Global Short-Term Forecasting of Covid-19 Cases

Thiago P. Oliveira[1] **and Rafael A. Moral**[2]

[1]NUI Galway and Insigth Centre for Data Analytics; [2]Maynooth University

Email: thiago.paulaoliveira@insigth-centre.org

Introduction

- The continuously growing number of COVID-19 cases pressures healthcare services worldwide [1].
- This pandemic is associated with:
  - → high basic reproduction numbers
  - → spreading with great speed
  - → a significant number of infected individuals remain asymptomatic
- Accurate short-term forecasting is thus vital to support country-level policy making [2].
  - → political leadership
  - → socioeconomic reality
  - → epidemic stage
  - → especially for systems on the brink of collapse
- Accounting for the hierarchical structure of the data and accommodating extra-variability is fundamental.

Motivation

- The main problem is that not only is this disease new, but there are also many factors acting in concert resulting in a seemingly unpredictable outbreak progression.
- Forecasting with great accuracy under these circumstances is very difficult.
- We propose a new modelling framework, based on a state-space hierarchical model:
  - → generate forecasts with very good accuracy for up to seven days ahead
  - → introduce a autoregressive parameter as a function of time, increasing predictive power and flexibility to adapt to each country
- Provide all results as an R Shiny Dashboard, including week-long forecasts for every country in the world

Autoregressive hierarchical state-space NB model

- We introduce a class of state-space hierarchical models for overdispersed count time series

$$
\begin{aligned}
Y_{it}|Y_{i,t-1} &\sim \mathrm{NB}(\mu_{it}, \psi) \\
\log \mu_{it} &= \gamma_{it} + \Omega_{it} \\
\gamma_{it} &= \phi_{it}\gamma_{it-1} + \eta_{it}, \text{ with } \eta_{it} \sim \mathrm{N}\left(0, \sigma^2_\eta\right) \\
\phi_{it} &= \sum_{q=0}^{Q} (\beta_q + b_{iq}) P_q(t), \text{ with } \boldsymbol{b}_i \sim \mathrm{N}_Q\left(\mathbf{0}, \Sigma_b\right) \\
\Omega_{it} &= \lambda_{it}\omega_{it}
\end{aligned}
$$

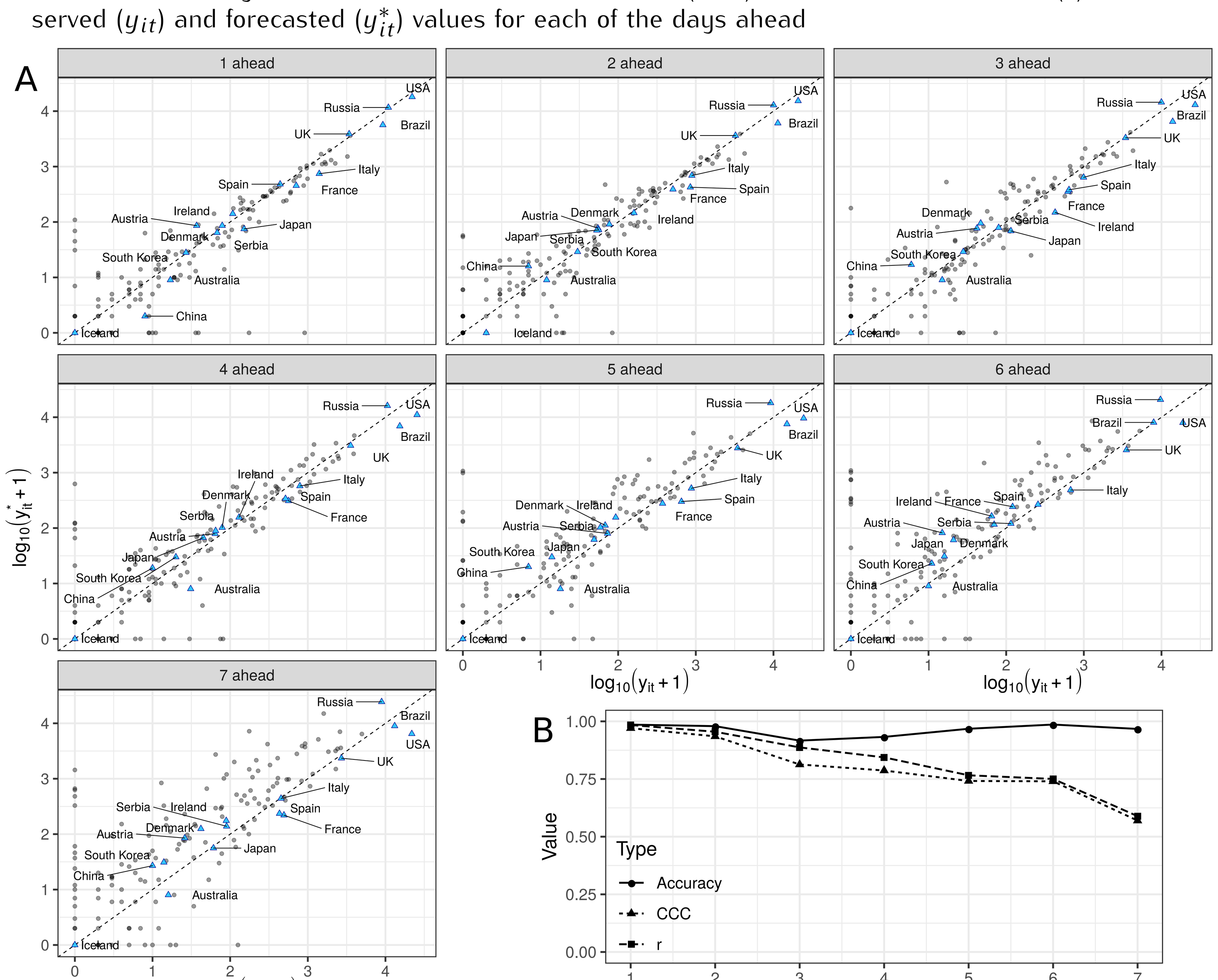
where $\lambda_{it} \sim \text{Bernoulli}(\pi)$ and $\omega_{it} \sim \mathrm{N}(0, \sigma^2_\omega)$. When $\lambda_{it} = 1$, then observation y_{it} is considered to be an outlier, and the extra variability is modelled by σ^2_ω The model is estimated using a Bayesian framework, and the prior distributions used are

