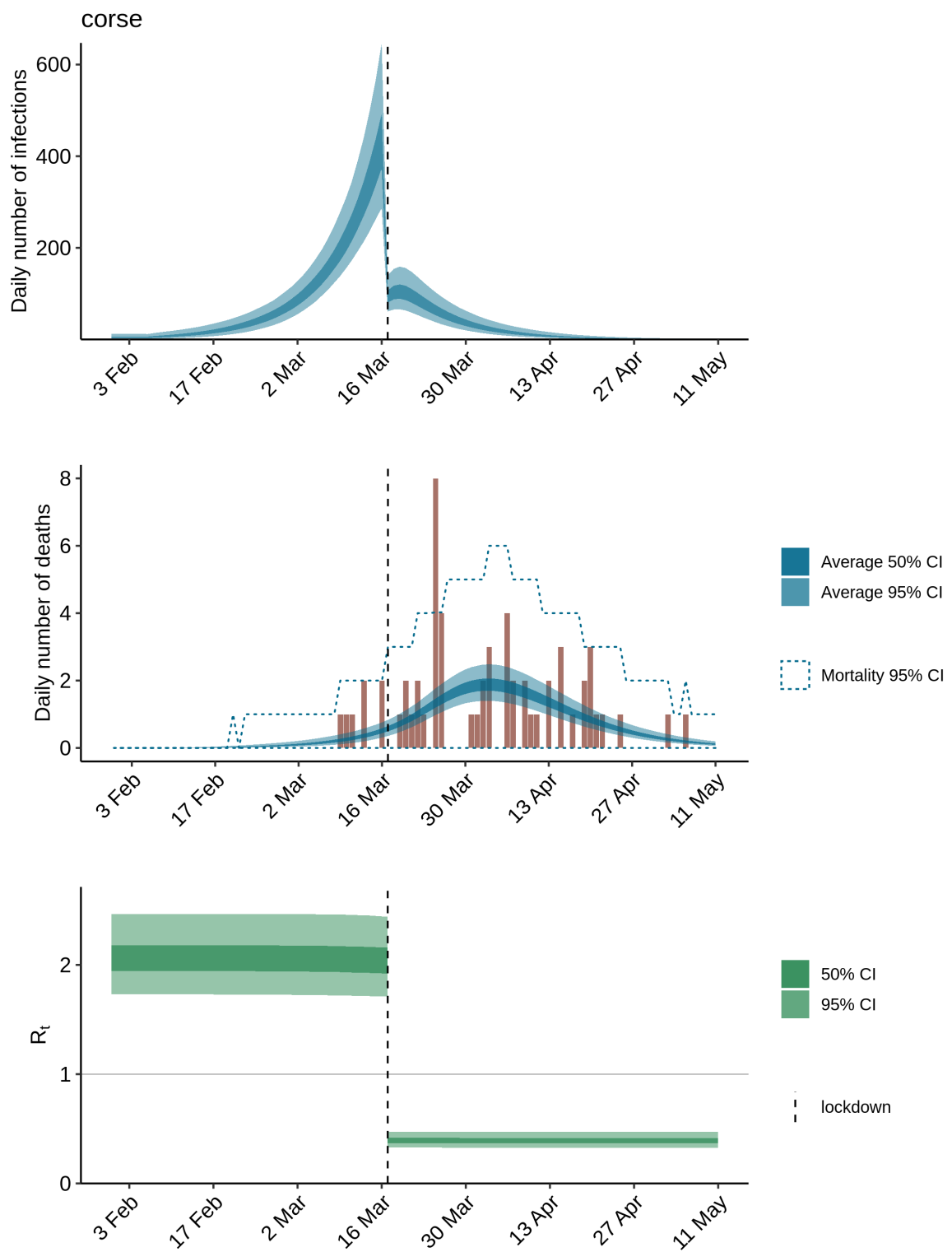
Figure 1: Region-wise estimates of daily new infection cases, death counts and reproduction number R_t

