Supplementary material

Figures

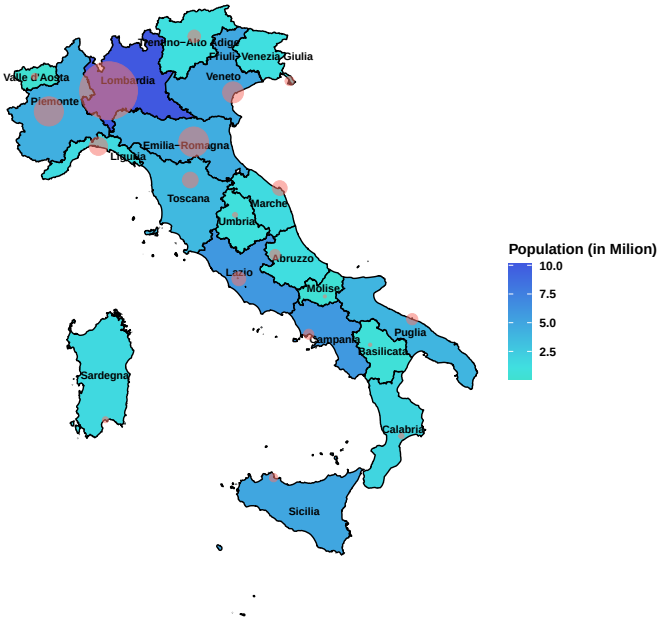

Figure S1. Italian regions colored according to the population size and with bubbles proportional to the number of COVID19-related deaths observed from August 1st 2020 to January 14th 2021, as reported by the Protezione Civile[1]

.

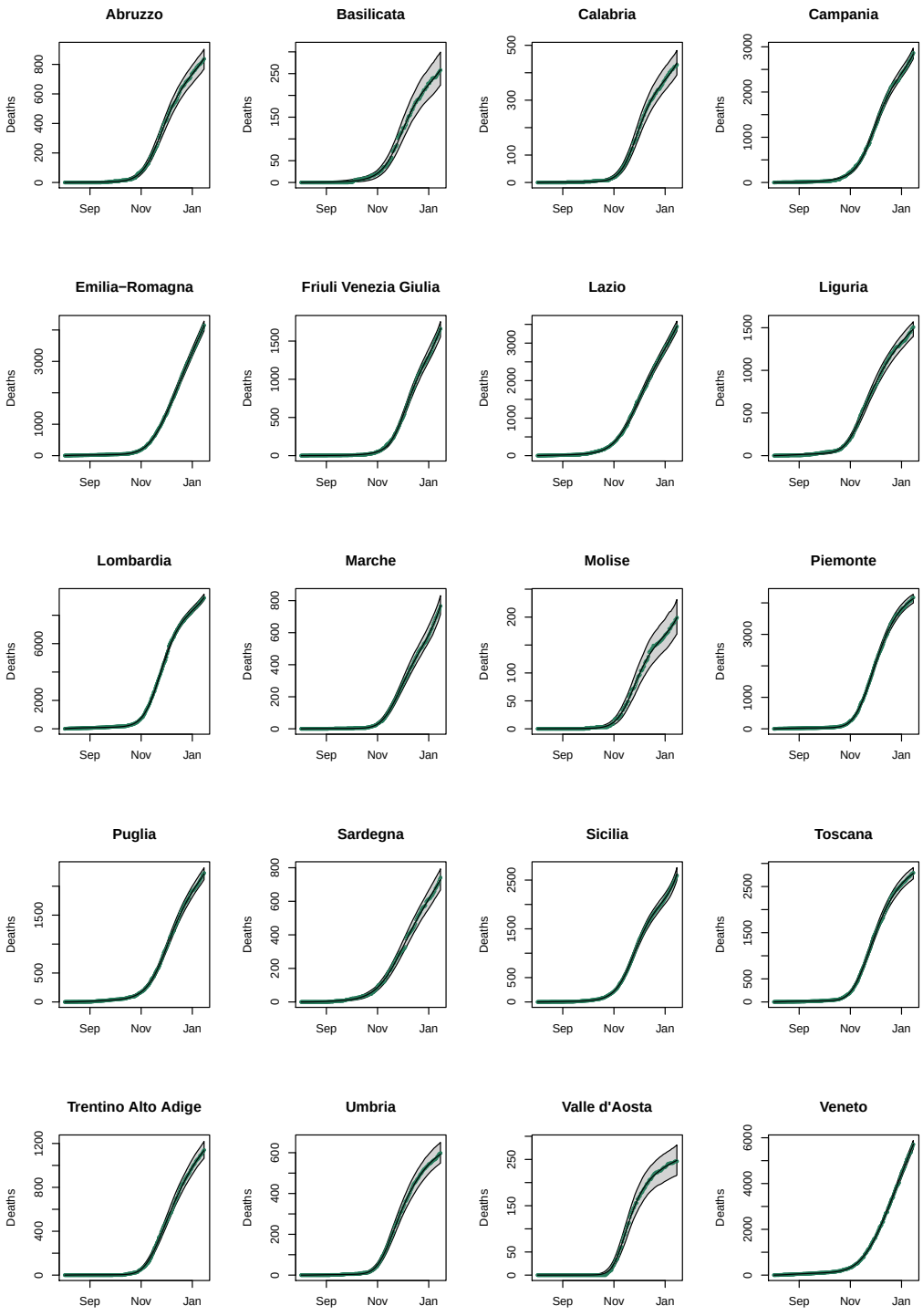

Figure S2. Estimated number of COVID19-related deaths with pointwise 90% confidence bands and observed number of COVID-19 deaths (green points), by region; $p = 1.15\%$, $T = 14$ days.

