StatGroup19 empirical models for the Italian epidemic indicators

P. Alaimo Di Loro   M. Mingione
F. Divino   A. Farcomeni   G. Jona Lasinio   G. Lovison   A. Maruotti

ECDC Forecast Hub Seminar

October 27, 2021

Outline

European COVID-19 ForecastHub

Common modeling strategies

Compartmental models - **SIR** (Diekmann et al., 2012)

- **Features**
  - **Epidemiologically** correct models
  - **Mechanistically** describe the underlying **dynamic**
  - Can model the pandemic **all-round**

European COVID-19 ForecastHub

Common modeling strategies

Compartmental models - **SIR** (Diekmann et al., 2012)

- **Features**
  - **Epidemiologically** correct models
  - **Mechanistically** describe the underlying **dynamic**
  - Can model the pandemic **all-round**
- **Pitfalls** (Ioannidis et al., 2020)
  - Based on **chaotic systems**
  - Rely on **hypothetical constants** and **require accurate data**
  - **Bias** and **misspecification** lead to completely different results

Common modeling strategies

Compartmental models - **SIR** (Diekmann et al., 2012)

- **Features**
  - **Epidemiologically** correct models
  - **Mechanistically** describe the underlying **dynamic**
  - Can model the pandemic **all-round**
- **Pitfalls** (Ioannidis et al., 2020)
  - Based on **chaotic systems**
  - Rely on **hypothetical constants** and **require accurate data**
  - **Bias** and **misspecification** lead to completely different results

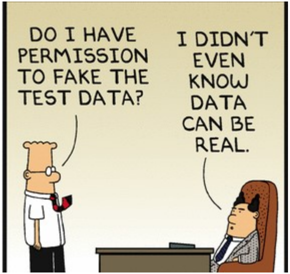
At the beggining of the Italian pandemic
most **forecasts** proved to be completely **wrong**!

European COVID-19 ForecastHub

Data issues

Errors and heterogeneity

- Different transmission and data collection system
- Measurement errors and errors in data entry
- Delays in reporting
- Temporal misalignment between indicators
- Tracking was highly symptoms driven



European COVID-19 ForecastHub

Phenomenological models (Chowell et al., 2016)

- **Features**
  - **Macroscopic** models
  - Draw inference from **global statistical properties** of the data
  - **Statistical artifacts** modeling the observables' mean

Common modeling strategies

Phenomenological models (Chowell et al., 2016)

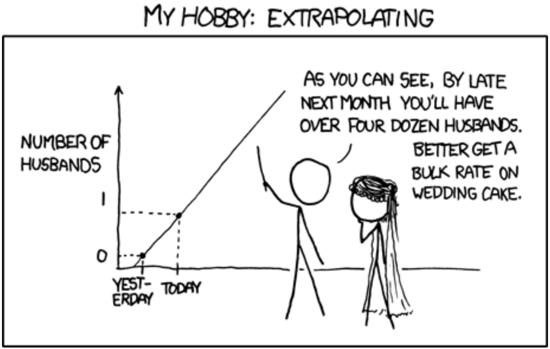
- **Features**
  - **Macroscopic** models
  - Draw inference from **global statistical properties** of the data
  - **Statistical artifacts** modeling the observables' mean
- **Pitfalls**
  - Entirely ignore the **nature of the data**
  - Assume **Gaussianity**
  - Use flexible (non-semiparametric) models to **extrapolate**

European COVID-19 ForecastHub

Common conceptual mistakes

Pick a model, validate its performances, but **be aware of its limitations**!

- Never ignore **confidence intervals**
- Good fit does not imply good **extrapolation**
- Predictions at **unreasonable time horizons** are **unreliable**



European COVID-19 ForecastHub

Italian available data

Italian available data

- **Data sources**: `GitHub` repository of *Italian Protezione Civile*

  https://github.com/pcm-dpc/COVID-19

- **Detail**: collected and updated daily by the regional Health Systems
- **Type** (main):
  - **Prevalence:** current positives, intensive care occupancy (stock)

$$Y_t = Y_{t-1} + I_t - O_t$$

  - **Incidence:** cumulative positives, cumulative deceased (flow)

$$Y_t^c = Y_{t-1}^c + \mathbf{Y_t}$$

European COVID-19 ForecastHub

Italian available data: incidence

- **Data sources**: `GitHub` repository of *Italian Protezione Civile*

  https://github.com/pcm-dpc/COVID-19

- **Detail**: collected and updated daily by the regional Health Systems
- **Type** (main):
  - **Prevalence:** current positives, intensive care occupancy (stock)

    $$Y_t = Y_{t-1} + I_t - O_t$$

  - **Incidence:** cumulative positives, cumulative deceased (flow)

    $$Y_t^c = Y_{t-1}^c + \mathbf{Y_t}$$

European COVID-19 ForecastHub

Logistic curves

- *S-shaped* curves
- Widely used to model various **growth** phenomena (biological, population, etc.)
- **Exponential** growth followed by a sudden **level off**

Cumulative incidence indicators: logistic growth

Logistic curves

- *S-shaped* curves
- Widely used to model various **growth** phenomena (biological, population, etc.)
- **Exponential** growth followed by a sudden **level off**

In the **epidemic** context

- Finite elements solutions/approximations to epidemiological **ode**
- Describe the **macroscopic** behavior of an **infection trajectory**
- Fit **globally** on the data respecting their **epidemic** nature

European COVID-19 ForecastHub

Richards' driven GLM

Cumulative counts follow a **Logistic Growth**

- Model the mean as a **modified Richards' curve**

$$\mathbb{E}[Y_t^c] = \lambda_\gamma(t) = b \cdot t + \frac{r}{\left[1 + 10^{h(p-t)}\right]^s}, \quad \gamma = [b, r, h, p, s]^\top$$