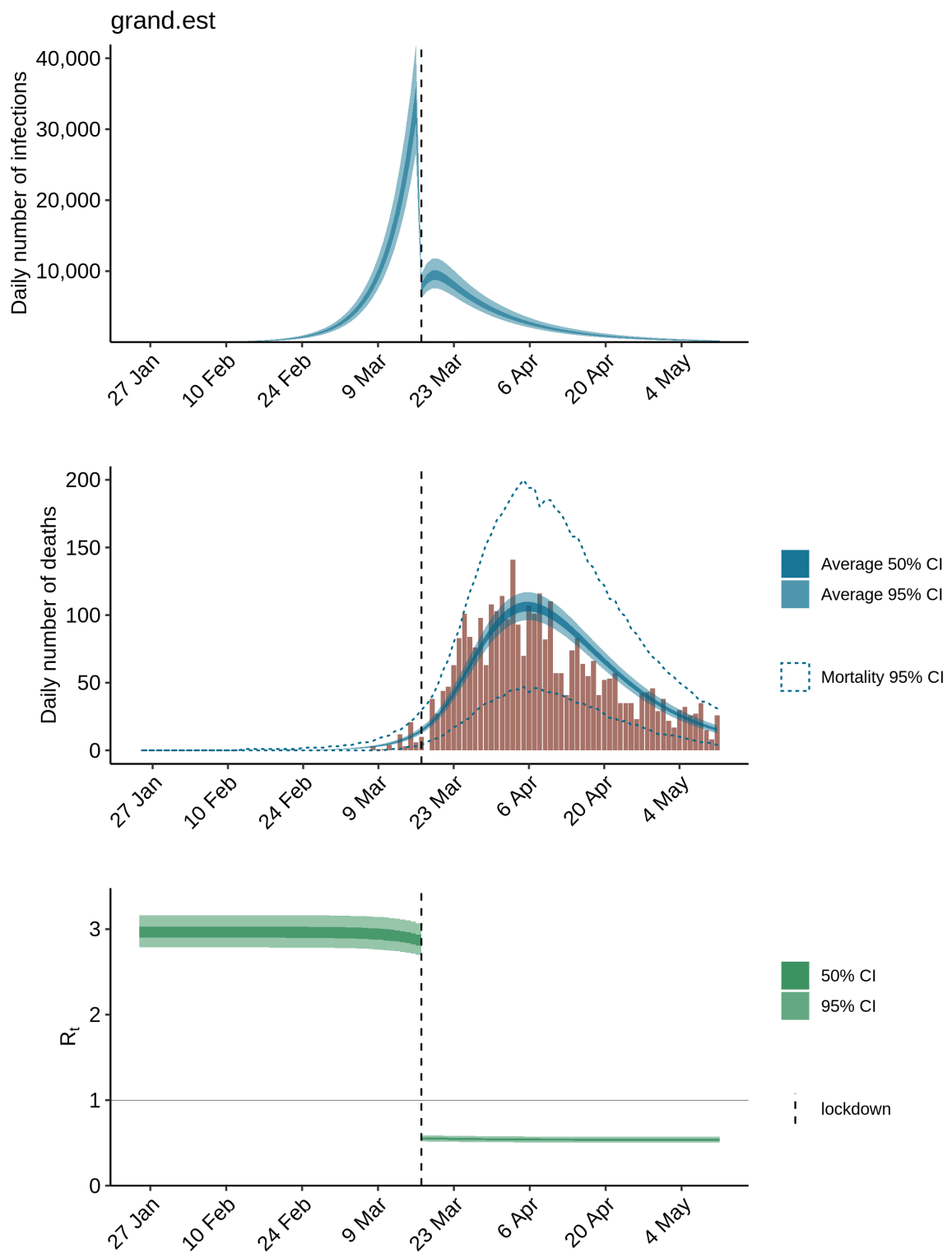


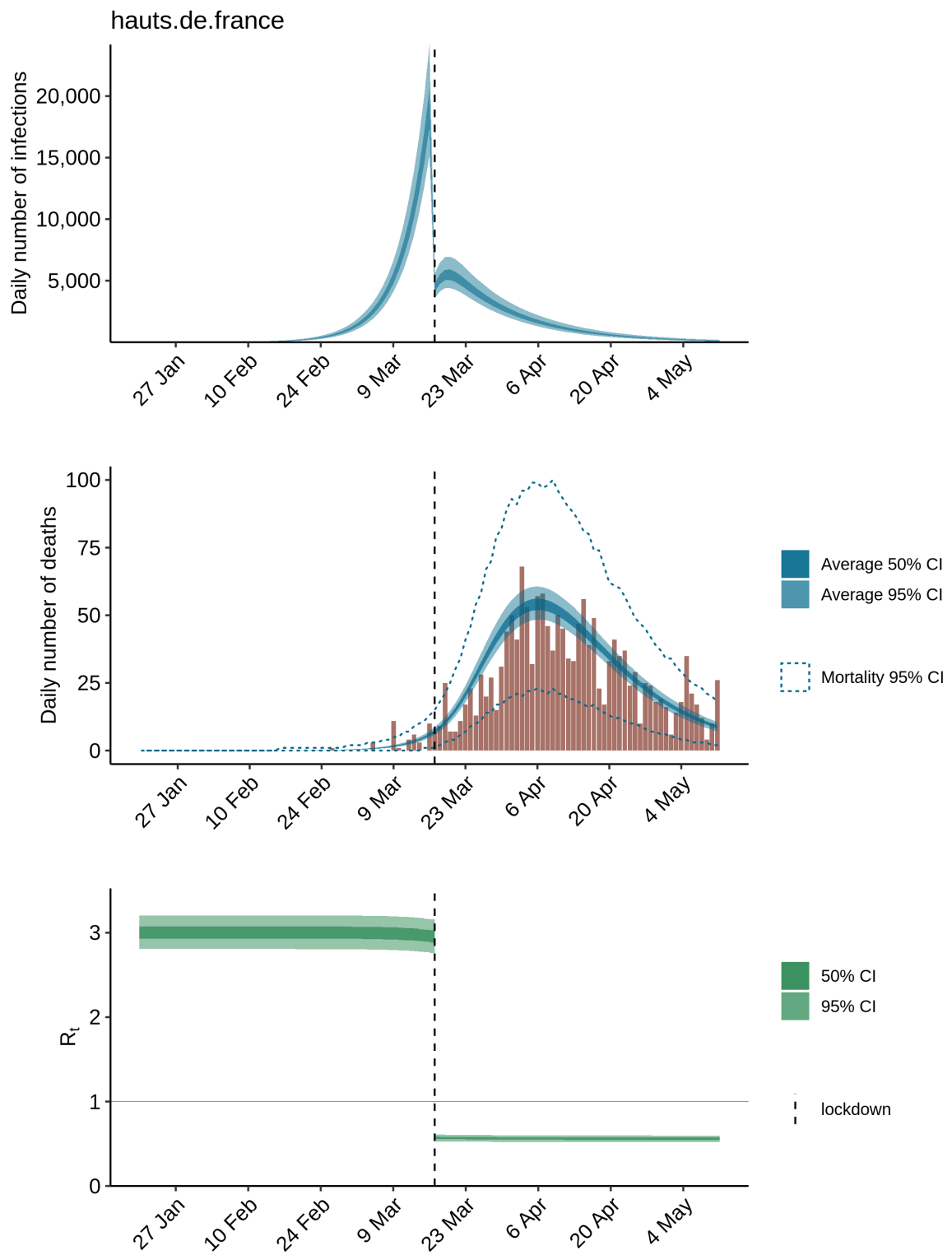


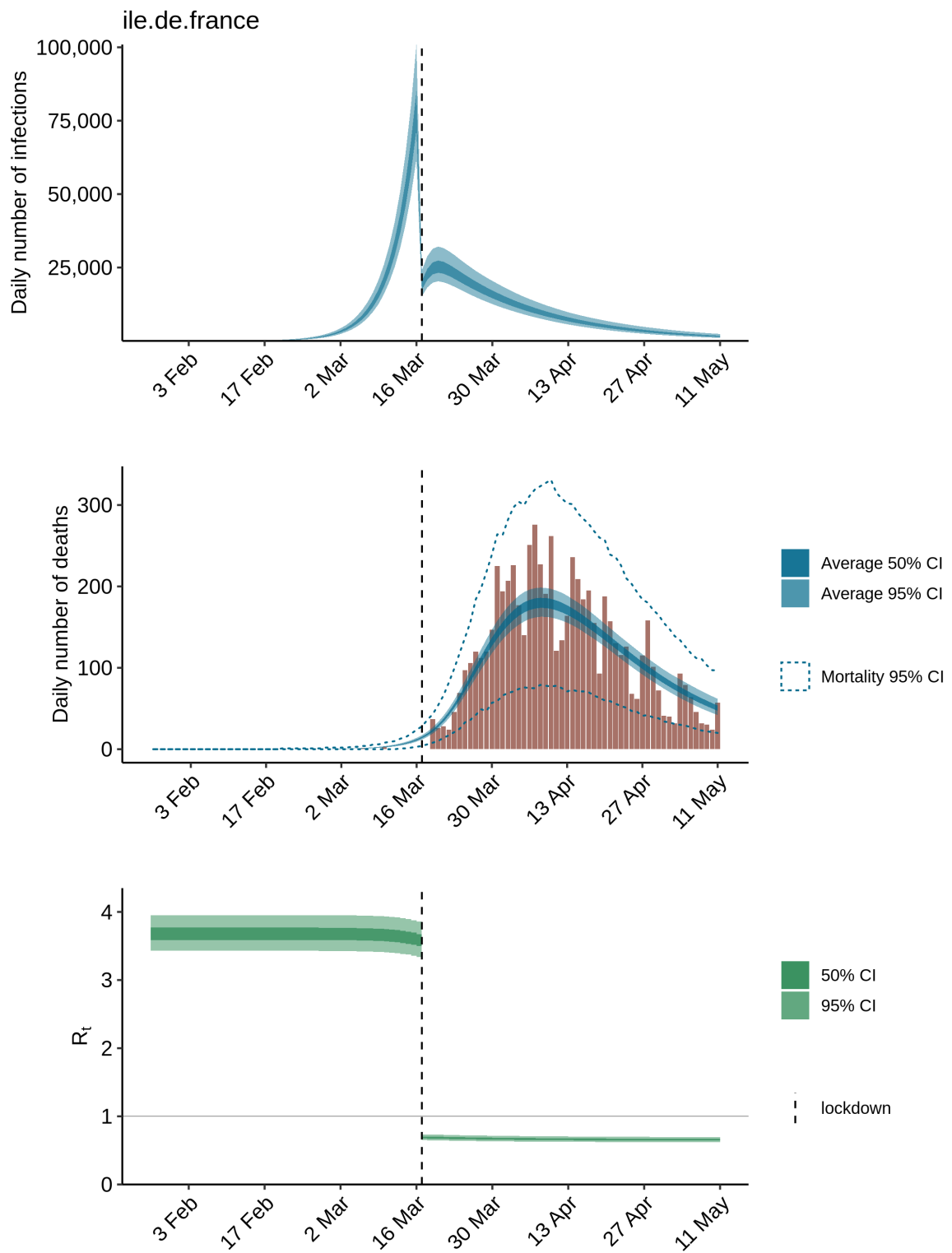


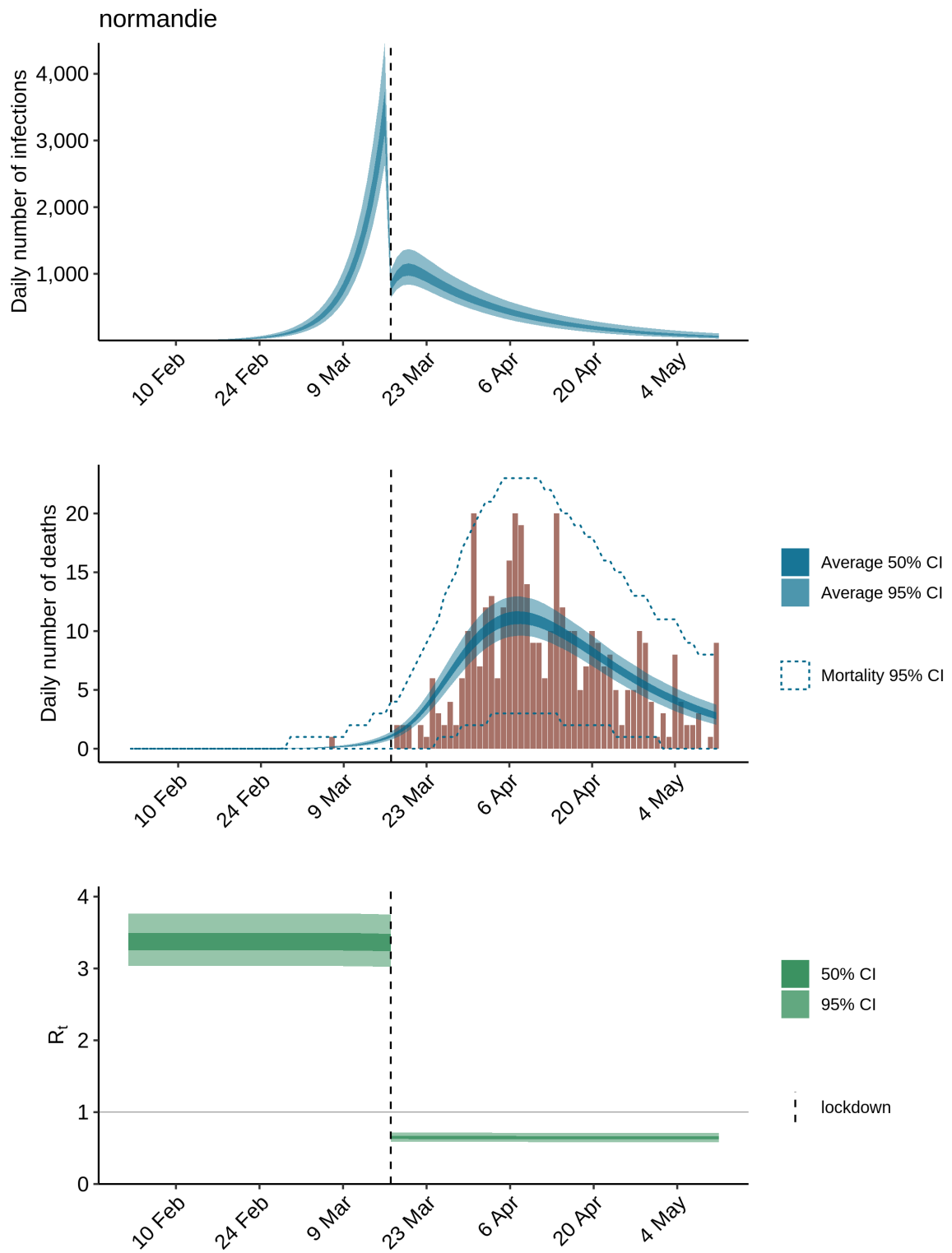


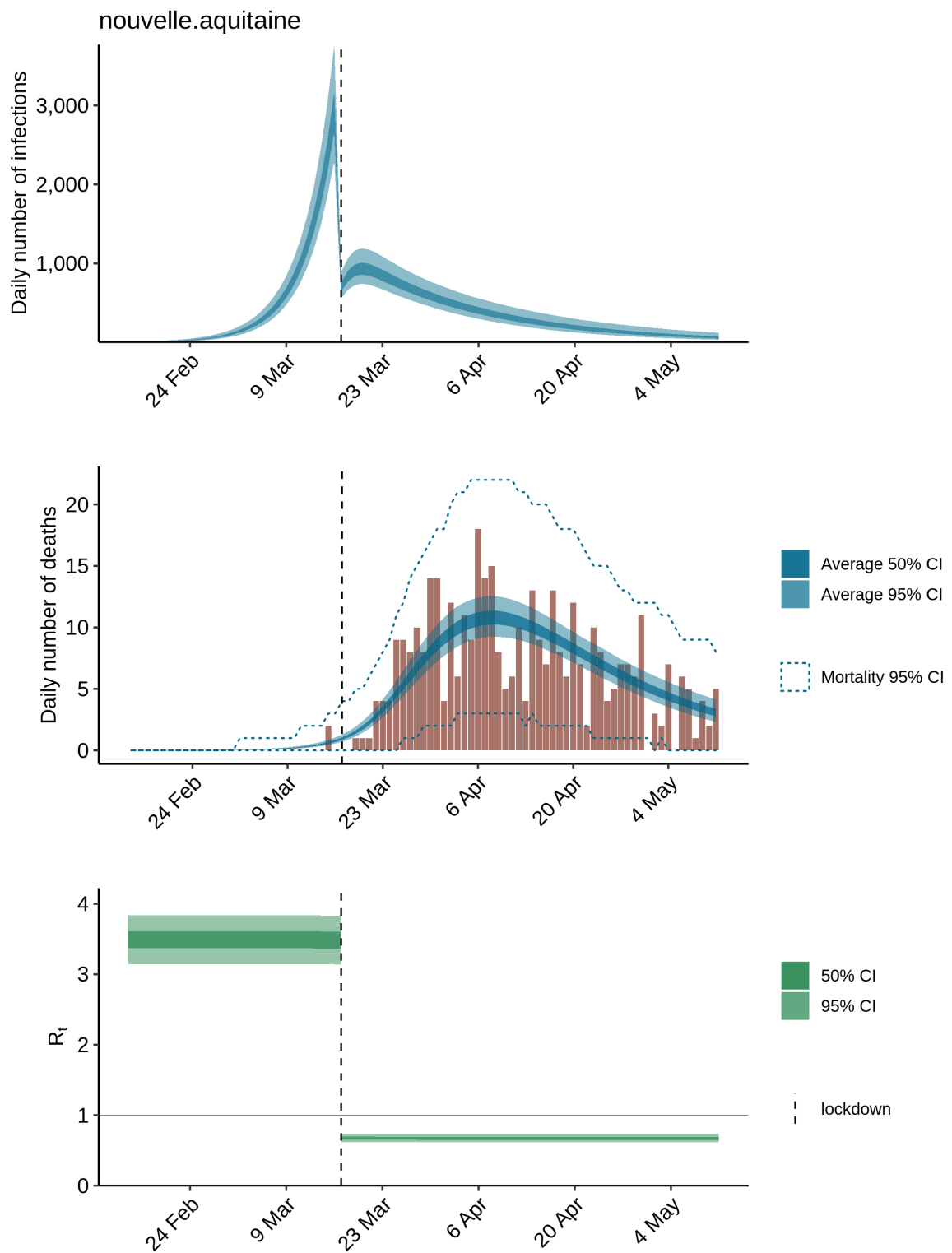


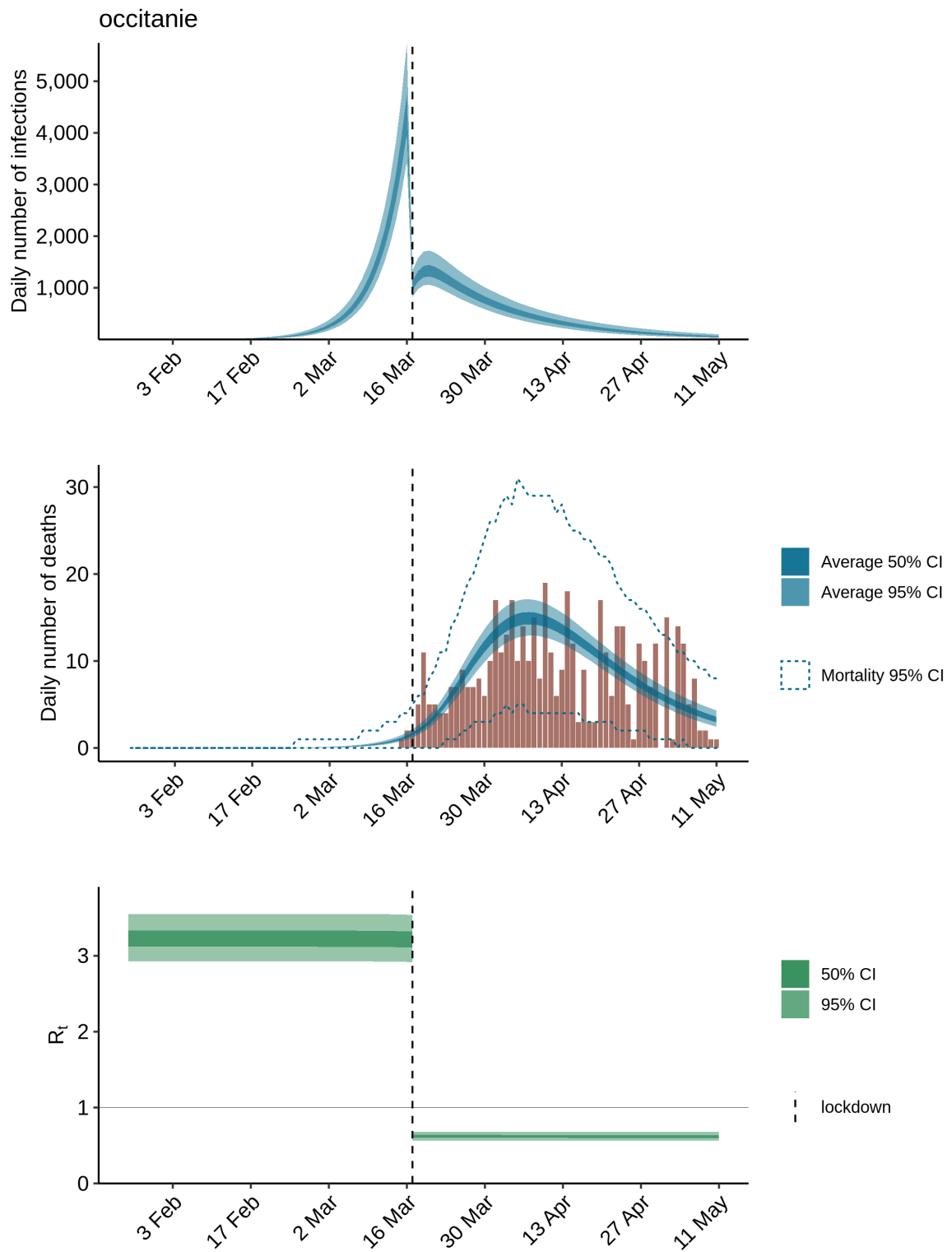


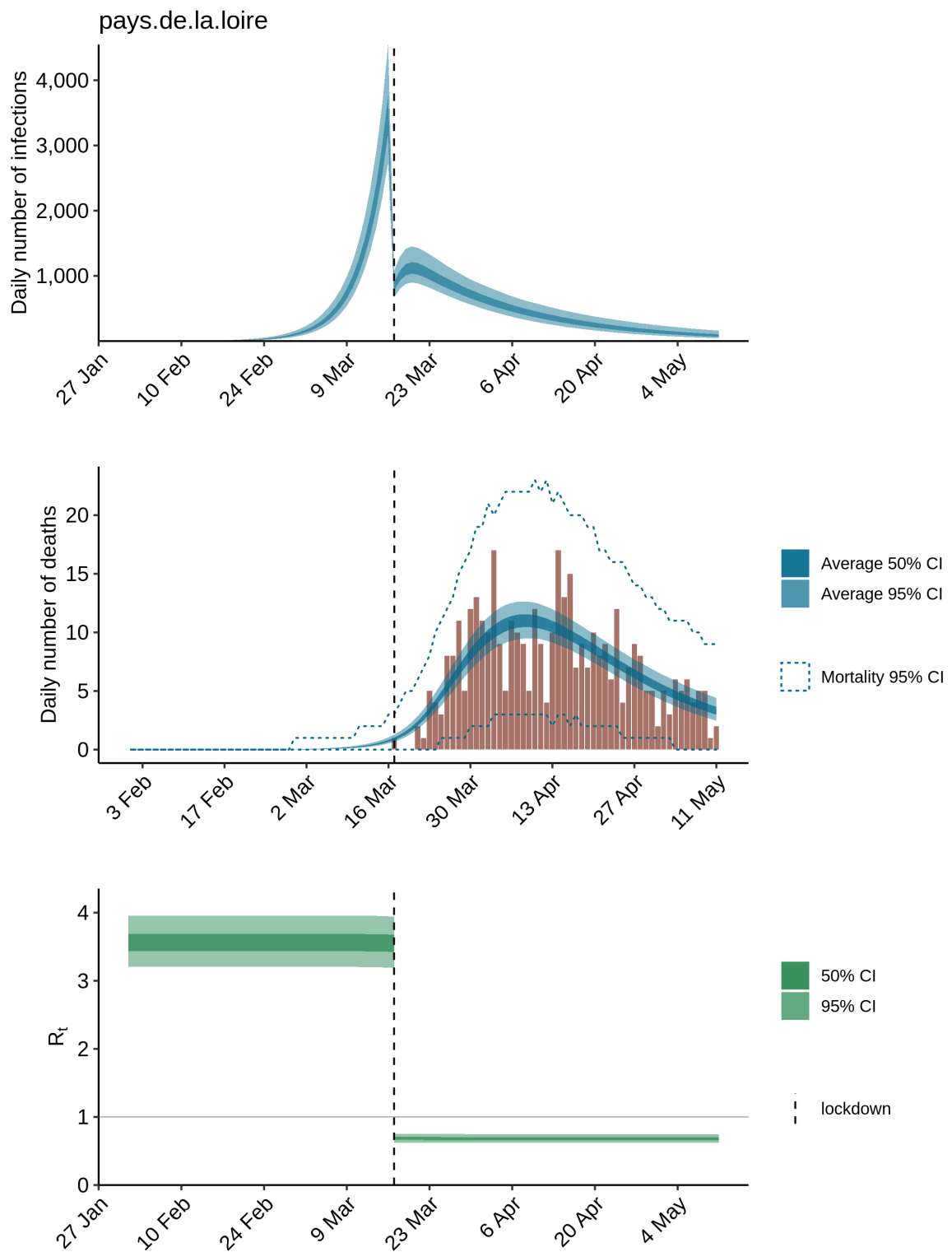


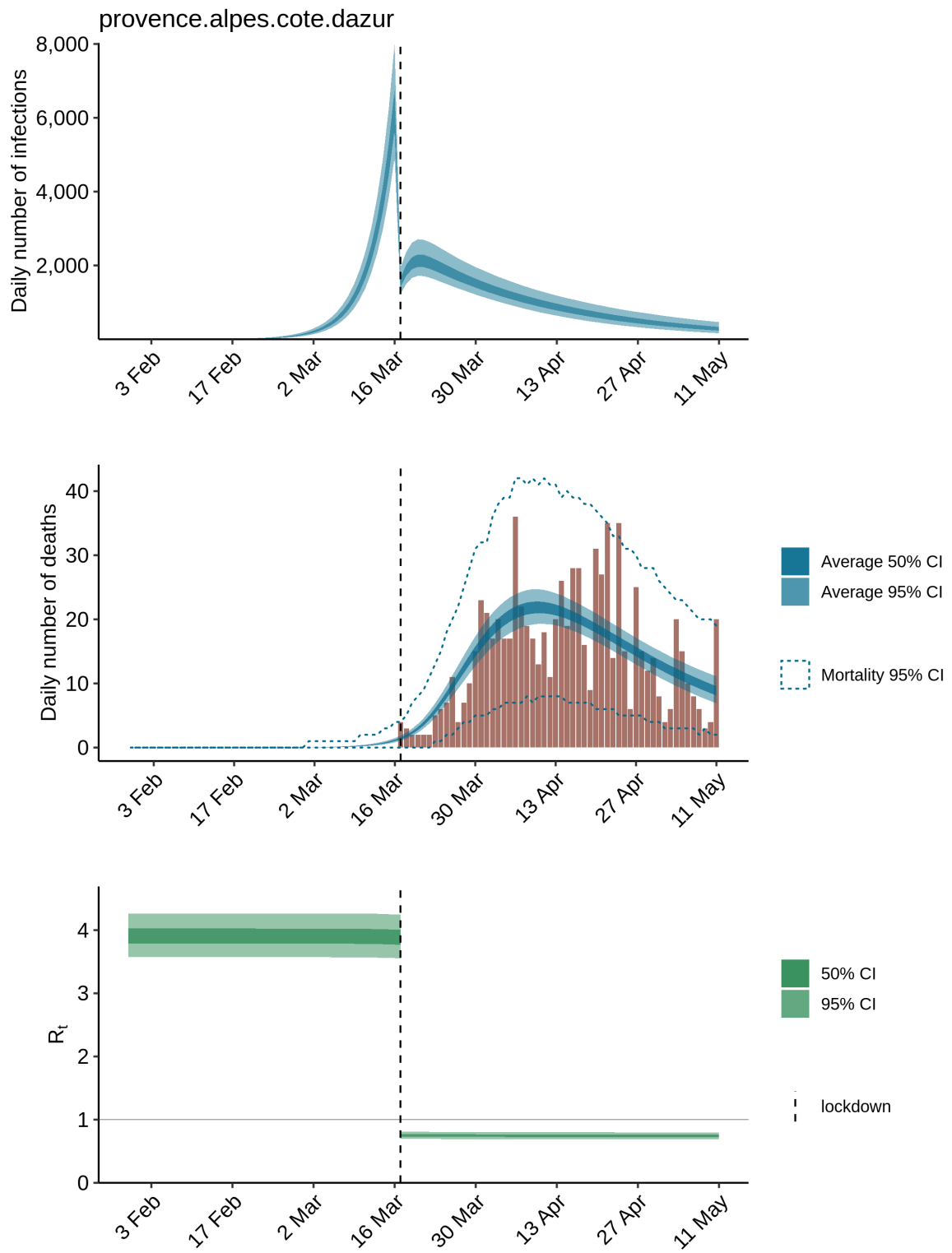












2 Heterogeneity between regions

2.1 Posterior decoding

MCMC inference in mixture models can be done in two ways: either by sampling mixture categories during the MCMC algorithm, or by integrating analytically over all mixture categories at each sample. The software Stan uses the latter approach, which means that one needs to use posterior-decoding to compute the probability for a data point to belong to a given category. Below we explain how one can use posterior-decoding to obtain region-wise estimates of parameters that depend on the mixture category.

$D_{t,m}$: the observed number of deaths on day t in region m

$D_{\bullet,m}$: the vector of observed number of deaths in region m over all days t

$\bar{D}_{t,m}$: the expected number of deaths on day t in region m

C_m : the mixture category of region m

α^k : the lockdown effect on regions in category $k \in 1, 2$

Ω : other parameters in the model

$$C_m - 1 \sim \text{Bern}(\theta)$$

$$\bar{D}_{t,m} = f(\alpha^1, \alpha^2, C_m, \Omega)$$

$$D_{t,m} \sim \text{NegBin}(\bar{D}_{t,m}, \phi)$$

The posterior probability that region m belongs to the category 1 of the mixture model is :

$$P(C_m = 1 | D_{\bullet,m}) = \int P(C_m = 1 | D_{\bullet,m}, \alpha^1, \alpha^2, \phi, \theta, \Omega) P(\alpha^1, \alpha^2, \phi, \theta, \Omega | D_{\bullet,m}) d\alpha^1 d\alpha^2 d\phi d\theta d\Omega \quad (1)$$

Note that $P(\alpha^1, \alpha^2, \phi, \theta, \Omega | D_{\bullet,m})$ is the posterior probability for all parameters other than C_m . In practice, the multidimensional integral on the right hand side of the equation above is approximated by a discrete sum over all n samples s obtained with a run of a Markov Chain Monte Carlo algorithm:

$$P(C_m = 1 | D_{\bullet,m}) \approx \frac{1}{n} \sum_s P(C_{m,s} = 1 | D_{\bullet,m}, \alpha_s^1, \alpha_s^2, \phi_s, \theta_s, \Omega_s) P(\alpha_s^1, \alpha_s^2, \phi_s, \theta_s, \Omega_s | D_{\bullet,m}) \quad (2)$$

Where the subscript s adjoined to a parameter name x means "the value of parameter x in sample s ".

Let's define $\mathcal{P}_{m,s}^1$ as:

$$\mathcal{P}_{m,s}^1 = P(C_{m,s} = 1 | D_{\bullet,m}, \alpha_s^1, \alpha_s^2, \phi_s, \theta_s, \Omega_s) \quad (3)$$

We can compute the probability $\mathcal{P}_{m,s}^1$ as follows :

$$\mathcal{P}_{m,s}^1 = \frac{P(D_{\bullet,m} | C_{m,s} = 1, \alpha_s^1, \alpha_s^2, \phi_s, \theta_s, \Omega_s) P(C_{m,s} = 1 | \alpha_s^1, \alpha_s^2, \phi_s, \theta_s, \Omega_s)}{P(D_{\bullet,m} | \alpha_s^1, \alpha_s^2, \phi_s, \theta_s, \Omega_s)} \quad (4)$$

$$= \frac{P(D_{\bullet,m} | C_{m,s} = 1, \alpha_s^1, \alpha_s^2, \phi_s, \theta_s, \Omega_s) P(C_{m,s} = 1 | \theta_s)}{\sum_{k=1,2} P(D_{\bullet,m} | C_{m,s} = k, \alpha_s^k, \phi_s, \theta_s, \Omega_s) P(C_{m,s} = k | \theta_s)} \quad (5)$$

$$= \frac{P(D_{\bullet,m} | C_{m,s} = 1, \alpha_s^1, \alpha_s^2, \phi_s, \theta_s, \Omega_s) \theta_s}{\sum_{k=1,2} P(D_{\bullet,m} | C_{m,s} = k, \alpha_s^k, \phi_s, \theta_s, \Omega_s) P(C_{m,s} = k | \theta_s)} \quad (6)$$

$$= \frac{\prod_t P(D_{t,m} | C_{m,s} = 1, \alpha_s^1, \alpha_s^2, \phi_s, \theta_s, \Omega_s) \theta_s}{\sum_{k=1,2} \prod_t P(D_{t,m} | C_{m,s} = k, \alpha_s^k, \phi_s, \theta_s, \Omega_s) P(C_{m,s} = k | \theta_s)} \quad (7)$$

It follows that $\mathcal{P}_{m,s}^2 = 1 - \mathcal{P}_{m,s}^1$.

For a given sample s , one can get a posterior-decoded region-specific estimate of $\alpha_{m,s}$ in two ways: either by summing, or by drawing from a Bernoulli distribution.

Summing involves computing:

$$\alpha_{m,s} = \sum_{k=1,2} \mathcal{P}_{m,s}^k \times \alpha_s^k \quad (8)$$

Our implementation makes use of the second approach, which involves first drawing a k value and then choosing the α_s^k accordingly:

$$k = 1 + \mathcal{B}(\mathcal{P}_{m,s}^1) \quad (9)$$

$$\alpha_{m,s} = \alpha_s^k \quad (10)$$

The two approaches converge to the same estimates as the number n of samples s approaches infinity.

Other region-wise parameters or estimates x that differ between the two mixture categories can be posterior-decoded using the same formulas, only changing the α_s^k above into x_s^k .

	region	Category 1	Category 2
1	auvergne.rhone.alpes	0.77	0.23
2	bourgogne.franche.comte	0.69	0.31
3	bretagne	0.55	0.45
4	centre.val.de.loire	0.64	0.36
5	corse	0.39	0.61
6	grand.est	0.57	0.43
7	hauts.de.france	0.93	0.07
8	ile.de.france	0.22	0.78
9	normandie	0.53	0.47
10	nouvelle.aquitaine	0.49	0.51
11	occitanie	0.81	0.19
12	pays.de.la.loire	0.61	0.39
13	provence.alpes.cote.dazur	0.68	0.32

Table 1: Posterior probability for each region to belong to either categories.