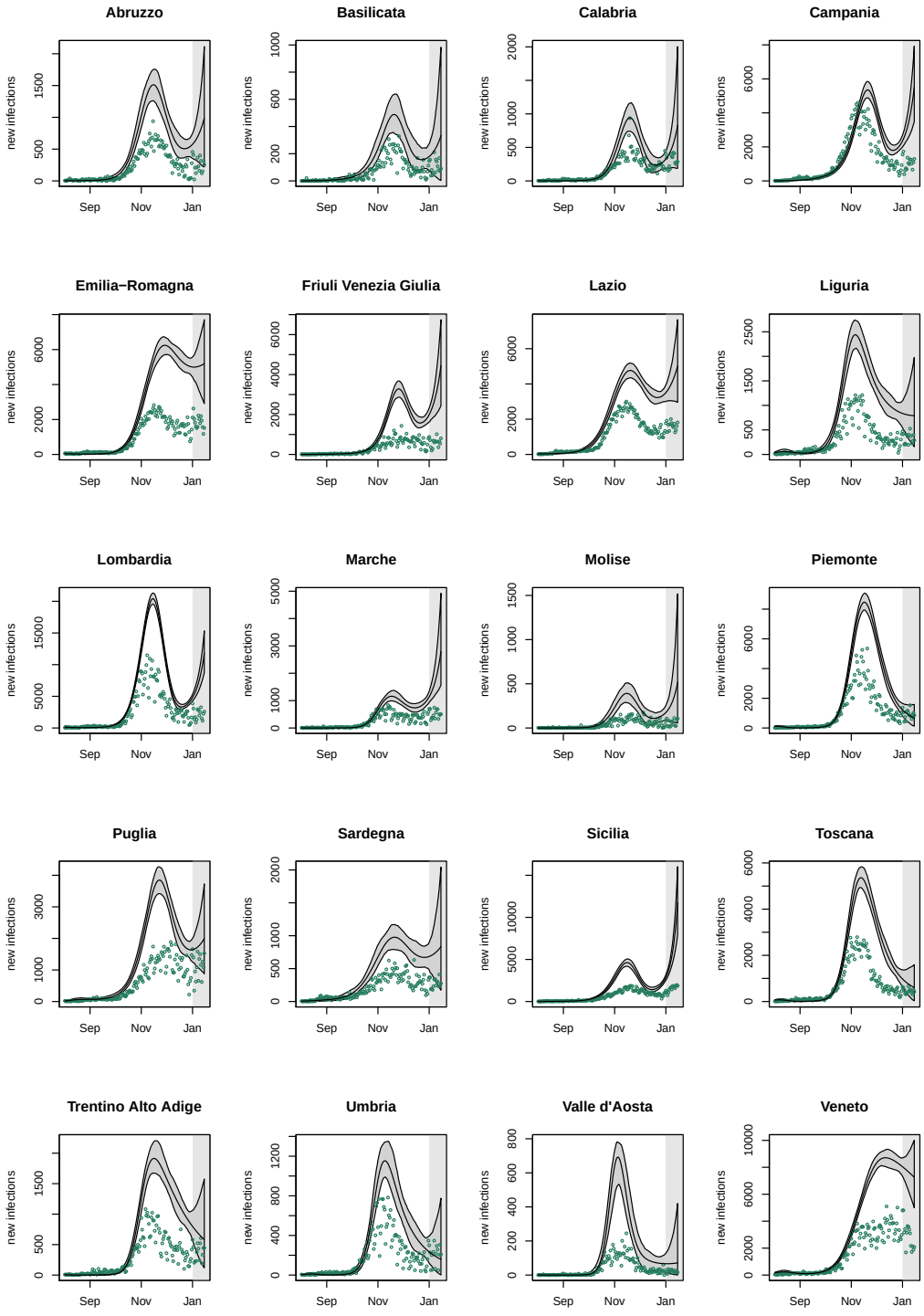

Figure S3. Estimated number of new infections with pointwise 90% confidence bands and observed number of new notified infections (green points), by region; $p = 1.15\%$, $T = 14$ days.