European COVID-19 ForecastHub

Richards' driven GLM

Cumulative counts follow a **Logistic Growth**

- Model the mean as a **modified Richards' curve**

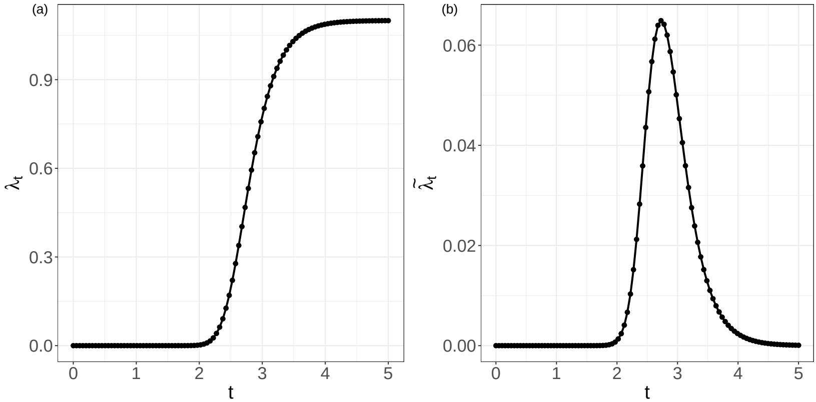
$$\mathbb{E}[Y_t^c] = \lambda_\gamma(t) = b \cdot t + \frac{r}{\left[1 + 10^{h(p-t)}\right]^s}, \quad \gamma = [b, r, h, p, s]^\top$$

- Innovations follow **first differences**

$$\mathbb{E}\left[\mathbf{Y_t}\right] = \mathbb{E}[Y_t^c] - \mathbb{E}[Y_{t-1}^c] = \lambda_\gamma(t) - \lambda_\gamma(t-1) \approx$$
$$\approx \frac{d}{dt}\lambda_\gamma(t) = \widetilde{\lambda}_\gamma(\mathbf{t}) = \mathbf{b} + \mathbf{f}_\lambda(\mathbf{t})$$

European COVID-19 ForecastHub

Richards' driven GLM



Example of a Richards' curve (a) and its first differences (b).

European COVID-19 ForecastHub

Modeling key-points of Alaimo Di Loro et al. (2021)

- Consider the effect of **covariates** through a link function

$$\eta_{\boldsymbol{\beta}}(\boldsymbol{X}) = \boldsymbol{\beta}\boldsymbol{X} \quad \Rightarrow \quad g_{\boldsymbol{\beta}}(\boldsymbol{X}) = \exp\left\{\eta_{\boldsymbol{\beta}}(\boldsymbol{X})\right\}$$

  - **Additive**:

$$\mu_{\boldsymbol{\theta}}(t, \mathbf{X}) = b_{\boldsymbol{\beta}}(\mathbf{X}) + r \cdot \widetilde{\lambda}_{\boldsymbol{\gamma}}(t), \quad b_{\boldsymbol{\beta}}(\mathbf{X}) = g_{\boldsymbol{\beta}}(\boldsymbol{X})$$

  - **Multiplicative**:

$$\mu_{\boldsymbol{\theta}}(t, \mathbf{X}) = b + r_{\boldsymbol{\beta}}(\mathbf{X}) \cdot \widetilde{\lambda}_{\boldsymbol{\gamma}}(t), \quad r_{\boldsymbol{\beta}}(\mathbf{X}) = g_{\boldsymbol{\beta}}(\boldsymbol{X})$$

European COVID-19 ForecastHub

Modeling key-points (Alaimo Di Loro et al., 2021)

- Behold to the **discrete** nature of counts
  Negative Binomial

$$Y_t|\boldsymbol{\theta}, \nu \sim NegBin(\mu_{\boldsymbol{\theta}}(t),\, \nu)$$

Modeling key-points (Alaimo Di Loro et al., 2021)

- Behold to the **discrete** nature of counts
  Negative Binomial

$$Y_t|\boldsymbol{\theta}, \nu \sim NegBin(\mu_{\boldsymbol{\theta}}(t),\ \nu)$$

- Given $\mu_{\boldsymbol{\theta}}(\cdot)$ the Y_t's is **stochastically independent** from Y^c_{t-1}:

$$Y_t \perp Y^c_\tau \quad \forall\, \tau < t$$

Cumulative counts **likelihood**:

$$f_{Y^c}(y^c_1, \ldots, y^c_T | y_0; \boldsymbol{\theta}) = \prod_{t=1}^{T} f_{Y^c_t}(y^c_t | y^c_{t-1}; \boldsymbol{\theta}) = \prod_{t=1}^{T} f_{Y_t}(y_t | \boldsymbol{\theta})$$

European COVID-19 ForecastHub

Estimation of Alaimo Di Loro et al. (2021)

Optimal parameters $\hat{\theta}$:

- **Log-Likelihood maximization**
- **Multi-start** routine based on **Fisher-scoring**
- Gradient $\nabla(\cdot)$ and Hessian $\mathcal{H}(\cdot)$ **computed analytically**

Estimation of Alaimo Di Loro et al. (2021)

Optimal parameters $\hat{\theta}$:

- **Log-Likelihood maximization**
- **Multi-start** routine based on **Fisher-scoring**
- Gradient $\nabla(\cdot)$ and Hessian $\mathcal{H}(\cdot)$ **computed analytically**