2.2 Results

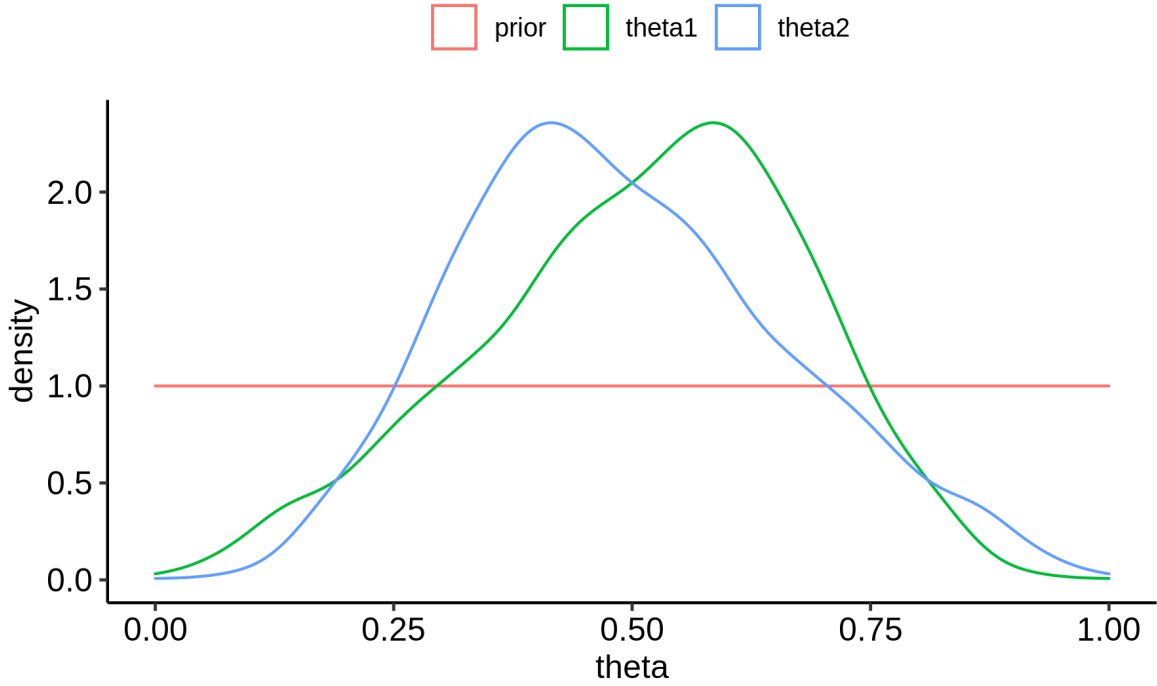
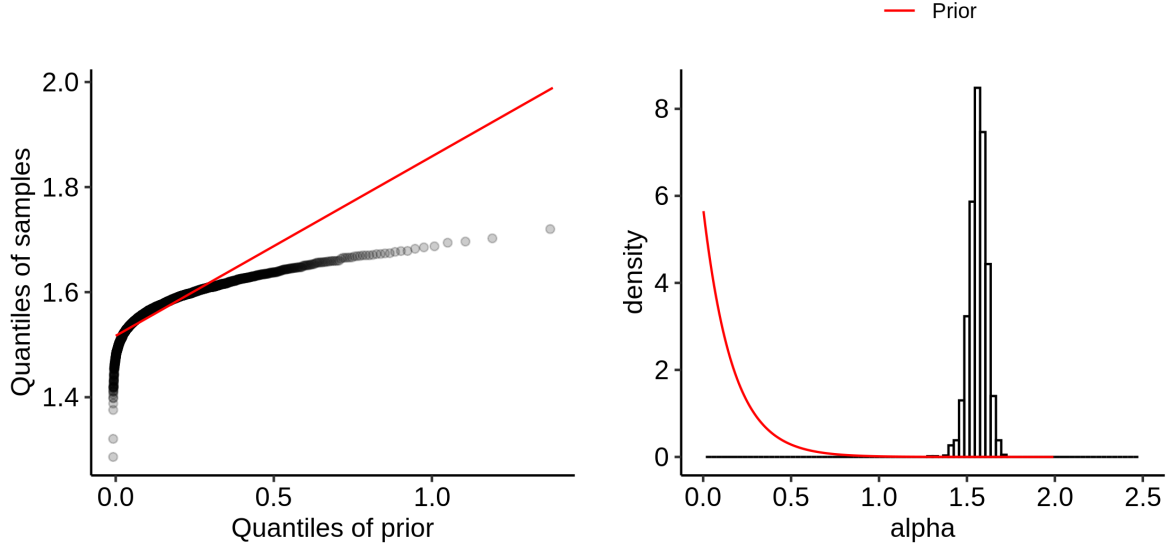
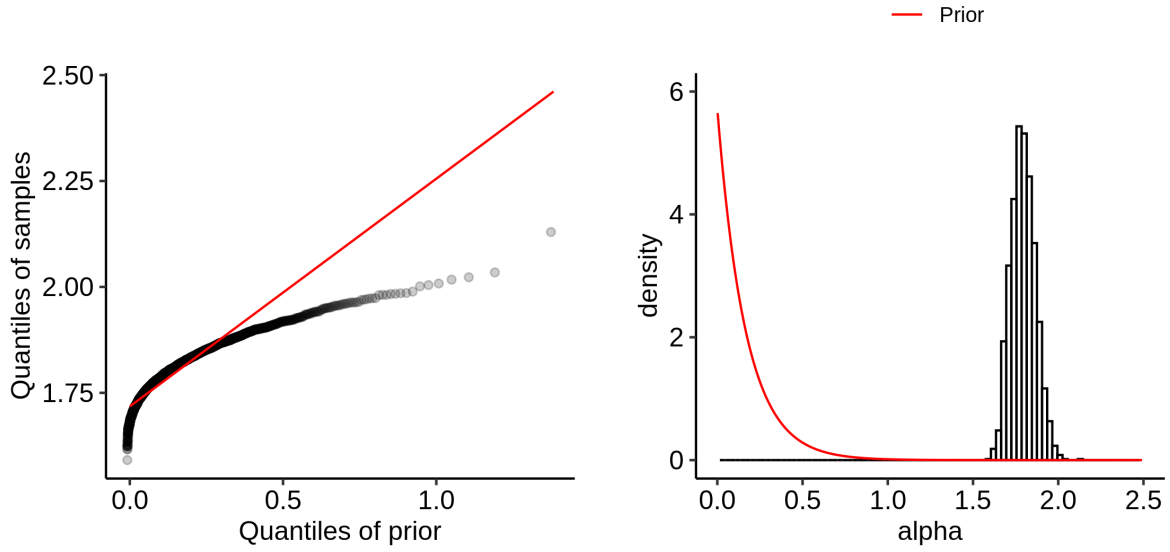


Figure 2: Posterior density of mixture proportions parameter θ



(a) Posterior distribution of α^1



(b) Posterior distribution of α^2

Figure 3: Posterior distributions of α parameter in both mixture conditions.

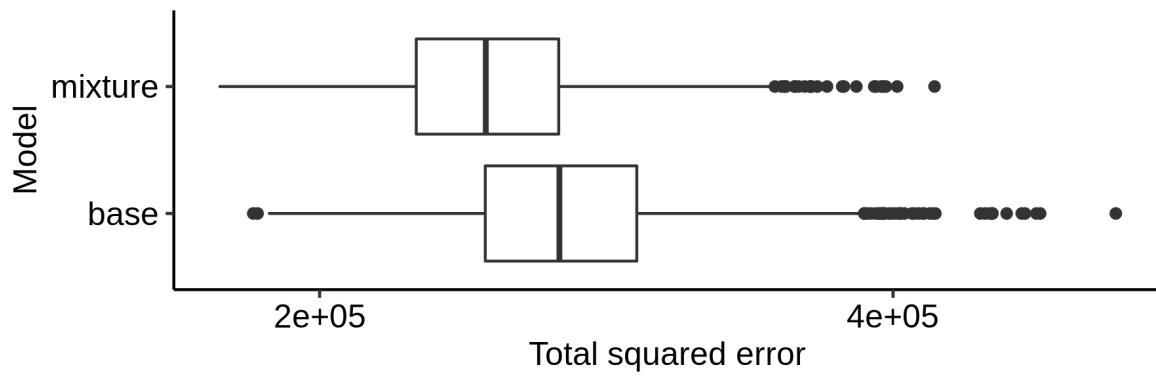


Figure 4: Comparison of total squared error between model fits up to May 11.

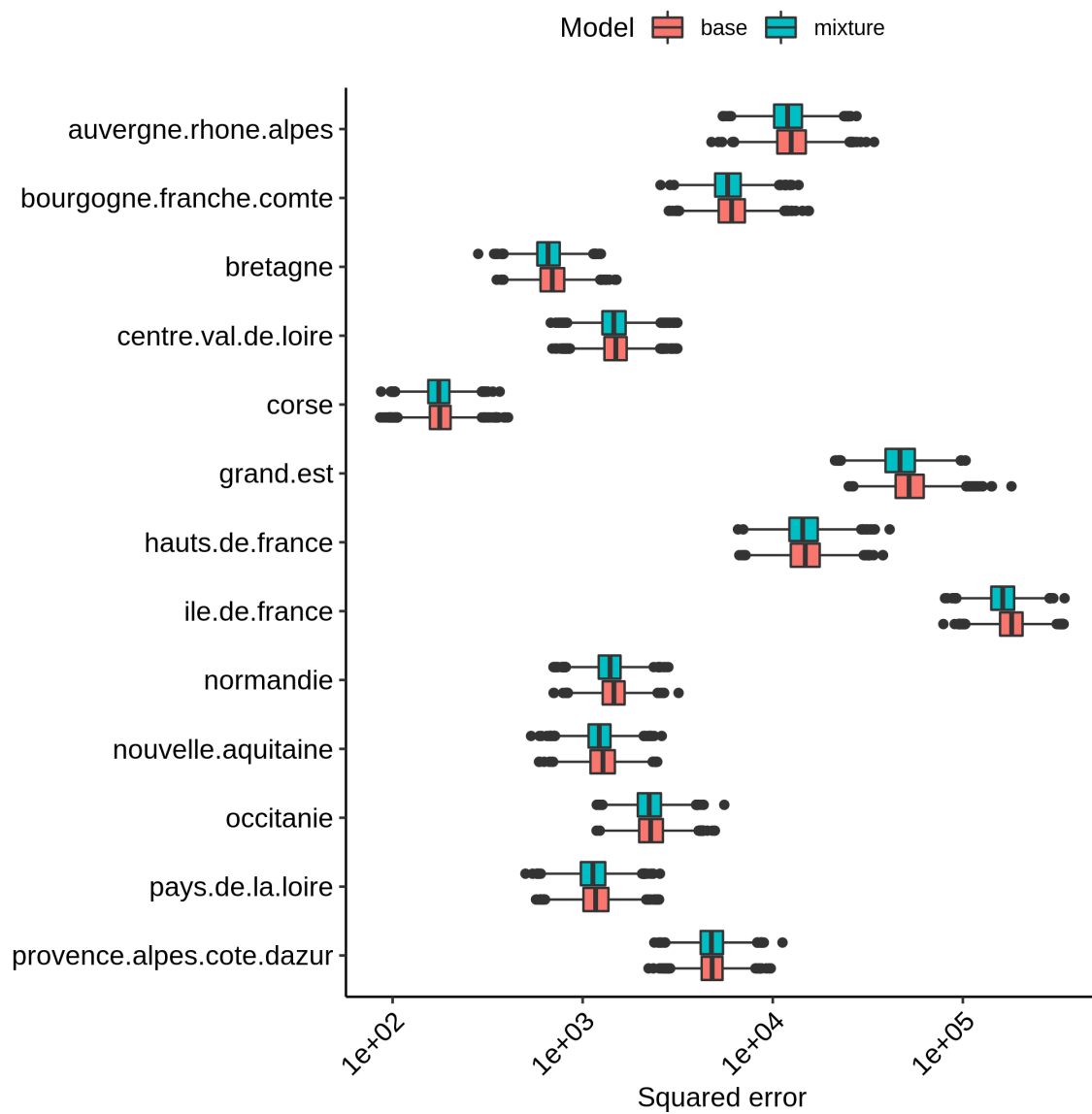


Figure 5: Total squared error per region of model fits up to May 11.

3 Posterior distributions of α parameters

The $\alpha_{lockdown}$ parameter is used to compute the reproduction number during the days when the lockdown is in place.

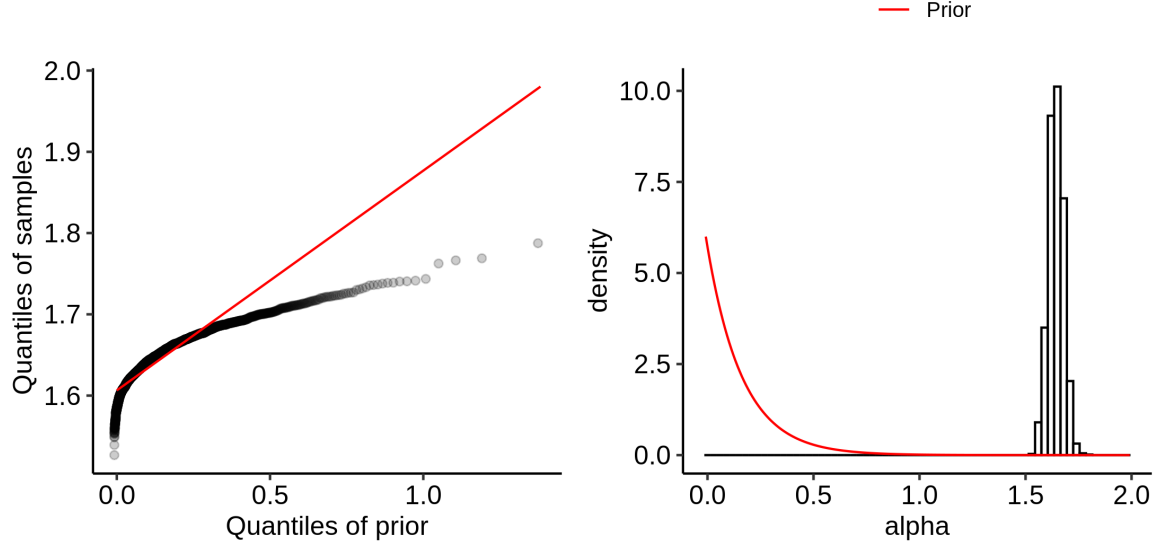


Figure 6: The posterior distribution of $\alpha_{lockdown}$ differs from its shifted Gamma prior. Left: Quantile-quantile plot between the prior and the posterior distributions; the line $y = x$ is in red. Right: Density plots of the prior (red line) and posterior distributions (black histogram).