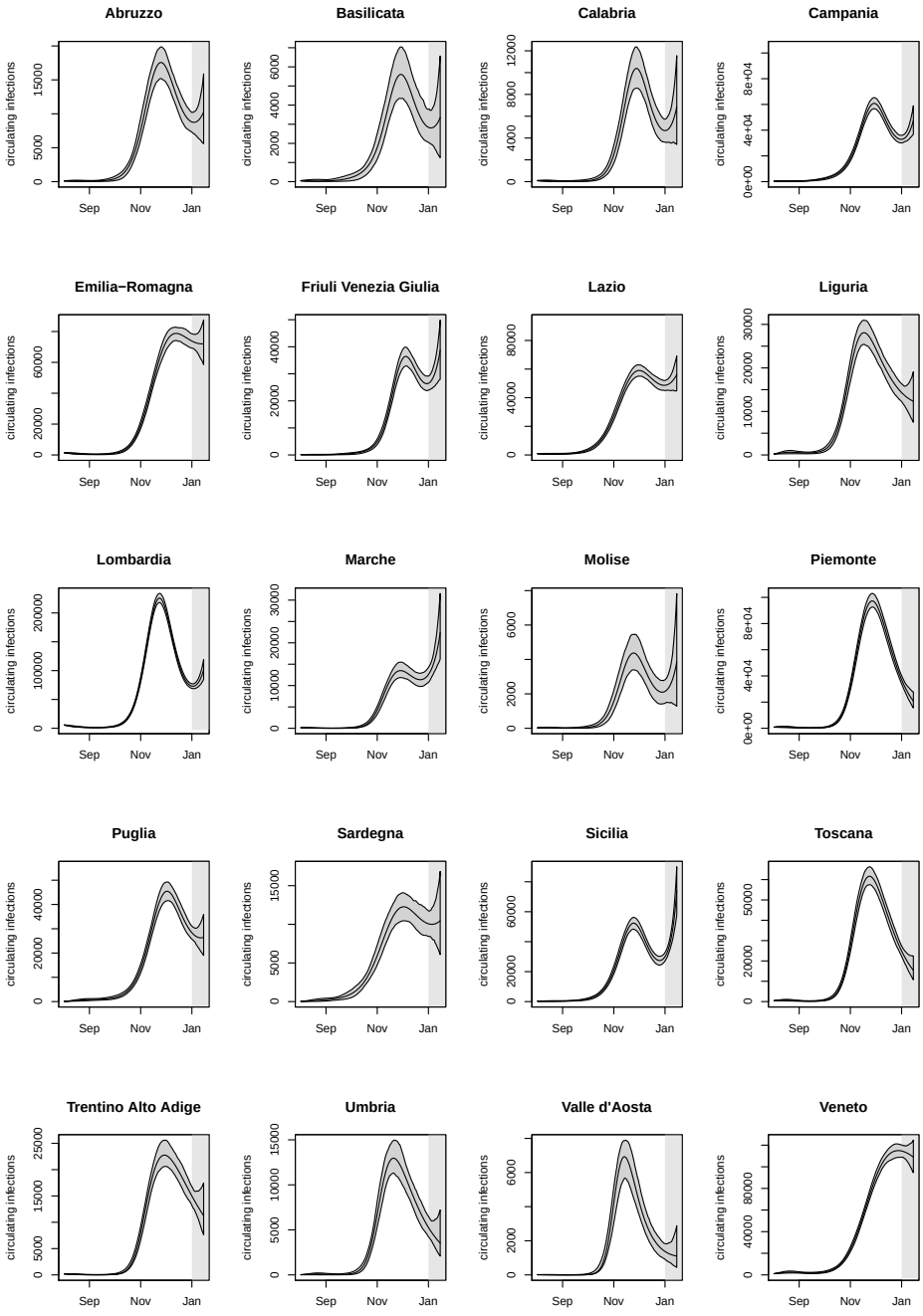

Figure S4. Estimated number of prevalent infections with pointwise 90% confidence bands, by region; $p = 1.15\%$, $T = 14$ days.

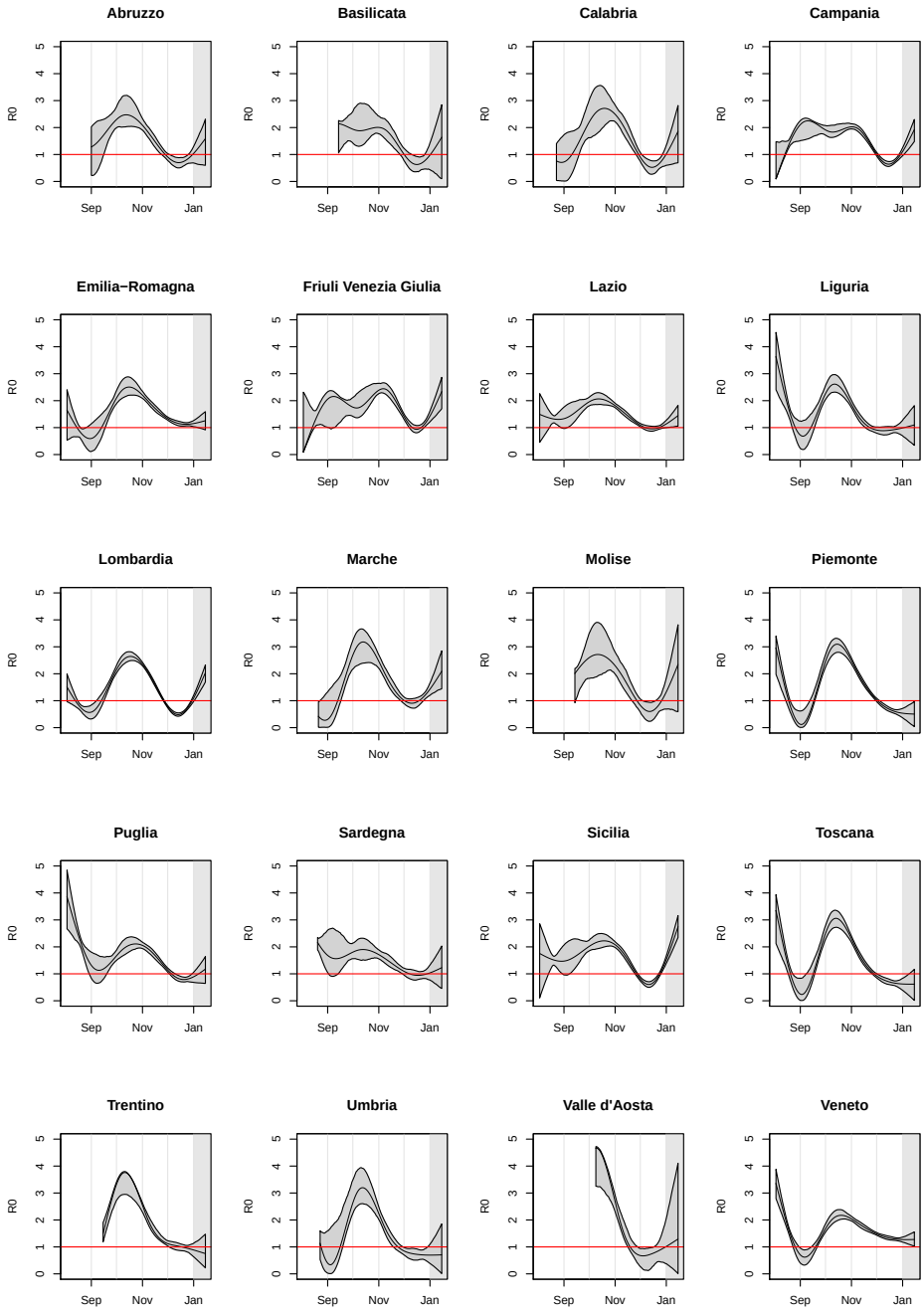

Figure S5. Estimated $R_0(t)$ with pointwise 90% confidence bands, by region; $p = 0.5\%$, $T = 14$ days.