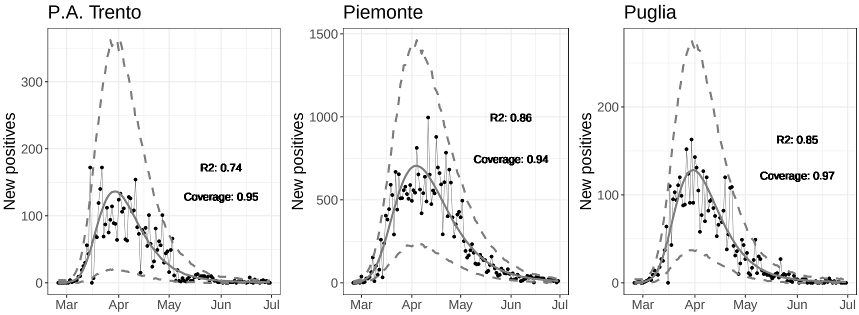
Uncertainty:

- Parameters' uncertainty quantified by the **Hessian**
- **Huber Sandwich** correction

$$\boldsymbol{V}\left[\hat{\boldsymbol{\theta}}\right] = \left(\mathcal{H}(\hat{\boldsymbol{\theta}})\right)^{-1} \nabla(\hat{\boldsymbol{\theta}})\nabla(\hat{\boldsymbol{\theta}})^{\top} \left(\mathcal{H}(\hat{\boldsymbol{\theta}})\right)^{-1}$$

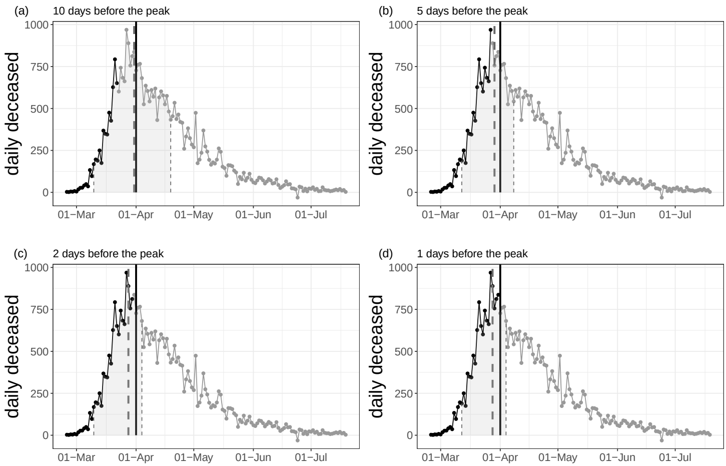
- Mean and prediction **intervals** obtained by bootstrap.

Model results - Fitting



Model fitting - Daily positives - Negative Binomial.

Model validation - peak detection



Estimation of the date of the peak for *daily deceased* at different steps-before.

European COVID-19 ForecastHub

Introducing dependence in Mingione et al. (2021)

Introducing dependence in Mingione et al. (2021)

Disease mapping

$$Y_{gt}|\mu_{gt} \sim Pois(\mu_{gt})$$
$$\log(\mu_{gt}) = \log(E_g) + \log(m_{gt}), \quad g = 1, \ldots, G, \quad t = 1, \ldots, T$$

- E_g is an offset accounting for area-specific exposures levels
- m_{gt} is a relative measure of the risk of area g at time t

Introducing dependence in Mingione et al. (2021)

Disease mapping

$$Y_{gt} | \mu_{gt} \sim Pois(\mu_{gt})$$
$$\log(\mu_{gt}) = \log(E_g) + \log(m_{gt}), \quad g = 1, \dots, G, \quad t = 1, \dots, T$$

- E_g is an offset accounting for area-specific exposures levels
- m_{gt} is a relative measure of the risk of area g at time t

The relative risk

$$\log(m_{gt}) = \log\left(\widetilde{\lambda}_{\gamma_g}(t)\right) + \boldsymbol{x}_{gt}^{\top}\boldsymbol{\beta} + \phi_{gt}$$

- $\widetilde{\lambda}_{\gamma_g}(t)$ logistic growth temporal trend
- $\boldsymbol{x}_{gt}^{\top}\boldsymbol{\beta}$ a linear predictor based on K covariates
- ϕ_{gt} is a random effect for the g-th area at time t

European COVID-19 ForecastHub

Introducing dependence in Mingione et al. (2021)

Spatial dependence (Stern and Cressie, 1999)

- Neighborhood graph $\boldsymbol{W}$ s.t. $w_{ii} = 0$, $w_{ij} > 0$ iff $i \sim j$
- **CAR proper prior**

$$\phi_t \sim \mathcal{N}_G \left(\mathbf{0}, \, \sigma^2 \cdot (\boldsymbol{D} - \alpha \boldsymbol{W})^{-1} \right)$$

Introducing dependence in Mingione et al. (2021)

Spatial dependence (Stern and Cressie, 1999)

- Neighborhood graph $\boldsymbol{W}$ s.t. $w_{ii} = 0$, $w_{ij} > 0$ iff $i \sim j$
- **CAR proper prior**

$$\phi_t \sim \mathcal{N}_G\left(\mathbf{0},\, \sigma^2 \cdot (\boldsymbol{D} - \alpha \boldsymbol{W})^{-1}\right)$$

Temporal dependence (Rushworth et al., 2014)

- 1-st order dependence on the past
- **AR(1)** over the space-vectors $\{\phi_t\}_{t=1}^T$

$$\phi_t \,|\, \phi_{1:t-1} \sim \mathcal{N}_G\left(\rho \cdot \phi_{t-1},\ \sigma^2 \cdot (\boldsymbol{D} - \alpha \boldsymbol{W})^{-1}\right)$$

European COVID-19 ForecastHub

Alternative settings of Mingione et al. (2021)

Common factors

- $\log(E_g) = \log(residents_g)$ scaled by a factor of 10^5
- Number of total weekly swabs (standardised) as covariate