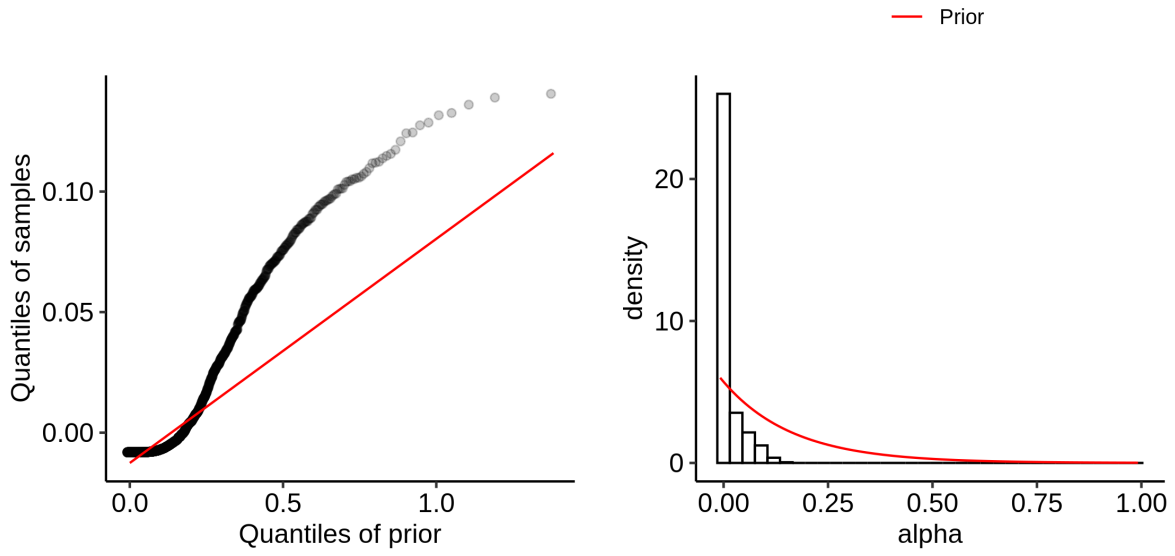


Figure 7: Posterior distribution of $\alpha_{weekend}$ is shifted towards values close to 0 compared to the prior distribution.

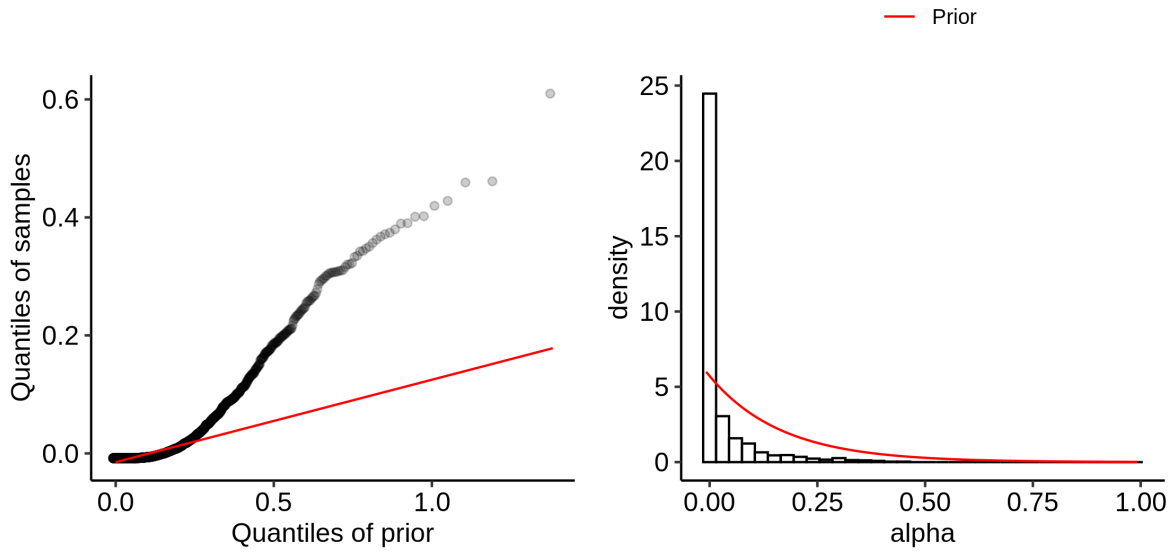
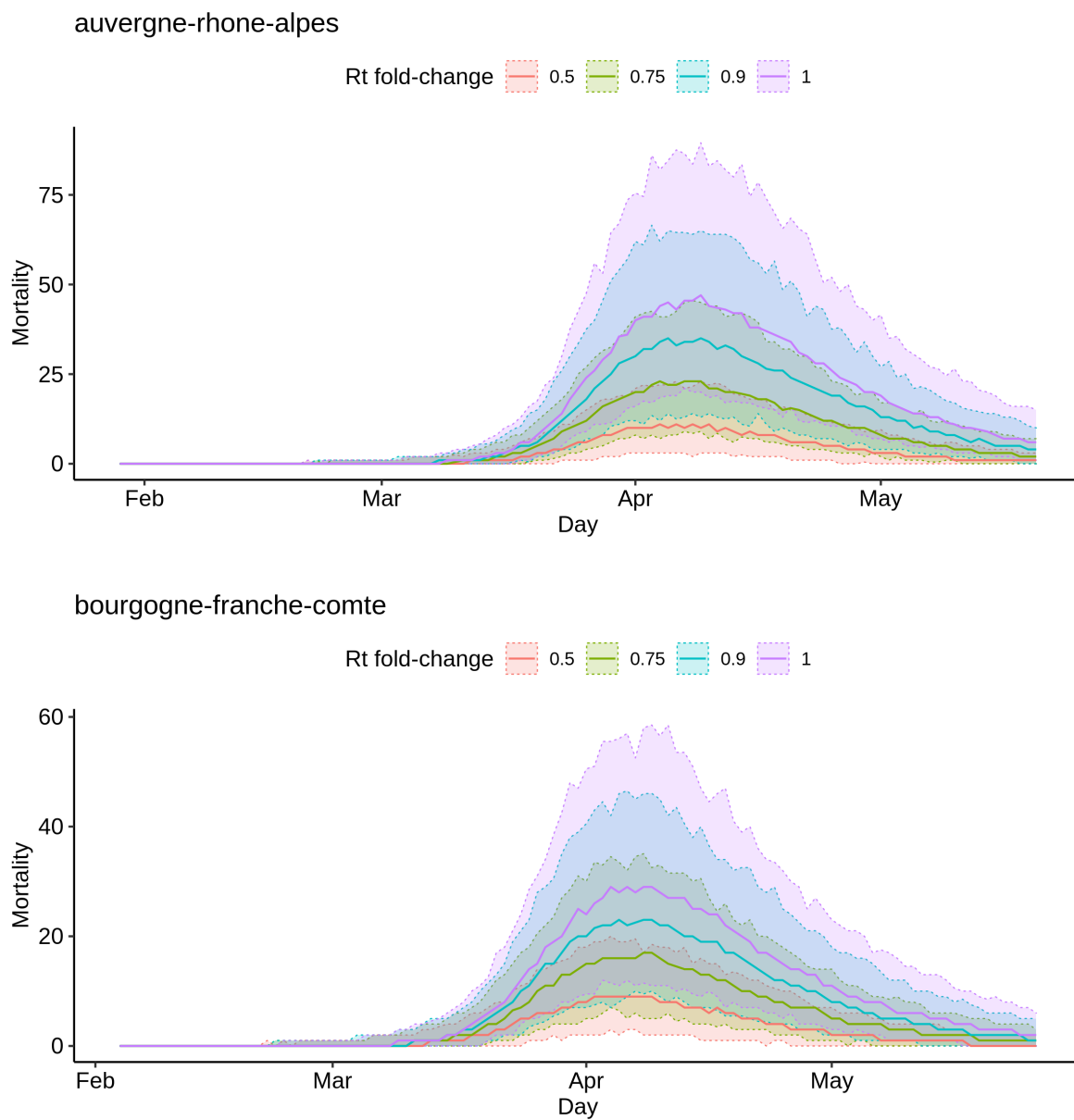


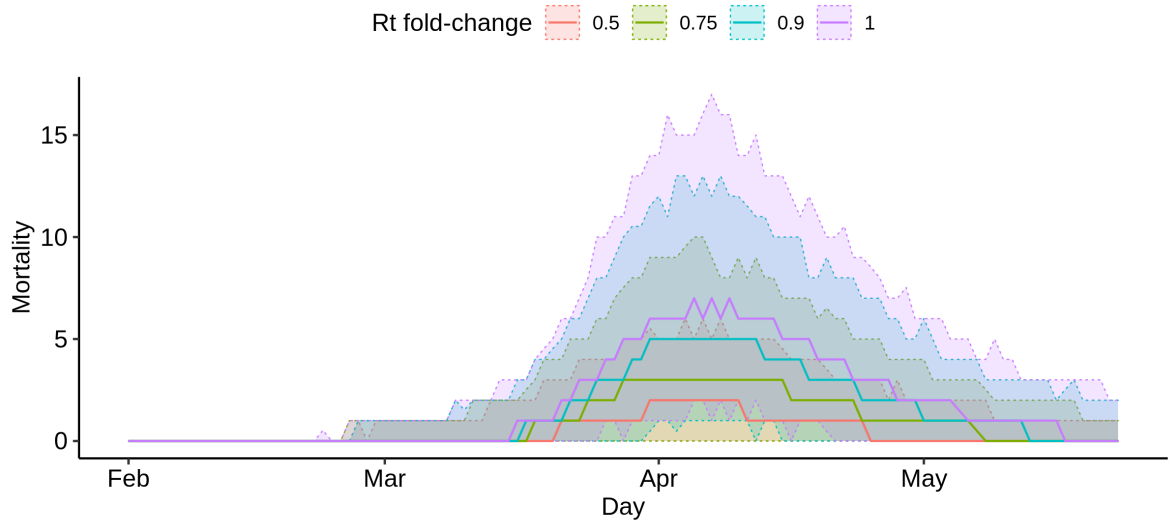
Figure 8: Posterior distribution of $\alpha_{elections}$ is shifted towards values close to 0 compared to the prior distribution.

4 Region-wise simulations for week-ends effect size

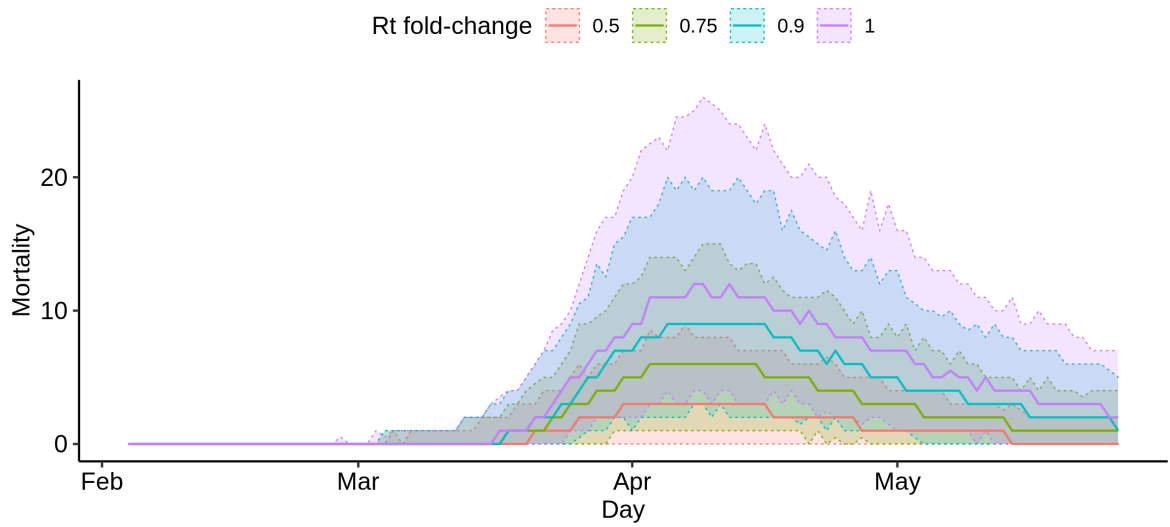
Figure 9: Region-wise simulations of daily mortality, assuming a range of R_t fold-changes during week-ends.