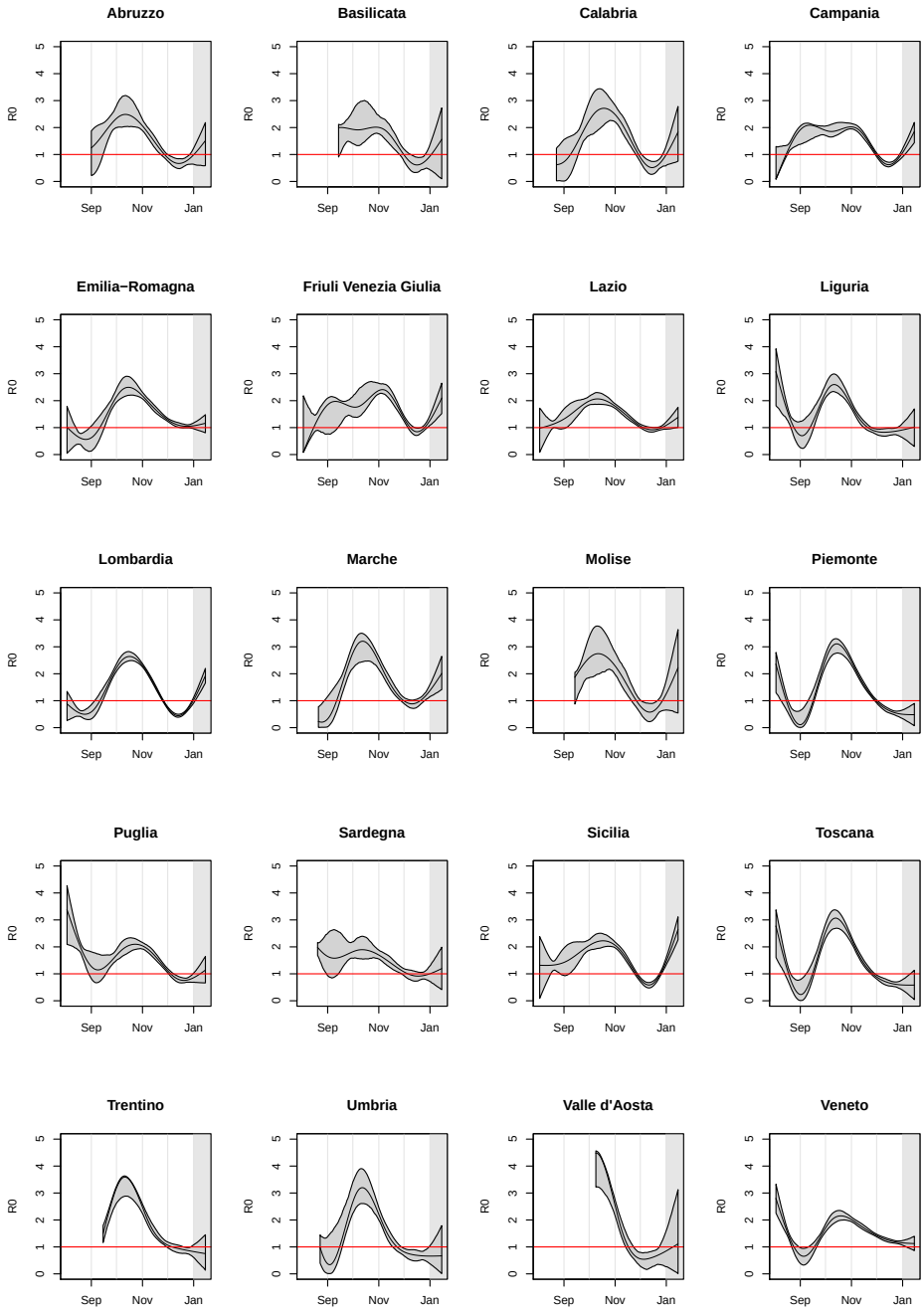

Figure S6. Estimated $R_0(t)$ with pointwise 90% confidence bands, by region; $p = 0.78\%$, $T = 14$ days.