Alternative settings of Mingione et al. (2021)

Common factors

- $\log(E_g) = \log(\textit{residents}_g)$ scaled by a factor of 10^5
- Number of total weekly swabs (standardised) as covariate

"Global VS Individual" growth

- One **common Richards** for all regions $\lambda_\gamma(\cdot)$
- **Individual Richards** for each region $\lambda_{\gamma_g}(\cdot)$

European COVID-19 ForecastHub

Alternative settings of Mingione et al. (2021)

Common factors

- $\log(E_g) = \log(residents_g)$ scaled by a factor of 10^5
- Number of total weekly swabs (standardised) as covariate

"Global VS Individual" growth

- One **common Richards** for all regions $\lambda_\gamma(\cdot)$
- **Individual Richards** for each region $\lambda_{\gamma_g}(\cdot)$

Alternative spatial dependence graphs $\boldsymbol{W}$

- $\boldsymbol{W}_{Ind} = \mathbf{0} \rightarrow$ spatial independence
- $\boldsymbol{W}_{Flow}$ based on proximity flows (direct HV trains, flights, ferries) as in Della Rossa et al. (2020)
- $\boldsymbol{W}_{Geo}$ based on regions' geographical position

European COVID-19 ForecastHub

Estimation in Mingione et al. (2021): Bayesian

Prior setting

$$\log(b), \log(r) \sim \mathcal{N}(0, 100) \quad \log(h), \log(s) \sim \mathcal{N}(0, 1)$$
$$p \sim \mathcal{N}\left(T/2,\ T/(2 \cdot 1.96)\right) \quad \boldsymbol{\beta} \sim \mathcal{N}_K\left(\mathbf{0}, 100 \cdot \boldsymbol{I}_K\right)$$

European COVID-19 ForecastHub

Estimation in Mingione et al. (2021): Bayesian

Prior setting

$$\log(b), \log(r) \sim \mathcal{N}(0, 100) \quad \log(h), \log(s) \sim \mathcal{N}(0, 1)$$
$$p \sim \mathcal{N}\left(T/2,\ T/(2 \cdot 1.96)\right) \quad \boldsymbol{\beta} \sim \mathcal{N}_K\left(\mathbf{0}, 100 \cdot \boldsymbol{I}_K\right)$$

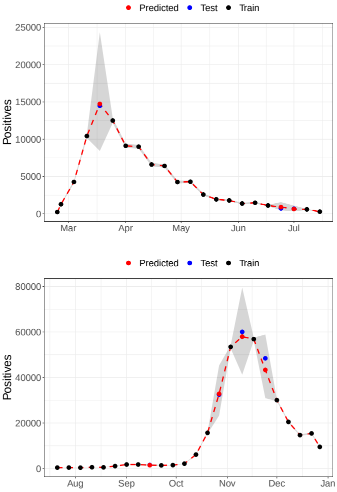
Implementation using `STAN` (Carpenter et al., 2017)

- Hamiltonian Monte Carlo → **NUTS** (Hoffman and Gelman, 2014)
- **Exact-sparse** CAR (Joseph, 2016) for computational efficiency
- 70%/30% **in-sample/out-of-sample** split

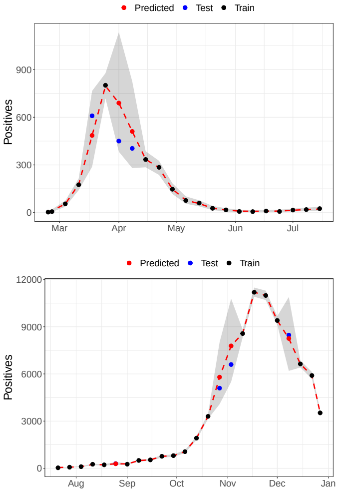
Codes at https://github.com/minmar94/Covid19-Spatial

European COVID-19 ForecastHub

Out-of-sample predictions



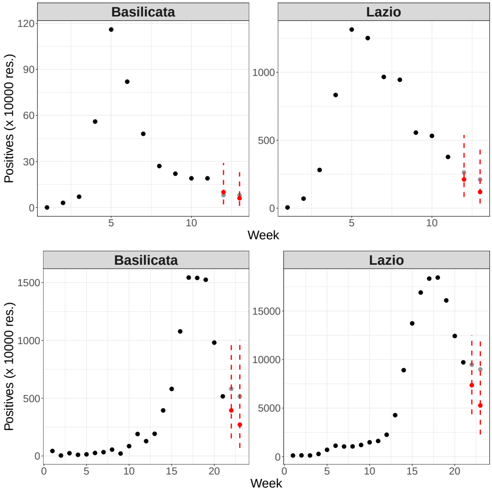
(a) Lombardia



(b) Sicilia

European COVID-19 ForecastHub

Forecasting ability of the best model



Black dots in-sample, grey dots out-of-sample, red dots predicted values, red dashed 95% prediction intervals.