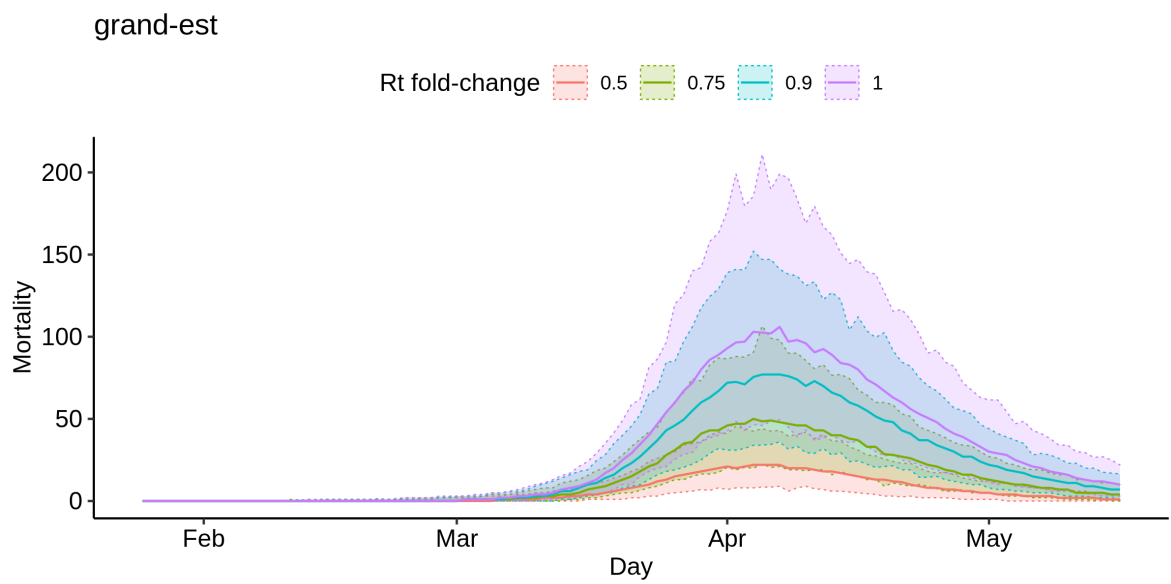
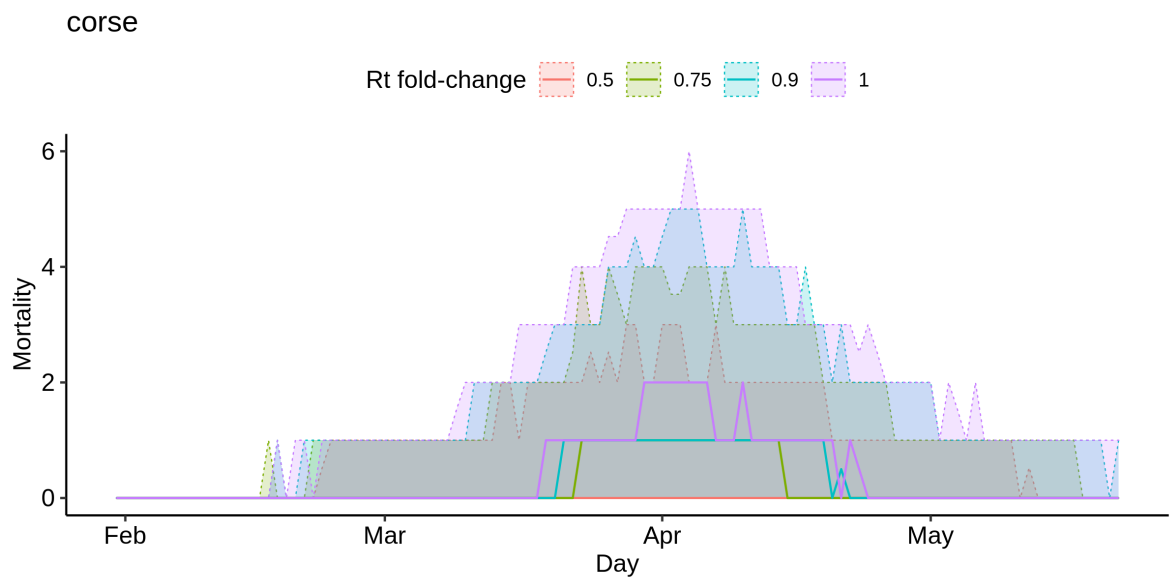