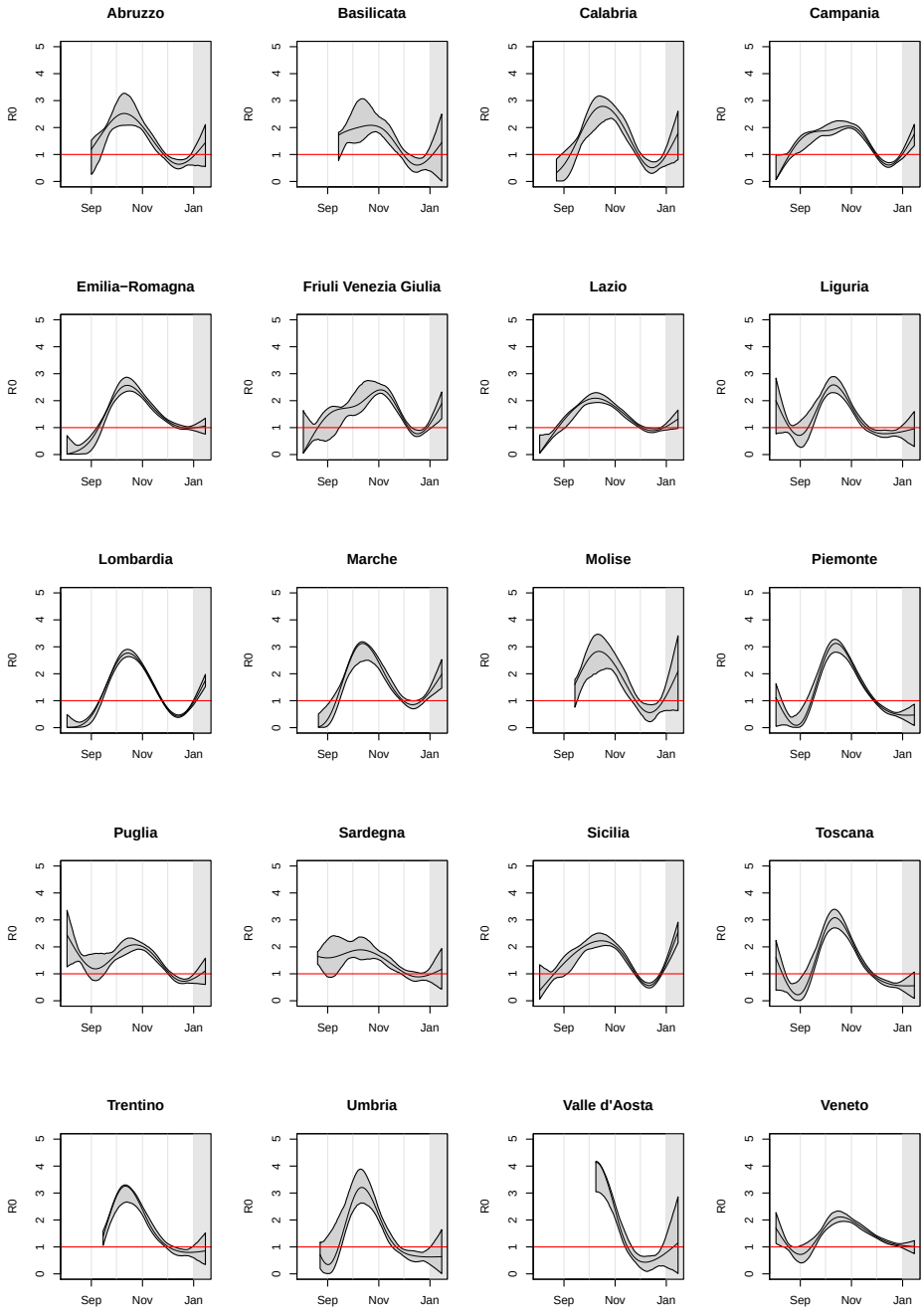

Figure S7. Estimated $R_0(t)$ with pointwise 90% confidence bands, by region; $p = 1.79\%$, $T = 14$ days.

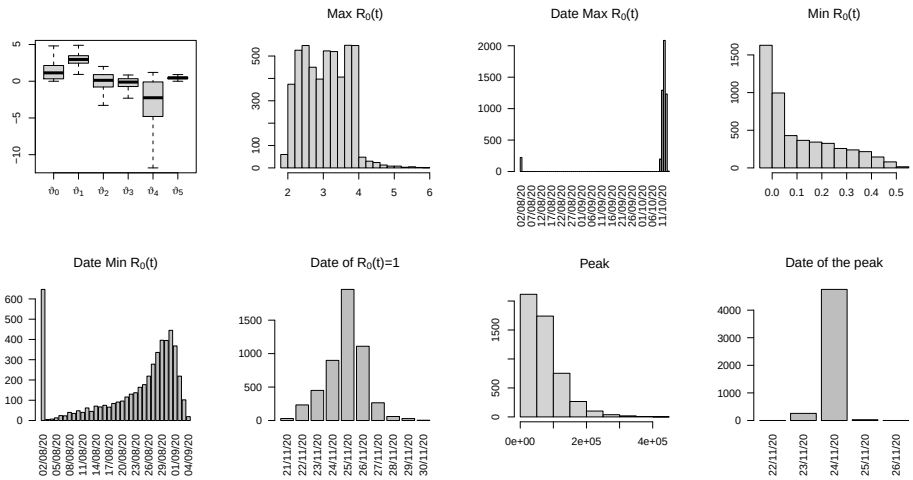

Figure S8. Monte Carlo distributions of the model outputs as the inputs vary obtained from the GSA.

Tables

Region	Population size	Deaths	i_0
Abruzzo	1305770	741	111
Basilicata	556934	228	50
Calabria	1924701	375	97
Campania	5785861	2409	395
Emilia-Romagna	4467118	3295	1496
Friuli-Venezia Giulia	1211357	1297	122
Lazio	5865544	2906	945
Liguria	1543127	1324	200
Lombardia	10103969	8317	6011
Marche	1518400	584	147
Molise	302265	168	27
Trentino	1074819	984	202
Piemonte	4356406	3793	799
Puglia	4008296	1920	95
Sardegna	1630474	613	33
Sicilia	4968410	2129	275
Toscana	3737000	2539	387
Umbria	880285	544	18
Valle d'Aosta	125501	233	12
Veneto	4907704	4465	1000

Table S1. Population size, observed number of COVID19-related deaths from August 1st to December 31st and observed number of circulating infections on July 31st (i_0), by region.