European COVID-19 ForecastHub

Italian available data: prevalence

Italian available data: prevalence

- **Data sources**: `GitHub` repository of *Italian Protezione Civile*

  https://github.com/pcm-dpc/COVID-19

- **Detail**: collected and updated daily by the regional Health Systems
- **Type** (main):
  - **Prevalence:** current positives, intensive care occupancy (stock)

    $$\mathbf{Y_t} = Y_{t-1} + I_t - O_t$$

  - **Incidence:** cumulative positives, cumulative deceased (flow)

    $$Y_t^c = Y_{t-1}^c + Y_t$$

European COVID-19 ForecastHub

Nowcasting of ICUs in Farcomeni et al. (2021)

Focus on **ICU occupancy**

- $Y_{gt} \in \{0, \ldots, B_g\}$ **occupied beds** at time t in area g, with **capacity** B_g
- **Key** for **planning and allocating** health resources

Provide **1 to 5 days-ahead** predictions

- Use only data from the most **recent two weeks**
- Optimal **ensemble** of two methods
- Validate **short-term** performances on the run

European COVID-19 ForecastHub

Nowcasting of ICUs in Farcomeni et al. (2021)

First model

- **Generalized Linear Mixed Effect Model (GLMM)**
- Fit through `glmer` in the `lme4` package (Bates et al., 2007)

Second model

GLMM

$$Y_{gt} \sim Poisson(\lambda_{gt})$$

where

$$\log(\lambda_{gt}/R_g) = \beta_{g0} + \beta_{g1} t + \beta_{g2} t^2$$

and

$$(\beta_{g0}, \beta_{g1}, \beta_{g2}) \sim MVN((\beta_0, \beta_1, \beta_2), \Sigma)$$

European COVID-19 ForecastHub

Nowcasting of ICUs in Farcomeni et al. (2021)

First model

- **Generalized Linear Mixed Effect Model (GLMM)**
- Fit through `glmer` in the `lme4` package (Bates et al., 2007)

Second model

- **INGARCH** (Agosto et al., 2016; Chen and Lee, 2016)
- Fit through `tsglm` in the `tscount` package (Liboschik et al., 2015)

INGARCH

$$Y_{gt} \sim Pois\,(\mu_{gt})$$
$$\log\,(\mu_{gt}) = \alpha_{0g} \cdot \log\,(\mu_{g,t-1}) + \alpha_{1g} \cdot \log\,(Y_{g,t-1} + 1) + \eta_{\boldsymbol{\beta}}(t)$$

where

$$\eta_{\boldsymbol{\beta}}(t) = \beta_{0g} + \beta_{1g}t + \beta_{2g}t^2 + \beta_{3g}t^3$$

Nowcasting of ICUs in Farcomeni et al. (2021)

First model

- **INGARCH** (Agosto et al., 2016; Chen and Lee, 2016)
- Fit through `tsglm` in the `tscount` package (Liboschik et al., 2015)

Second model

- **Generalized Linear Mixed Effect Model**
- Fit through `glmer` in the `lme4` package (Bates et al., 2007)

Model averaging

- **Convex combination** of predictions
- **Leave-last-out** optimal weight

European COVID-19 ForecastHub

Model averaging

Ensemble predictor

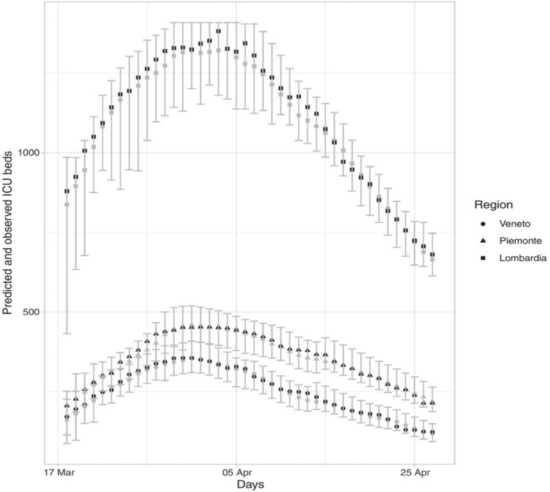
$$\hat{Y}_{g,t+1} = w_{gt}\,\hat{Y}_{g,t+1}^{(INGARCH)} + (1 - w_{gt})\,\hat{Y}_{g,t+1}^{(GLMM)}$$

where

- $w_{gt} \in (0, 1)$ and $w_{gt} = 0.5$ for $t < 15$
- **Prediction intervals** are obtained as the **weighted average** of the limits of prediction intervals
- **Jensen's inequality** show that this conservatively guarantees the **nominal level**

European COVID-19 ForecastHub

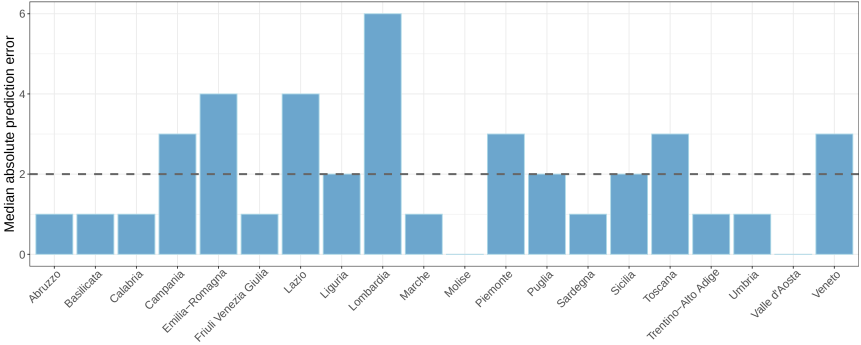
Results (Farcomeni et al., 2021)



Predicted (grey) vs observed (black) number of ICU beds during the first outbreak in Lombardia, Vento and Piemonte. Grey solid lines are 99% confidence intervals.

European COVID-19 ForecastHub

Results



Median absolute prediction error by region for ICU occupancy since March 12, 2020.

European COVID-19 ForecastHub

Concluding remarks

Concluding remarks

Conclusion

- ✓ Development of a coherent framework for the growth dynamic of counts
- ✓ Inclusion of desirable space-time dependence in the residuals
- ✓ Development of a reliable ensemble predictor of ICU occupancy

European COVID-19 ForecastHub

Concluding remarks

Conclusion

- ✓ Development of a coherent framework for the growth dynamic of counts
- ✓ Inclusion of desirable space-time dependence in the residuals
- ✓ Development of a reliable ensemble predictor of ICU occupancy

Work in progress

- × Growth model for prevalence indicators using INAR perspective
- × Include space-time dependence in the ICU ensemble model
- × Include the effects of external policies in the model

Main references

A. Agosto, G. Cavaliere, D. Kristensen, and A. Rahbek. Modeling corporate defaults: Poisson autoregressions with exogenous covariates (parx). *Journal of Empirical Finance*, 38:640–663, 2016.