bretagne

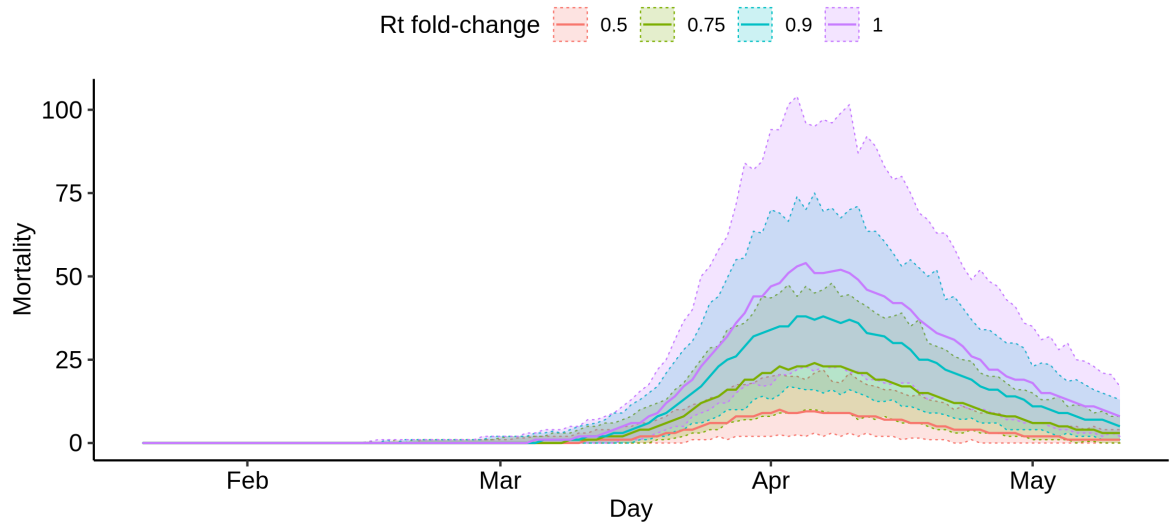


centre-val-de-loire

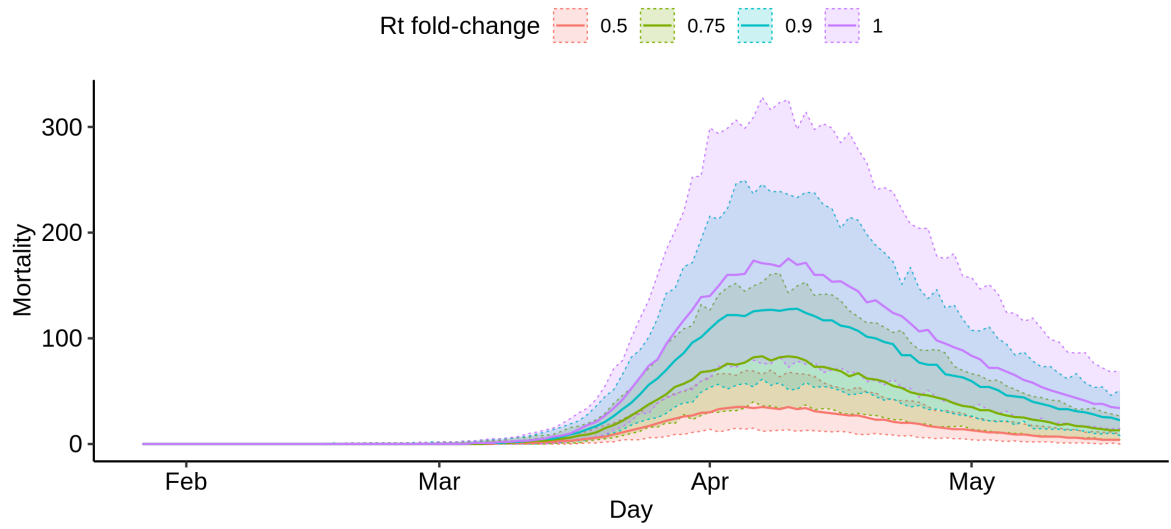




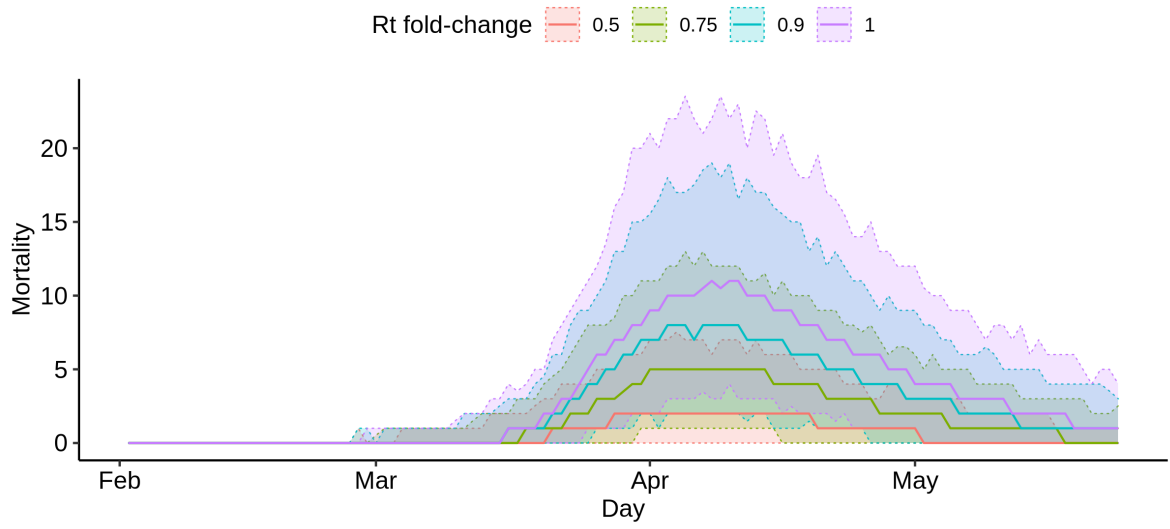
hauts-de-france



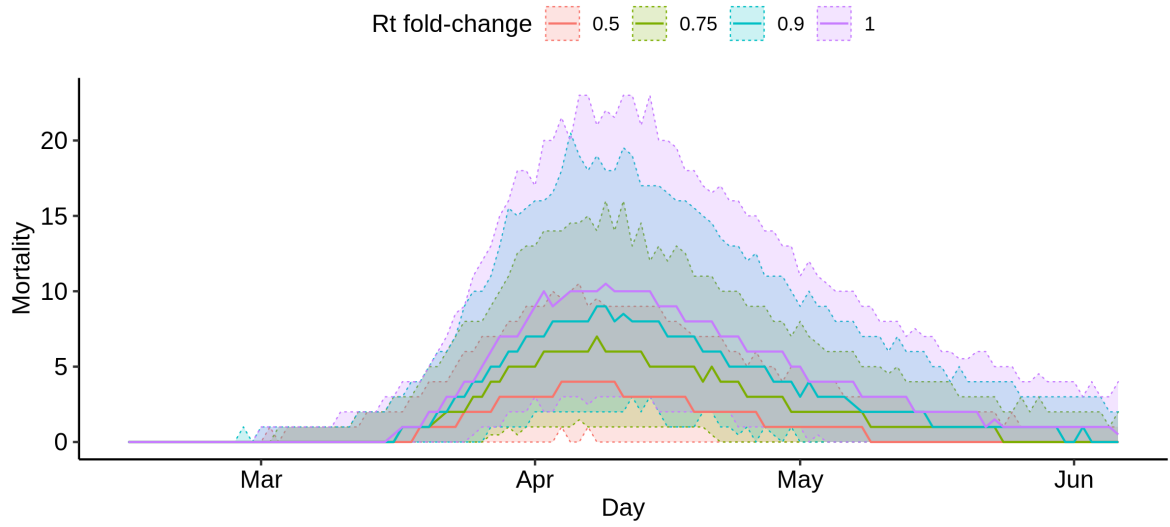
ile-de-france



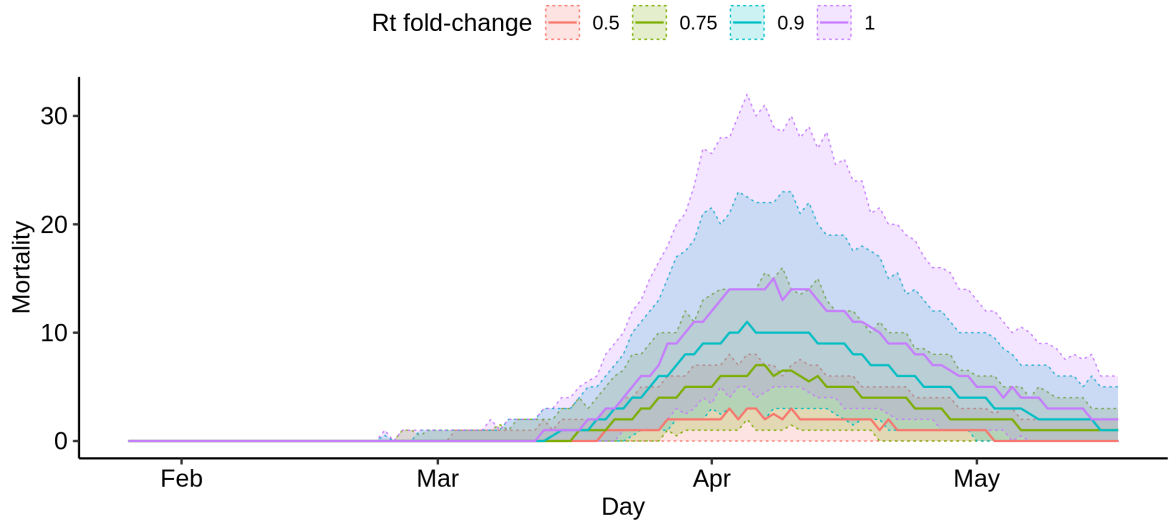
normandie



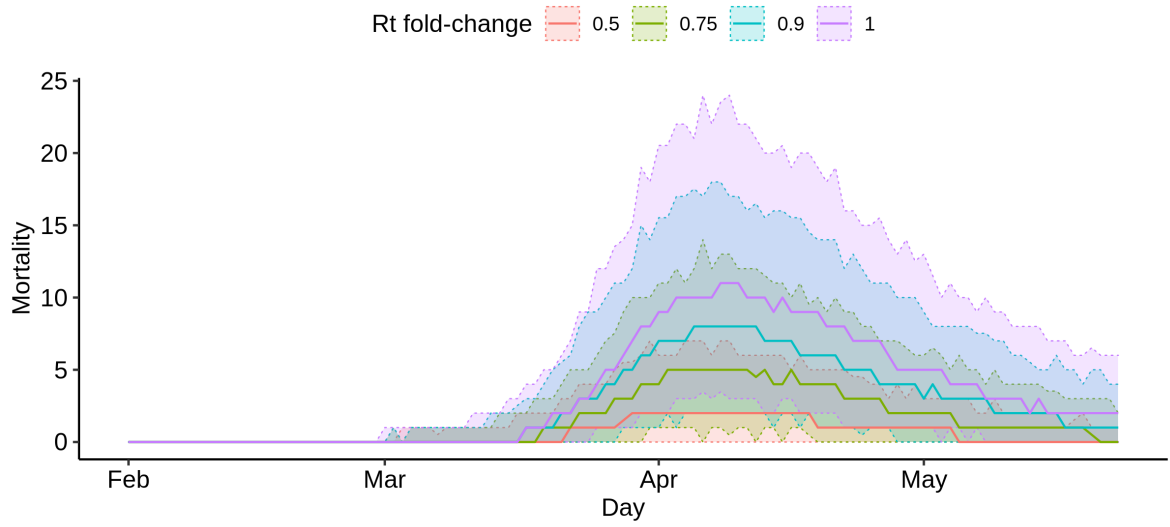
nouvelle-aquitaine



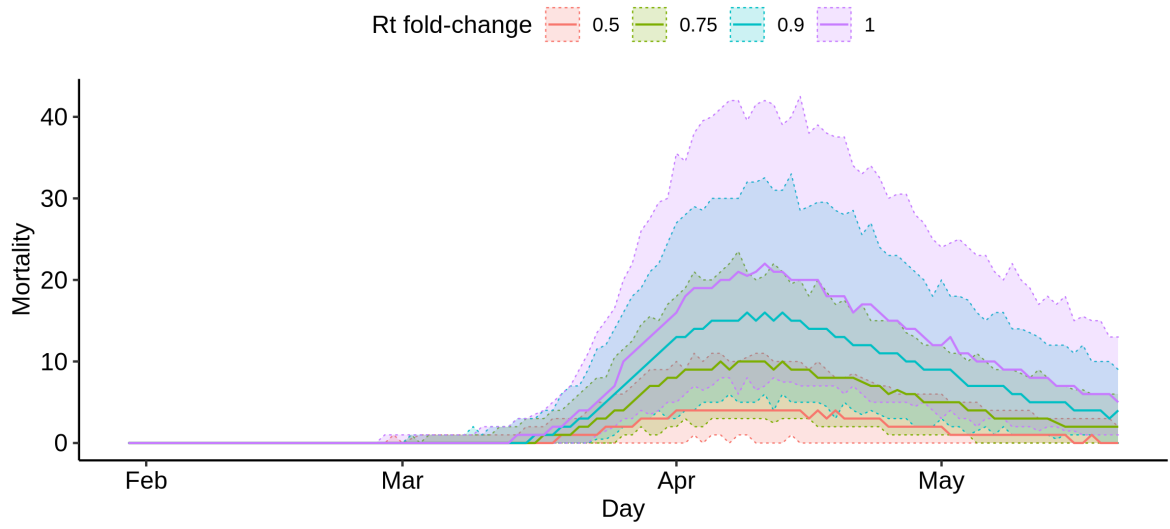
occitanie



pays-de-la-loire

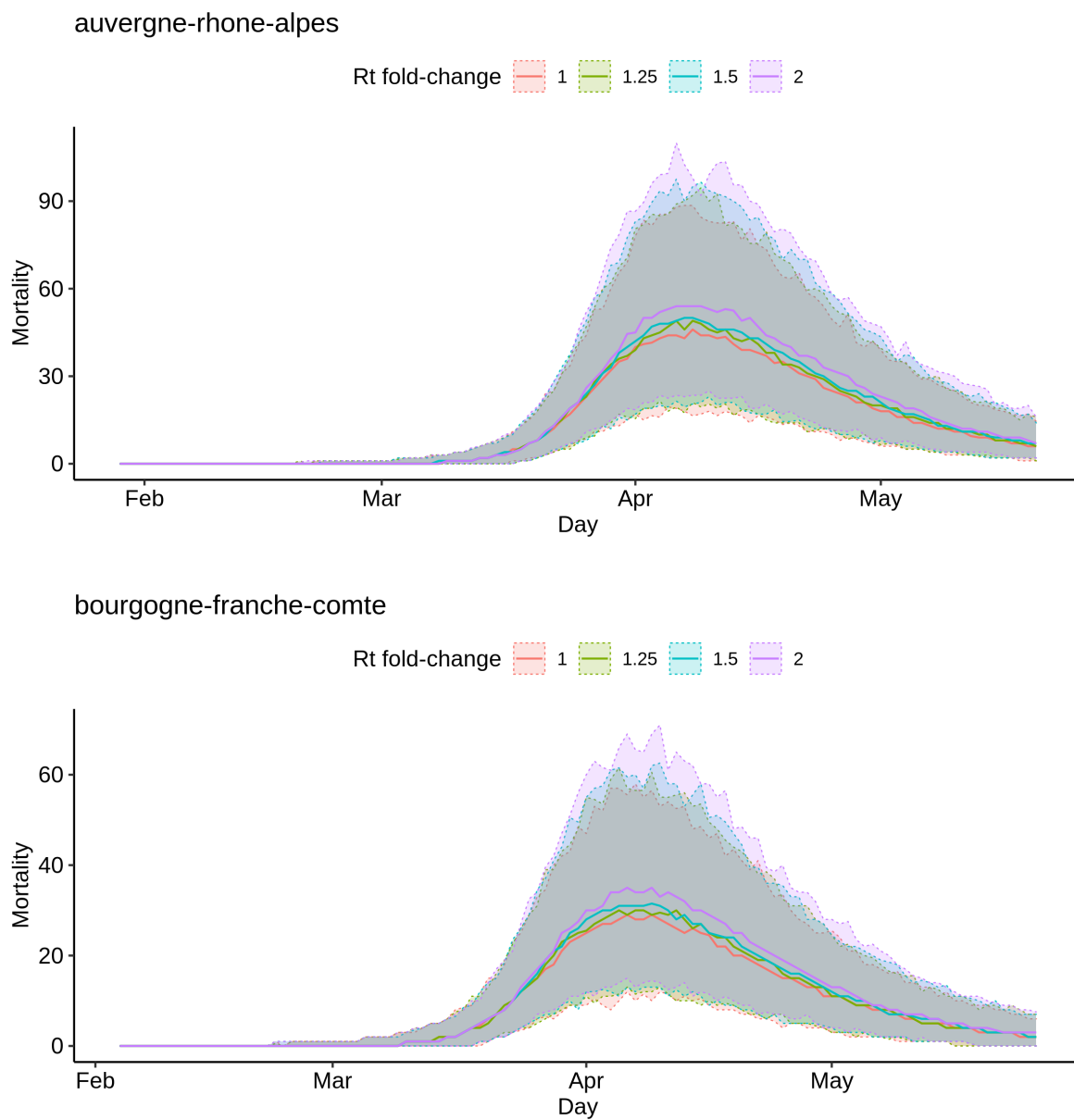


provence-alpes-cote-dazur

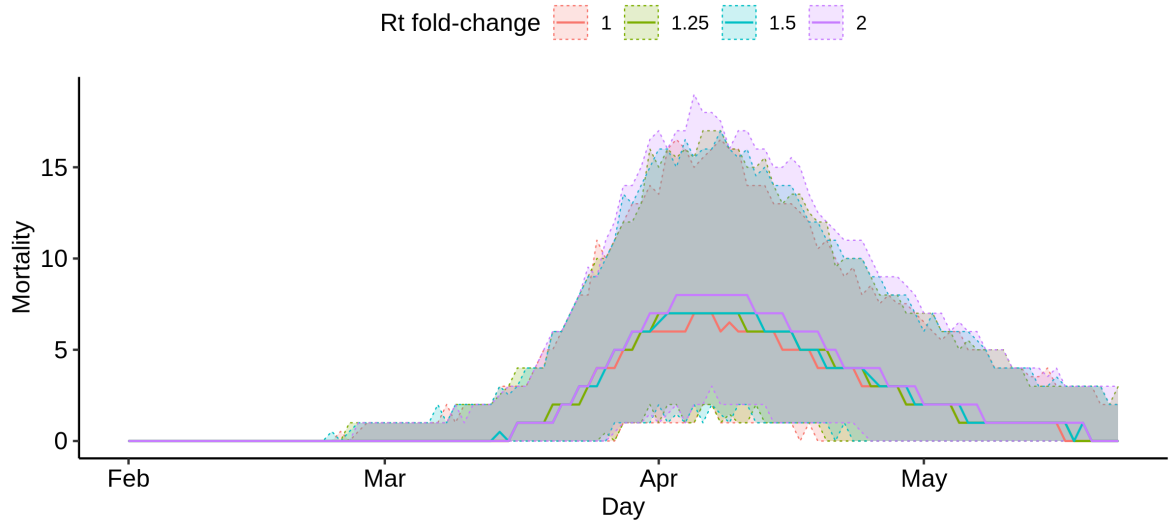


5 Region-wise simulations for elections effect size

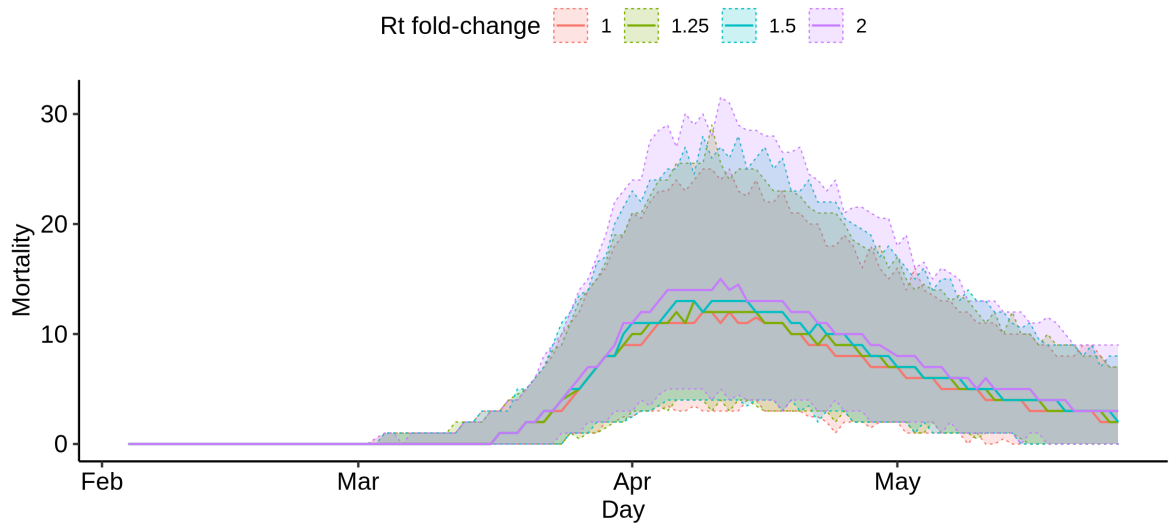
Figure 10: Region-wise simulations of daily mortality, assuming a range of R_t fold-changes during elections.

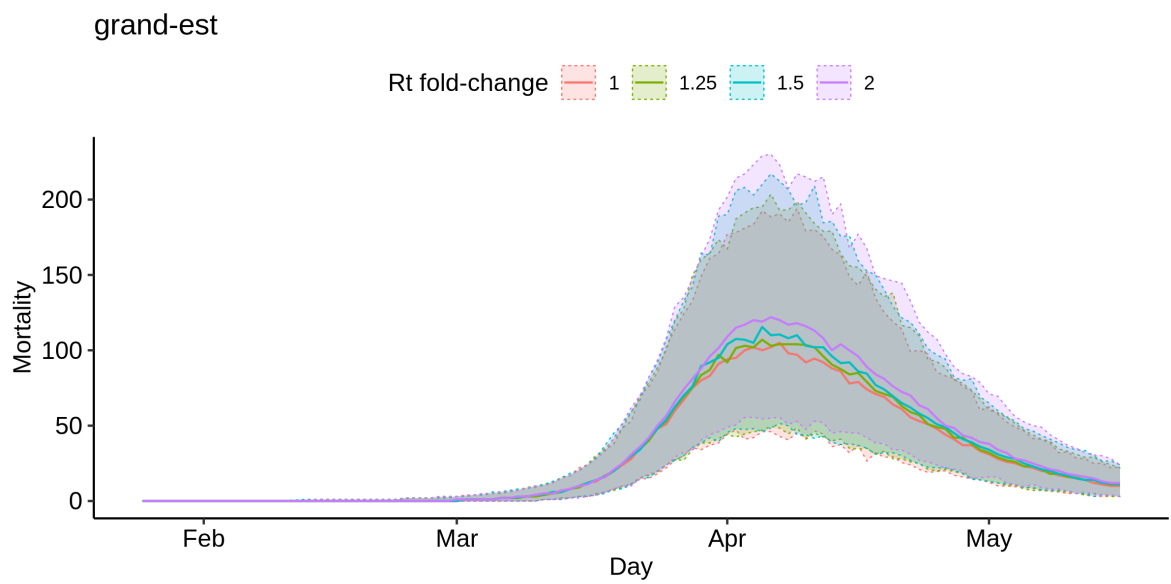
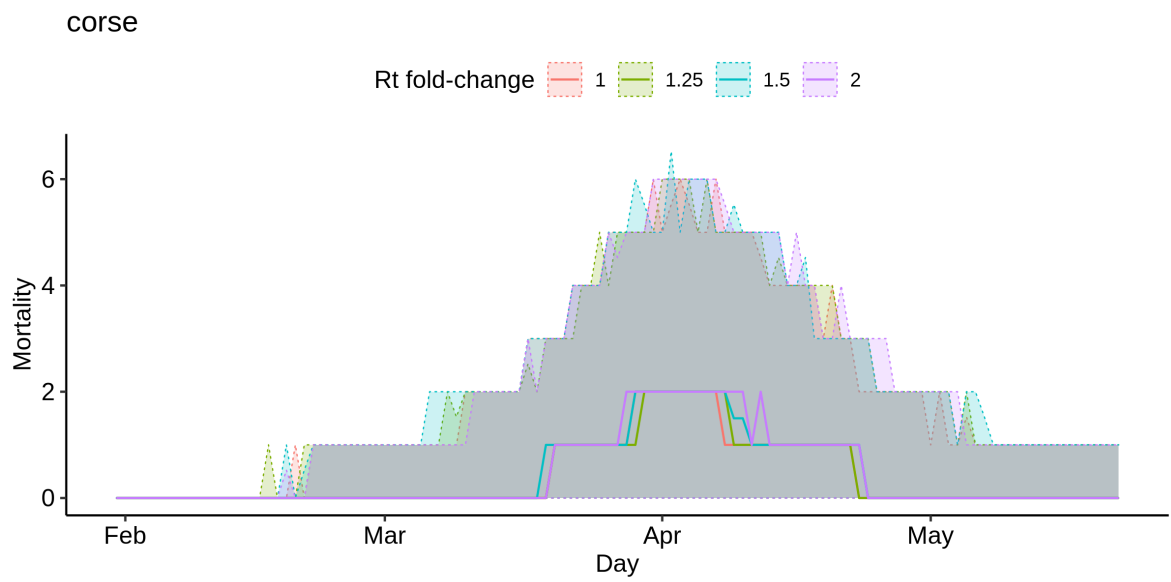