Determinant	Measure	Year	Source
General health	Percentage of people with at least two chronic diseases	2019	ISTAT
Health infrastructures	Number of general practitioners (per 10'000 residents)	2018	ISTAT
	Number of pediatricians (number per 10'000 children)	2018	
	Number of hospitals (per 1'000 inhabitants)	2019	
	Percentage of public hospitals (over all hospitals)	2019	
Population age	Aging index [a]	2020	ISTAT
	Mean age in the population	2020	
Household size	Average number of persons per household	2020	ISTAT
Education	People with an academic degree (per 10'000 people 24-65)	2020	ISTAT
	Schooling rate [b]	2020	
Kindergarten	Children of 0-3 attending kindergartens (percentage of the 0-2)	2017	ISTAT
Public Transport	Percentage of people using public transport to go to work (%)	2020	ISTAT
	Percentage of people (0-34) using public transport to go to school (%)	2020	
Economic status	Poverty index [c]	2019	ISTAT
	Employment rate [d]	2018	
Tourism	Tourism rate [e]	2018	ISTAT
Energy consumption	Electric energy consumption of industries and manifactures (GWH)	2018	ISTAT
Density	Percentage of inhabitants of high urbanization areas	2011	ISTAT
Temperature	Average temperature (period October-December)	2020	www.Ilmeteo.it

Table S2. Description of the meta-regressors and corresponding source.
a. Number of over 65 per 100 individuals younger than 15.
b. Percentage of the individuals aged 20-24 who have at least a high school diploma.
c. People living in households below the poverty threshold (%).
d. Employed persons 15-64 (%, annual average).
e. Days of presence of tourists in a year in the region per inhabitant.

Region	School beginning
Abruzzo	2020-09-24
Basilicata	2020-09-24
Calabria	2020-09-24
Campania	2020-09-24
Emilia-Romagna	2020-09-14
Friuli Venezia Giulia	2020-09-16
Lazio	2020-09-14
Liguria	2020-09-14
Lombardia	2020-09-14
Marche	2020-09-14
Molise	2020-09-14
Trentino	2020-09-07
Piemonte	2020-09-14
Puglia	2020-09-24
Sardegna	2020-09-22
Sicilia	2020-09-14
Toscana	2020-09-14
Umbria	2020-09-14
Valle d'Aosta	2020-09-14
Veneto	2020-09-14

Table S3. Date of schools re-opening, by region (studenti.it).

Region	September 1st			October 1st			November 1st			December 1st			January 1st		
	prev.	90% CI		prev.	90% CI		prev.	90% CI		prev.	90% CI		prev.	90% CI	
Abruzzo	0.06	0.03	0.13	0.36	0.13	0.65	6.22	5.12	7.66	12.73	11.07	14.31	6.71	5.50	7.83
Basilicata	0.07	0.02	0.18	0.41	0.10	0.74	3.44	2.30	4.84	9.95	7.79	12.44	5.06	3.62	6.81
Calabria	0.01	0.01	0.04	0.05	0.02	0.12	1.42	0.98	1.89	5.24	4.38	6.20	2.44	1.87	2.99
Campania	0.07	0.05	0.11	0.44	0.29	0.54	3.36	3.00	3.78	10.40	9.69	11.15	5.70	5.22	6.24
Emilia-Romagna	0.10	0.06	0.15	0.26	0.15	0.38	4.67	4.13	5.29	16.55	15.38	17.65	16.46	15.46	17.55
Friuli Venezia Giulia	0.09	0.06	0.19	0.47	0.17	0.71	4.45	3.54	5.38	29.64	26.61	32.43	22.03	19.88	24.23
Lazio	0.13	0.09	0.19	0.62	0.46	0.77	4.89	4.40	5.45	10.06	9.38	10.74	8.31	7.69	8.89
Liguria	0.34	0.19	0.50	0.84	0.53	1.20	13.29	11.48	14.84	15.63	14.17	17.24	9.16	7.77	10.48
Lombardia	0.14	0.10	0.18	0.39	0.29	0.47	9.05	8.58	9.60	19.67	18.96	20.43	7.22	6.82	7.62
Marche	0.01	0.01	0.04	0.05	0.04	0.14	3.08	2.39	3.69	8.81	7.75	10.10	8.48	7.24	9.59
Molise	0.03	0.01	0.12	0.19	0.05	0.48	5.63	3.76	7.86	13.31	10.35	16.88	7.20	4.89	9.50
Piemonte	0.02	0.02	0.07	0.11	0.10	0.26	8.13	6.76	9.36	21.08	19.18	23.73	13.68	11.95	15.35
Puglia	0.09	0.05	0.15	0.16	0.13	0.27	8.07	7.34	8.82	21.68	20.62	22.93	8.18	7.45	8.90
Sardegna	0.20	0.11	0.30	0.43	0.30	0.61	3.65	3.16	4.12	11.34	10.35	12.30	6.98	6.34	7.67
Sicilia	0.16	0.05	0.27	0.65	0.34	0.98	3.82	3.06	4.69	7.52	6.42	8.65	6.16	5.18	7.18
Toscana	0.08	0.04	0.12	0.38	0.26	0.51	4.53	4.01	5.10	9.70	8.94	10.47	6.23	5.65	6.79
Trentino	0.10	0.05	0.15	0.19	0.13	0.31	7.91	7.06	8.80	15.43	14.37	16.47	6.44	5.74	7.14
Umbria	0.08	0.01	0.18	0.17	0.06	0.38	8.24	6.58	9.96	13.01	11.17	15.08	5.81	4.69	7.21
Valle d'Aosta	0.02	0.01	0.08	0.17	0.15	0.68	36.45	26.60	43.92	33.00	27.02	41.14	10.68	6.89	14.50
Veneto	0.45	0.33	0.57	0.56	0.42	0.74	5.15	4.66	5.69	18.66	17.74	19.69	23.32	22.17	24.40

Table S4. Prevalence of infections (number of infections over 1,000 inhabitants) with 90% confidence intervals on September 1st, October 1st, November 1st, December 1st, and January 1st, by region; $p = 1.15\%$, $T = 14$ days.

Spline coefficients						$R_0(t)$					Infections	
ϑ_0	ϑ_2	ϑ_3	ϑ_4	ϑ_5	ϑ_6	Max	Max (date)	Min	Min (date)	$= 1$ (date)	Peak	Date Peak
0.86	0.26	14.04	2.91	1.08	0.44	0.21	0.21	1.24	0.47	0.01	0.74	0.002

Table S5. Coefficient of variations of the model outputs as the inputs change.