P. Alaimo Di Loro, F. Divino, A. Farcomeni, G. Jona Lasinio, G. Lovison, A. Maruotti, and M. Mingione. Nowcasting covid-19 incidence indicators during the italian first outbreak. *Statistics in Medicine*, 40(16):3843–3864, 2021.

D. Bates, D. Sarkar, M. D. Bates, and L. Matrix. The lme4 package. *R package version*, 2(1):74, 2007.

B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: a probabilistic programming language. *Grantee Submission*, 76(1):1–32, 2017.

C. W. Chen and S. Lee. Generalized poisson autoregressive models for time series of counts. *Computational Statistics & Data Analysis*, 99:51–67, 2016.

G. Chowell, D. Hincapie-Palacio, J. Ospina, B. Pell, A. Tariq, S. Dahal, S. Moghadas, A. Smirnova, L. Simonsen, and C. Viboud. Using phenomenological models to characterize transmissibility and forecast patterns and final burden of zika epidemics. *PLoS currents*, 8, 2016.

F. Della Rossa, D. Salzano, A. Di Meglio, F. De Lellis, M. Coraggio, C. Calabrese, A. Guarino, R. Cardona-Rivera, P. De Lellis, D. Liuzza, et al. A network model of Italy shows that intermittent regional strategies can alleviate the COVID-19 epidemic. *Nature communications*, 11(1):1–9, 2020.

O. Diekmann, H. Heesterbeek, and T. Britton. *Mathematical tools for understanding infectious disease dynamics*, volume 7. Princeton University Press, 2012.

A. Farcomeni, A. Maruotti, F. Divino, G. Jona-Lasinio, and G. Lovison. An ensemble approach to short-term forecast of COVID-19 intensive care occupancy in italian regions. *Biometrical Journal*, 63:503–513, 2021.

M. D. Hoffman and A. Gelman. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.

J. P. Ioannidis, S. Cripps, and M. A. Tanner. Forecasting for COVID-19 has failed. *International Journal of Forecasting*, 2020. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2020.08.004. URL http://www.sciencedirect.com/science/article/pii/S0169207020301199.

M. Joseph. Exact sparse CAR models in Stan, 2016. *URL http://mc-stan. org/users/documentation/case-studies/mbjoseph-CARStan. html*, 2016.

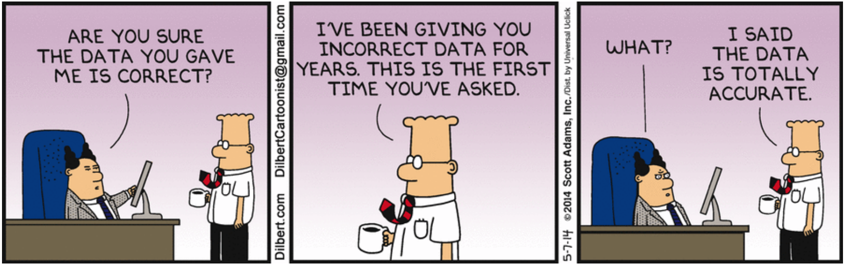
T. Liboschik, K. Fokianos, and R. Fried. *tscount: An R package for analysis of count time series following generalized linear models*. Universitätsbibliothek Dortmund Dortmund, Germany, 2015.

M. Mingione, P. Alaimo Di Loro, A. Farcomeni, F. Divino, G. Lovison, G. J. Lasinio, and A. Maruotti. Spatial modelling of covid-19 incident cases using richards' curve: an application to the italian regions. *arXiv preprint arXiv:2106.05067*, 2021.

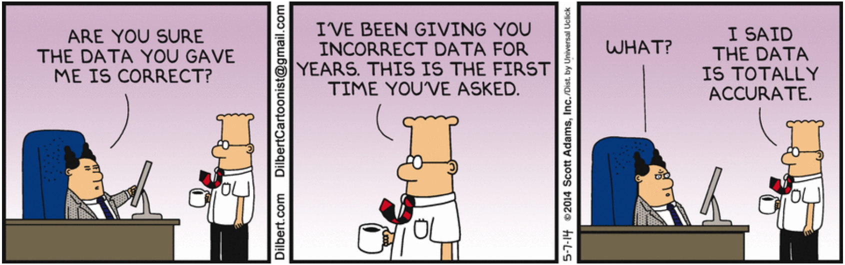
A. Rushworth, D. Lee, and R. Mitchell. A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London. *Spatial and spatio-temporal epidemiology*, 10:29–38, 2014.

H. Stern and N. A. Cressie. Inference for extremes in disease mapping. 1999.

THANK YOU FOR YOUR ATTENTION!



THANK YOU FOR YOUR ATTENTION!



Questions?