bretagne

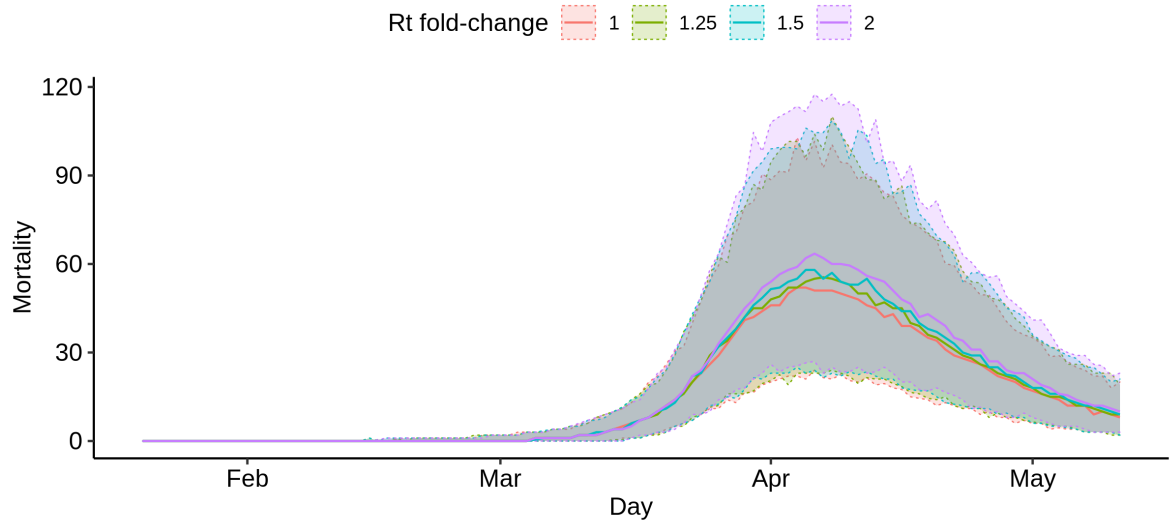


centre-val-de-loire

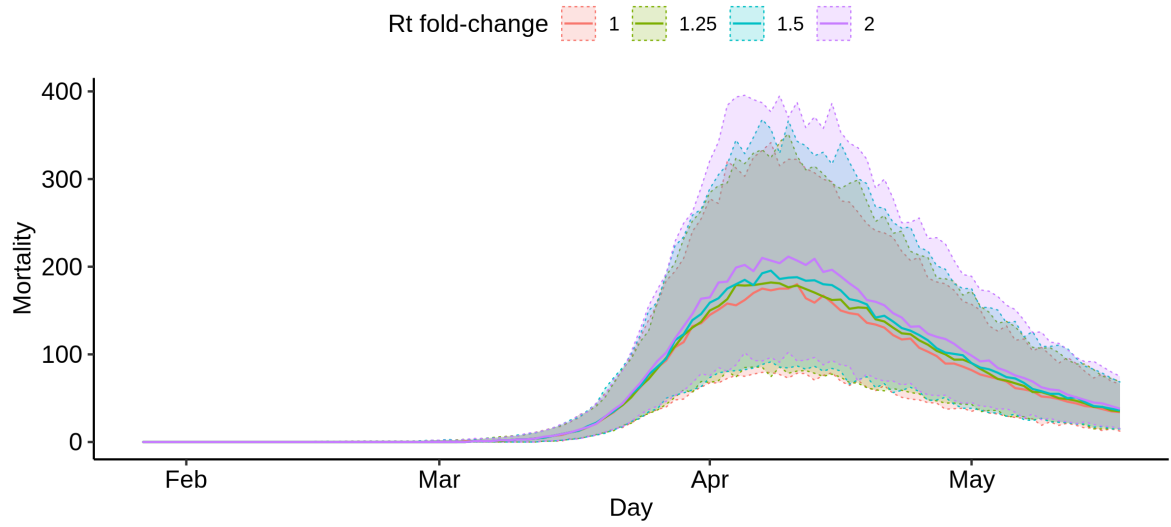




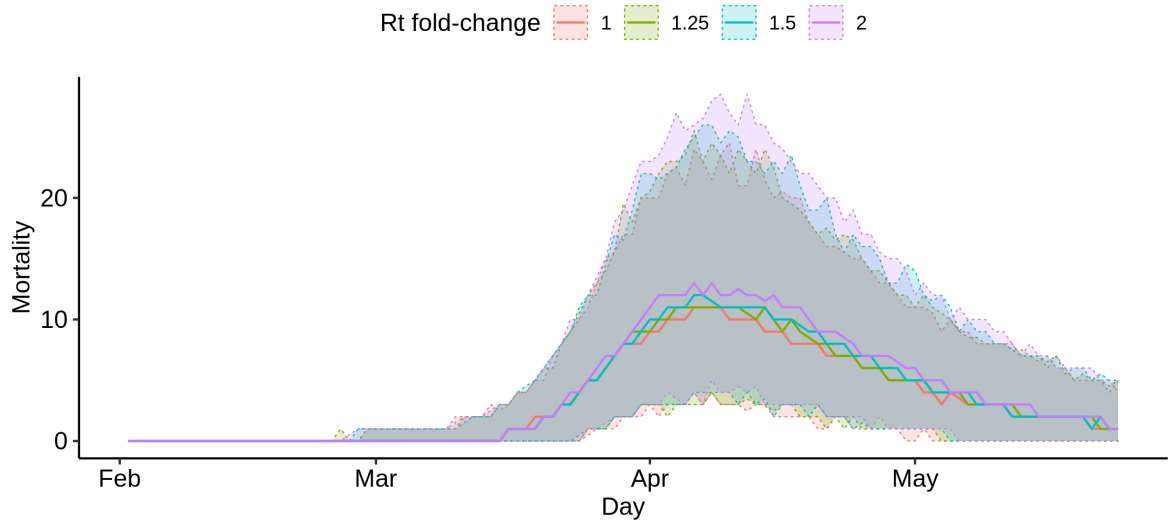
hauts-de-france



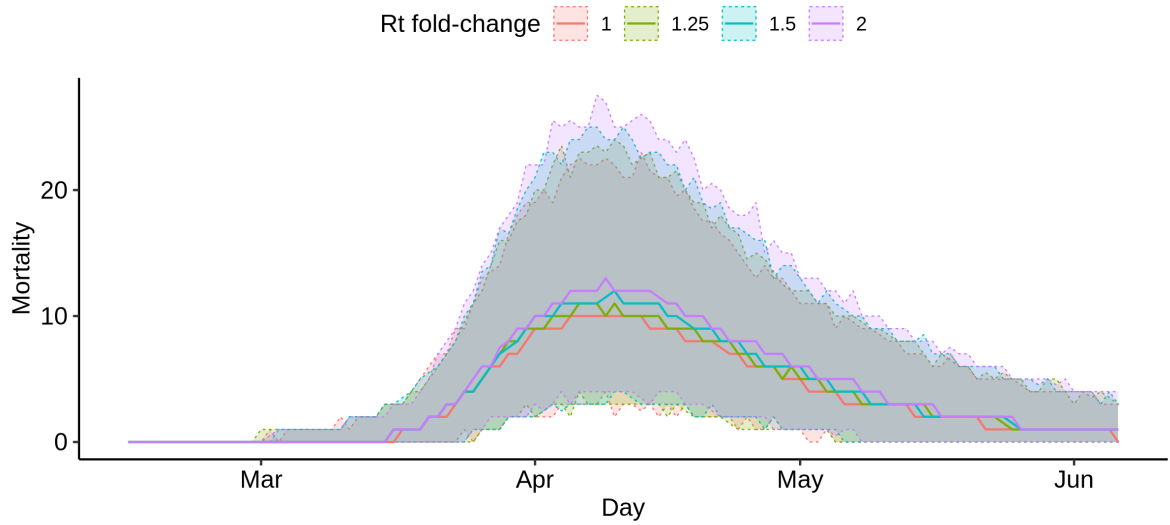
ile-de-france



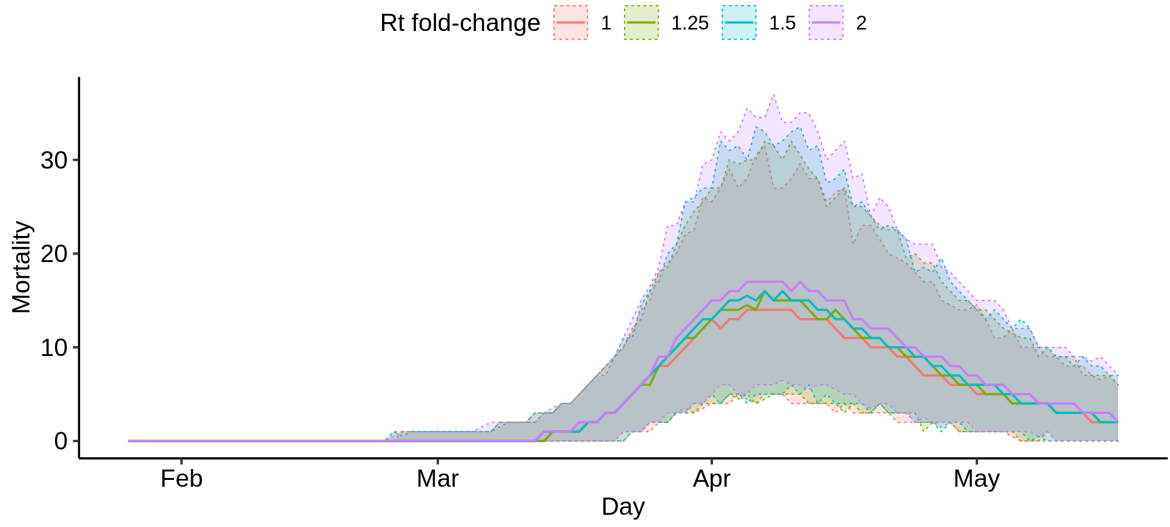
normandie



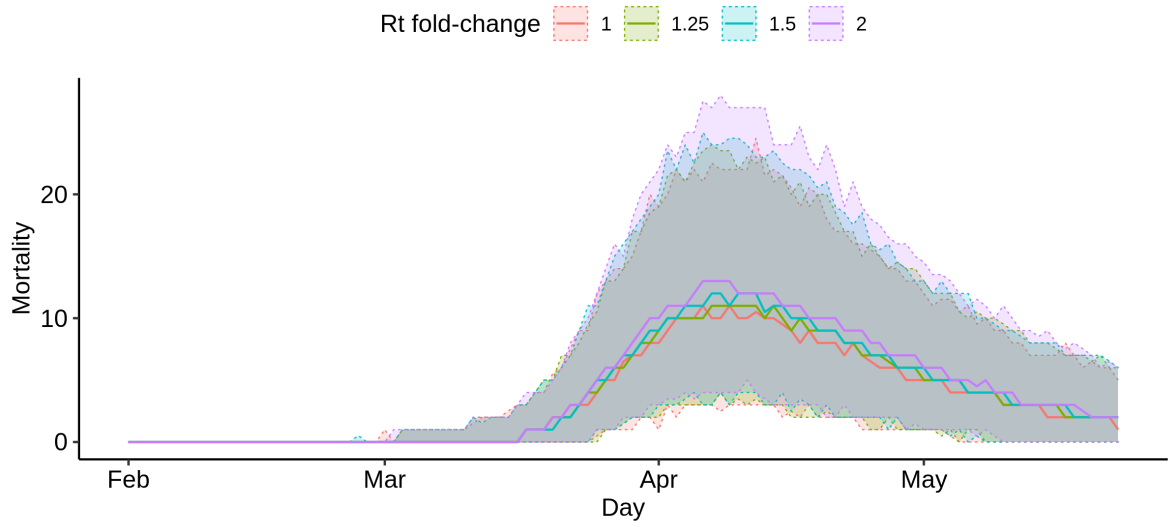
nouvelle-aquitaine



occitanie



pays-de-la-loire



provence-alpes-cote-dazur