Appendix

S1 Time discretization in the SIRD model

Compartmental models are described by a system of differential equations as in Equation 1 (main document). The underlying continuous time of the process can be discretized evaluating the size of the compartments at fixed points in time by considering constant intervals Δt. It follows that when $\Delta t = 1$ the system of differential equations is replaced by a system of difference equations:

$$\begin{cases} S(t) & = S(t-1) - \beta(t)\frac{S(t-1)}{S(0)}I(t-1) \\ I(t) & = I(t-1) + \beta(t)\frac{S(t-1)}{S(0)}I(t-1) - \alpha I(t-1) - \delta I(t-1) \\ R(t) & = R(t-1) + \alpha I(t-1) \\ D(t) & = D(t-1) + \delta I(t-1) \end{cases}. \tag{1}$$

The system of equations in (1) results from a two-fold approximation. First, we assume that the rates of transition are constant in each time interval. It follows that, denoting by λ the rate of occurrence of a given event, i.e. a transition from a compartment to another one, and by τ the time at which the event occurs, the size of each compartment depends on the probability $\Pr\big(\tau \in [t, t+\Delta t)\big)$. In particular, assuming that the waiting times are exponentially distributed, $\Pr\big(\tau \in [t, t+\Delta t)\big) = 1 - \exp(-\lambda \Delta t)$. When $\Delta t = 1$, this leads to the following system of difference equations:

$$\begin{cases} S(t) & = S(t-1) - \big(1 - \exp\big(\beta(t)\frac{I(t-1)}{S(0)}\big)\big)S(t-1) \\ I(t) & = I(t-1) + \big(1 - \exp\big(\beta(t)\frac{I(t-1)}{S(0)}\big)\big)S(t-1) - \big(1 - \exp\big(-\frac{1}{T}\big)\big)I(t-1) \\ R(t) & = R(t-1) + (1-p)\big(1 - \exp\big(-\frac{1}{T}\big)\big)I(t-1) \\ D(t) & = D(t-1) + p\big(1 - \exp\big(-\frac{1}{T}\big)\big)I(t-1) \end{cases}. \tag{2}$$

Note that a rate of transition $\lambda = \frac{1}{T}$ corresponds to the parameter of the exponential distribution modelling the minimum between time to death and time to recovery.

Finally, a second source of approximation is introduced considering $1 - \exp(-\lambda) \approx \lambda$. This approximation applies only for small values of the transition rates.

S2 Parametric bootstrap

We performed a parametric bootstrap [2, Ch 6.5] to compute confidence intervals around the point estimates of the coefficients of the cubic splines $\boldsymbol{\vartheta}$.

We relied on a parametric approach since it is a common practice in compartmental models to assume that the daily increments of deaths are distributed according to Negative Binomial distributions.

In particular, we assumed:

$$d(t) \sim NB(\lambda_d(t), \omega_d(t)) \tag{3}$$

where $d(t) = D(t) - D(t-1)$ and $NB(\lambda, \omega)$ is a Negative Binomial distribution with mean λ and clumping parameter ω.

Starting from the estimates of the model parameters $\widehat{\boldsymbol{\vartheta}}$ obtained via calibration, we computed the size of the corresponding SIRD compartments as $\widehat{S}, \widehat{I}, \widehat{R}, \widehat{D}$ and, from the estimated time series $\widehat{D}(t)$, the daily increments $\widehat{d}(t) = \widehat{D}(t) - \widehat{D}(t-1)$. Then, we set in (3):

$$\widehat{\lambda_d}(t) = \widehat{d}(t), \;\; \widehat{\omega_d}(t) = \widehat{d}(t-1). \tag{4}$$

It should be noticed that, being the variance of the Negative Binomial equal to $\lambda(1 + \lambda/\omega)$, the distribution (3) accounts for over-dispersion, especially at the beginning and at the end of the study period, when the number of deaths and new infections is small[3].

In order to get n bootstrap replications of the model parameters we adopted a simplified version of the bootstrap procedure detailed in Baccini et al.[4] and proposed by Chowell et al.[5]:

- For each t, we sampled increments $d^*(t)$ from the following Negative Binomial distribution:

$$d^*(t) \sim NB(\widehat{\lambda_d}(t), \widehat{\omega_d}(t)),$$

- We derived one bootstrap replication for the cumulative sum of these sampled increments obtaining one bootstrap sample of $D^*(t)$,
- We performed a calibration procedure assuming the bootstrap time series $D^*(t)$ as observed, thus obtaining a bootstrap estimate for $\boldsymbol{\vartheta}$.

This procedure was repeated n times to obtain n bootstrap replicates of the unknown parameters of the SIRD model and compute their percentile intervals [2, Ch 13]. We fixed $n = 500$.

S3 Global sensitivity analysis

Given K_X mutually independent inputs $(X_1, X_2, ..., X_{K_X})$ and a model which, given the inputs, returns K_Y outputs $(Y_1, Y_2, ..., Y_{K_Y})$, the Global Sensitivity Analysis (GSA) based on the Sobol's decomposition of the variance[6], allows computing both the first order and the total effect index for each input X_i and output Y (suppressing, for sake of simplicity, the index of the output variable).

The first order index for the output Y and the input X_i is defined as the ratio $S_i = \frac{\mathrm{Var}(\mathrm{E}(Y|X_i))}{\mathrm{Var}(Y)}$, corresponding to the fraction of the total variance attributable to the main effect of X_i.