European COVID-19 ForecastHub

APPENDIX

Deviance residuals (Svetliza and Paula, 2003)

Poisson

$$\hat{d}_t^{\text{Poi}} = \text{sgn}\left(y_t - \mu_{\hat{\boldsymbol{\theta}}}(t)\right) \cdot \sqrt{2y_t \log\left(\frac{y_t}{\mu_{\hat{\boldsymbol{\theta}}}}\right) - \left(y_t - \mu_{\hat{\boldsymbol{\theta}}}(t)\right)}$$

Negative Binomial

$$\hat{d}_t^{\text{NB}} = \text{sgn}\left(y_t - \mu_{\hat{\boldsymbol{\theta}}}(t)\right) \cdot \sqrt{2\left[y_t \log\left(\frac{y_t}{\mu_{\hat{\boldsymbol{\theta}}}(t)}\right) - (y_t + \nu) \cdot \log\left(\frac{y_t + \nu}{\mu_{\hat{\boldsymbol{\theta}}}(t) + \nu}\right)\right]}$$

Pearson residuals

$$\hat{\phi}_t = \frac{y_t - \hat{y}_t}{\widehat{\mathrm{Var}}\left[Y_t\right]}, \qquad t = 1, \ldots, T.$$

where:

$$\widehat{\mathrm{Var}}_{\mathrm{Poi}}\left[Y_t\right] = \mu_{\hat{\boldsymbol{\theta}}}(t), \qquad \widehat{\mathrm{Var}}_{\mathrm{NB}}\left[Y_t\right] = \mu_{\hat{\boldsymbol{\theta}}}(t) + \frac{\mu_{\hat{\boldsymbol{\theta}}}(t)^2}{\hat{\nu}}$$

European COVID-19 ForecastHub

Model validation

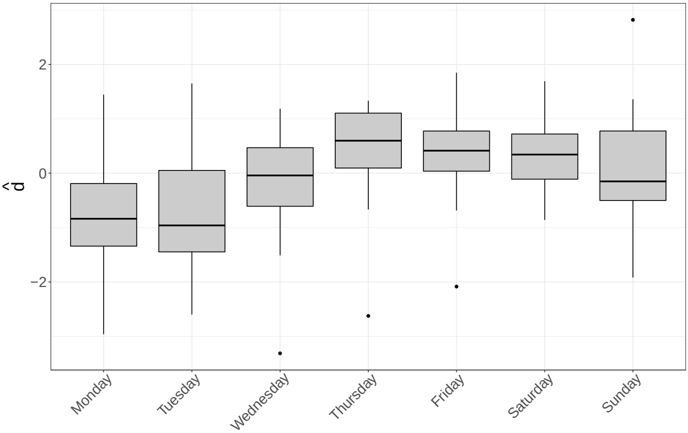


Figure: Deviance residuals distribution aggregated by day of the week for *daily positives*.

European COVID-19 ForecastHub

Model validation

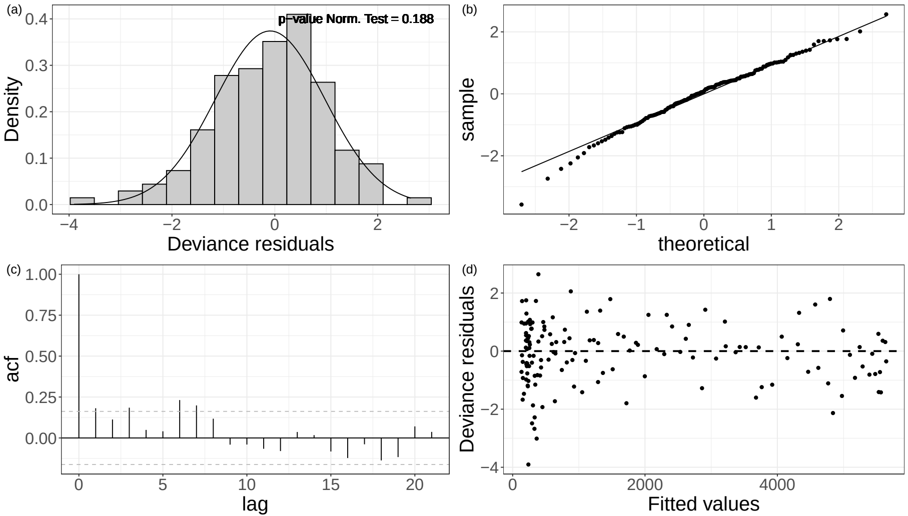


Figure: Residual diagnostic - Daily positives - Negative Binomial.

European COVID-19 ForecastHub

Model results

Table: *Parameters' point estimates and 95% confidence intervals for the additive model on daily positives.*

Parameter	Point estimate	95% Interval
β_0	5.26	$(5.18, 5.34)$
β_{wd}	-0.46	$(-0.53, -0.38)$
r	224.57×10^3	$(224.13 \times 10^3, 225.01 \times 10^3)$
h	0.0289	$(0.0287, 0.0291)$
p	-23.26	$(-29.64, -16.88)$
s	44.42	$(-35.67, 124.51)$
ν	22.01	$(21.35, 22.70)$

European COVID-19 ForecastHub

Parameters' estimates

Wave	Param.	M_{Ind}	M_{Flow}	M_{Geo}
I	α	–	**0.14 (0.02, 0.21)**	**0.76 (0.71, 0.81)**
	ρ	0.89 (0.87, 0.91)	**0.88 (0.90, 0.93)**	**0.86 (0.85, 0.89)**
	β	0.36 (0.26, 0.44)	0.34 (0.25, 0.42)	0.21 (0.14, 0.29)
II	α	–	**0.93 (0.92, 0.95)**	**0.87 (0.85, 0.90)**
	ρ	0.88 (0.86, 0.90)	**0.87 (0.85, 0.89)**	**0.82 (0.80, 0.85)**
	β	0.42 (0.38, 0.46)	0.27 (0.24, 0.30)	0.13 (0.09, 0.16)

Comparison of parameters' estimates for the spatial (α) and temporal (ρ) auto-correlation, and for the swabs' effect in the first and the second wave.

European COVID-19 ForecastHub

Model selection and validation

Wave	Metric	M_{Ind}	M_{Flow}	M_{Geo}
I	Coverage	0.98	0.98	0.98
	PIW	1535	1178	1144
	RMSE	423	**184**	**272**
	WAIC	2869	**2650**	**2774**
	LOO	3087	**2849**	**2982**
II	Coverage	0.96	0.97	0.92
	PIW	33393	4497	4046
	RMSE	12841	**910**	**995**
	WAIC	4112	**3820**	**3971**
	LOO	4393	**4080**	**4252**

Out-of-sample predictive performances for the first and the second wave.

European COVID-19 ForecastHub

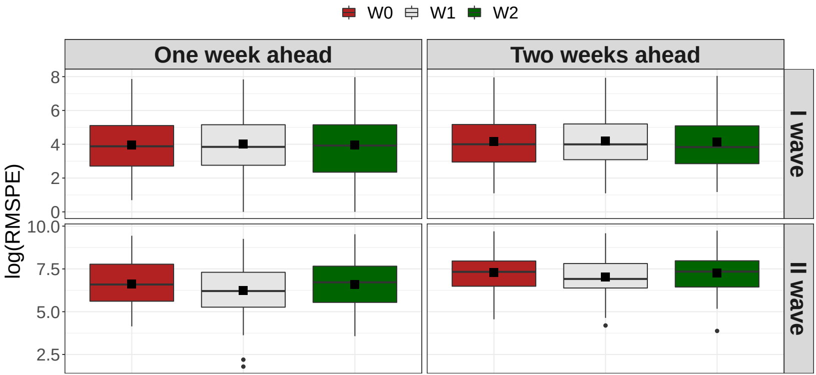
Parameters' estimates

Wave	Model	h	s
I	M_0	0.62 (0.60, 0.64)	7.8 (6.3, 9.9)
	M_1	0.62 (0.59, 0.65)	7.9 (5.5, 9.3)
	M_2	0.61 (0.58, 0.65)	7.8 (5.2, 9.3)
II	M_0	3.46 (3.26, 3.63)	0.06 (0.05,0.07)
	M_1	2.72 (2.33, 3.08)	0.09 (0.07,0.10)
	M_2	3.50 (3.20, 3.70)	0.06 (0.05, 0.07)

Wave	Model	b	r	p
I	M_0	0.05 (0.04, 0.06)	23 (20, 27)	2.0 (1.5, 2.5)
	M_1	0.06 (0.05, 0.07)	20 (17, 22)	2.2 (1.7, 2.8)
	M_2	0.05 (0.04, 0.06)	26 (21, 31)	2.2 (1.5, 2.9)
II	M_0	$7 \cdot 10^{-5}$ ($1 \cdot 10^{-6}$, $1 \cdot 10^{-3}$)	158 (143, 172)	23.2 (23.1, 23.3)
	M_1	$2 \cdot 10^{-4}$ ($3 \cdot 10^{-5}$, $7 \cdot 10^{-3}$)	178 (127, 215)	22.9 (22.8, 23.2)
	M_2	$4 \cdot 10^{-4}$ ($3 \cdot 10^{-6}$, $1 \cdot 10^{-2}$)	194 (163, 220)	23.1 (22.9, 23.2)

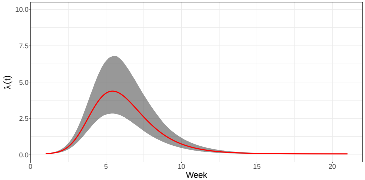
Table: Parameters' estimates of the Richards' curve for the waves I and II.

Forecasting comparisons between graphs

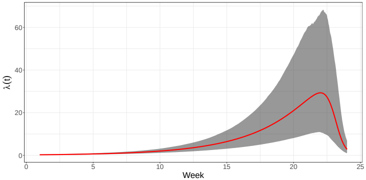


Prediction error (on the log scale) at different steps ahead, for each specification of ***W*** and for each wave.

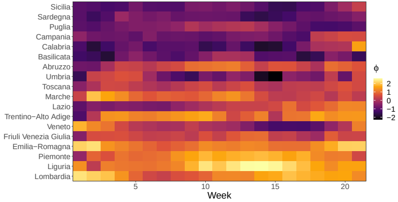
Common Richards' and latent effects



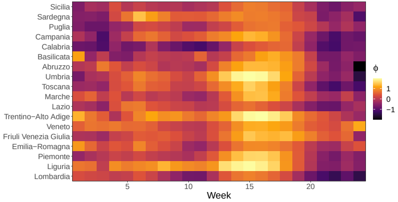
(a) M_{Flow} - I wave - $\widetilde{\lambda}_{\gamma}(t)$



(b) M_{Flow} - II wave - $\widetilde{\lambda}_{\gamma}(t)$



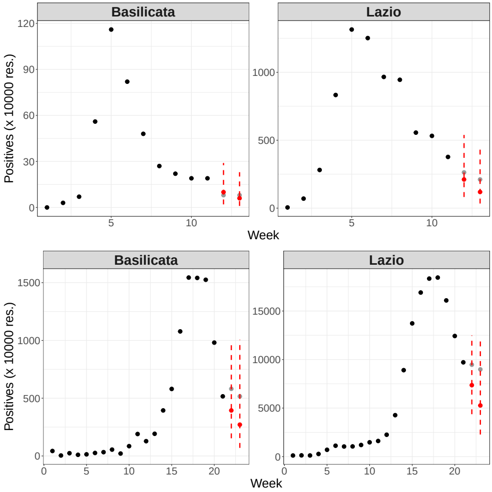
(c) M_{Flow} - I wave - ϕ_{gt}



(d) M_{Flow} - II wave - ϕ_{gt}

European COVID-19 ForecastHub

Forecasting ability of the best model



Black dots in-sample, grey dots out-of-sample, red dots predicted values, red dashed 95% prediction intervals.