The total effect index represents the proportion of the total variance of Y which is due to the main effect of the input X_i and all its interactions with the other inputs. Denoted as S_i^{tot}, it is defined as

$$S_i^{\mathrm{tot}} = \frac{\mathrm{E}(\mathrm{Var}(Y|\mathbf{X}_{\sim \mathbf{i}}))}{\mathrm{Var}(Y)} \tag{5}$$

$$\tag{6}$$

where $\mathbf{X}_{\sim \mathbf{i}}$ denotes the vector $(X_1, X_2, ...X_{i-1}, X_{i+1}, ...X_{K_X})$. According to the notation adopted in Saltelli[7], the outer operator E is applied to $\mathbf{X}_{\sim \mathbf{i}}$. In the case of a multivariate output ($K_Y > 1$), an overall index can be obtained for each input by averaging the single-output indexes weighted by the variance of their corresponding output[8,9].

S4 Bayesian meta-analysis

S4.1 Multivariate Bayesian meta-analysis

Let $\boldsymbol{\vartheta}_r$ be the vector of the spline coefficients for region r arising from the SIRD model, and Σ_r the associated estimate of the variance-covariance matrix resulting from the bootstrap replicates. We assumed the following Normal-Normal multivariate model:

$$\begin{aligned} \boldsymbol{\vartheta}_r &\sim N(\boldsymbol{\mu}_{\vartheta r}, \Sigma_r) \\ \boldsymbol{\mu}_{\vartheta r} &\sim N(\boldsymbol{\vartheta}_{all}, \Gamma), \quad r = 1, 2, ..., 20, \end{aligned}$$

where $\boldsymbol{\vartheta}_{all}$ and Γ are the overall meta-analytic vector of the spline coefficients and the between-region variance-covariance matrix, respectively. We used a MCMC algorithm to get a sample from the joint posterior distribution of $\boldsymbol{\vartheta}_{all}$ and Γ, through the function *mvma.bayesian* of the `altmeta` library of R software[10]. This function assigns vague non informative priors on $\boldsymbol{\vartheta}_{all}$ and Γ[10]. The same multivariate meta-analysis approach was also used to combine the regional estimates of the monthly average prevalence of infection from September to December (in this case, we combined vectors of 4 estimates).

S4.2 Bayesian univariate meta-analyses and meta-regressions

Let q_r and s_r^2 be the estimates of the quantity of interest and of its variance in the region r. The random effects meta-analysis model assumed:

$$\begin{aligned} q_r &\sim N(\mu_{q_r}, s_r^2) \\ \mu_{q_r} &\sim N(\zeta, \tau^2), \end{aligned}$$

where ζ is the overall meta-analytic quantity and τ^2 the between-region heterogeneity variance. Non-informative priors were assumed on the model hyperparameters: $\zeta \sim N(0, 10^3)$, $\frac{1}{\tau^2} \sim IG(10^2, 10^2)$. The percentage of variability due to the between region heterogeneity was expressed in terms of I^2 index ($I^2 = 100 \times \frac{\tau^2}{\tau^2 + \hat{\sigma}^2}$, where $\hat{\sigma}^2$ is the harmonic mean of the variances s_r^2). The previous model can be extended to include the meta-regressor x_r:

$$\begin{aligned} q_r &\sim N(\mu_{q_r}, s_r^2) \\ \mu_{q_r} &\sim N(\zeta_0 + \zeta_1 x_r, \tau_{res}^2), \end{aligned}$$

where x_r, ζ_0 is the overall meta-analytic estimate of the quantity of interest (in our analysis, the average $R_0(t)$ from October 1st to December 31st) when $x_r = 0$, ζ_1 is the meta-regression coefficient, and τ^2_{res} is the residual heterogeneity variance (the between-region heterogeneity which is not explained by x_r). Non-informative priors were assumed on the model hyperparameters: $\zeta_0 \sim N(0, 10^3)$, $\zeta_1 \sim N(0, 10^3)$), $\tau^2_{res} \sim IG(10^2, 10^2)$.

We used the software WinBUGS and the interface library R2WinBUGS of R software[11] to get a sample from the joint posterior distribution of the meta-analysis and meta-regression hyperparameters via MCMC algorithms (3 chains of 7'000 iterations, after a burning of 3000 iterations, and a thinning of 5).

References

1. Protezione Civile. Repository of COVID-19 outbreak data for Italy. https://github.com/pcm-dpc/COVID-19.

2. Efron, B. & Tibshirani, R. *An Introduction to the Bootstrap* (CRC press, Boca Raton, 1994).

3. Grenfell, B. T., Bjørnstad, O. N. & Finkenstädt, B. F. Dynamics of measles epidemics: Scaling noise, determinism, and predictability with the TSIR model. *Ecol. Monogr.* **72**, 185–202 (2002).

4. Baccini, M., Cereda, G. & Viscardi, C. The first wave of the SARS-CoV-2 epidemic in Tuscany (Italy): A SI2R2D compartmental model with uncertainty evaluation. *PLoS ONE* **16**, e0250029– (2021).

5. Chowell, G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infect. Dis. Model.* **2**, 379–398 (2017).

6. Sobol, I. M. Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.* **1**, 407–414. (1993).

7. Saltelli, A. *et al. Global Sensitivity Analysis: The Primer* (Wiley, 2008).

8. Lamboni, M., Monod, H. & Makowski, D. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliab. Eng. Syst. Saf.* **96**, 450–459 (2011).

9. Gamboa, F., Janon, A., Klein, T. & Lagnoux, A. Sensitivity analysis for multidimensional and functional outputs. *Electron. J. Stat.* **8**, 575–603 (2014).

10. Lin, L. & Chu, H. *altmeta: Alternative Meta-Analysis Methods*. https://CRAN.R-project.org/package=altmeta (2021).

11. Sturtz, S., Ligges, U. & Gelman, A. R2WinBUGS: A Package for Running WinBUGS from R. *J. Stat. Softw.* **12**, 1–16 (2005).