$$
\beta_i \sim \mathrm{N}_Q(\mathbf{0}, \mathbf{I}_Q \times 1000), \quad \sigma^{-2}_{b_q}, \sigma^{-2}_\eta, \sigma^{-2}_\omega \sim \text{Gamma}(0.001, 0.001), \quad \pi \sim \text{Uniform}(0, 1)
$$

- 3 MCMC chains
- 2,000 adaptation iterations
- 50,000 as burn-in
- 50,000 iterations per chain with a thinning of 25

Model Validation

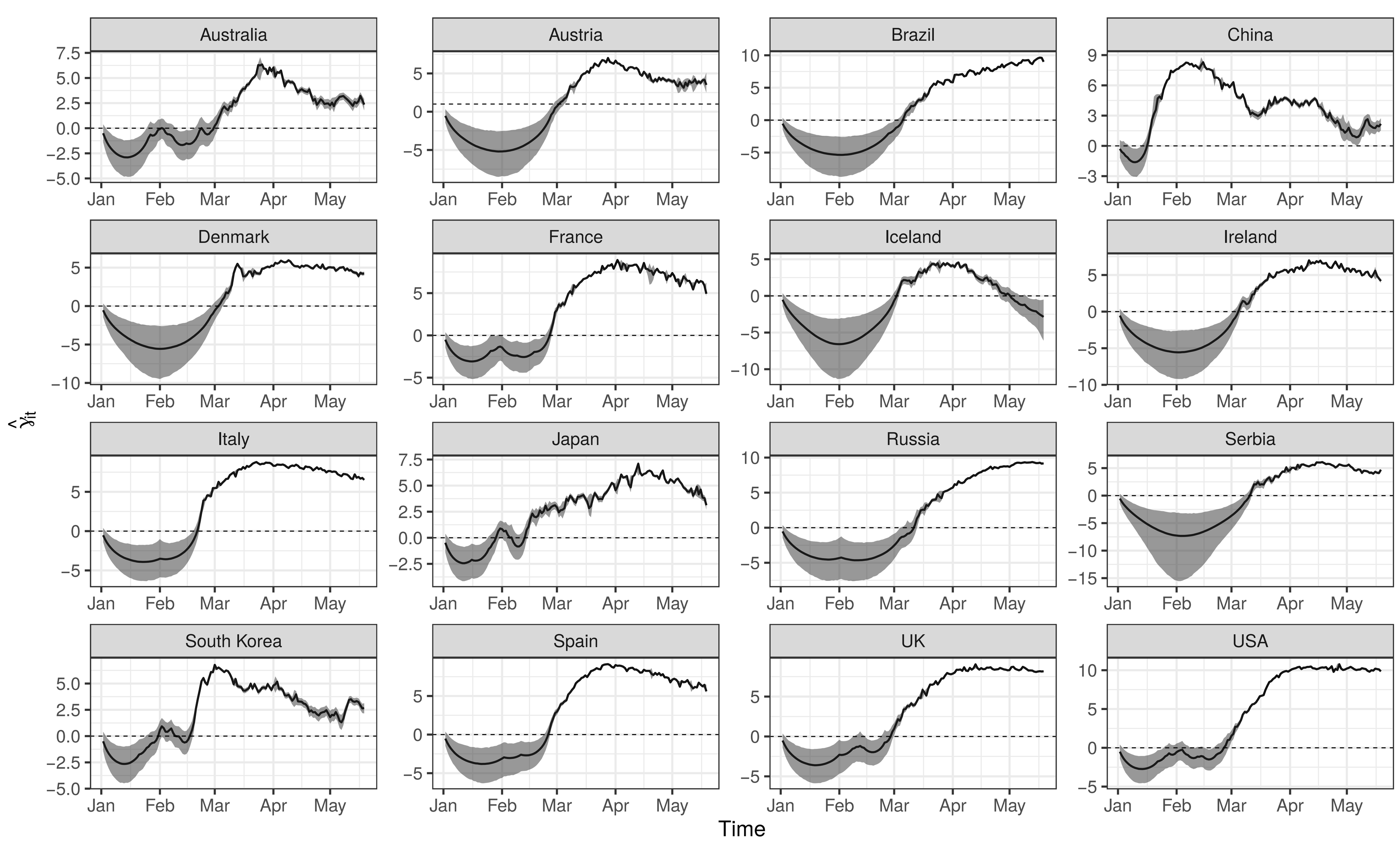
- Logarithm of the observed y_{it} versus the forecasted daily number of cases y^*_{it} for each country, for up to seven days ahead, where each day ahead constitutes one panel
- Observed accuracy, concordance correlation coefficient (CCC) and Pearson correlation (r) between observed (y_{it}) and forecasted (y^*_{it}) values for each of the days ahead



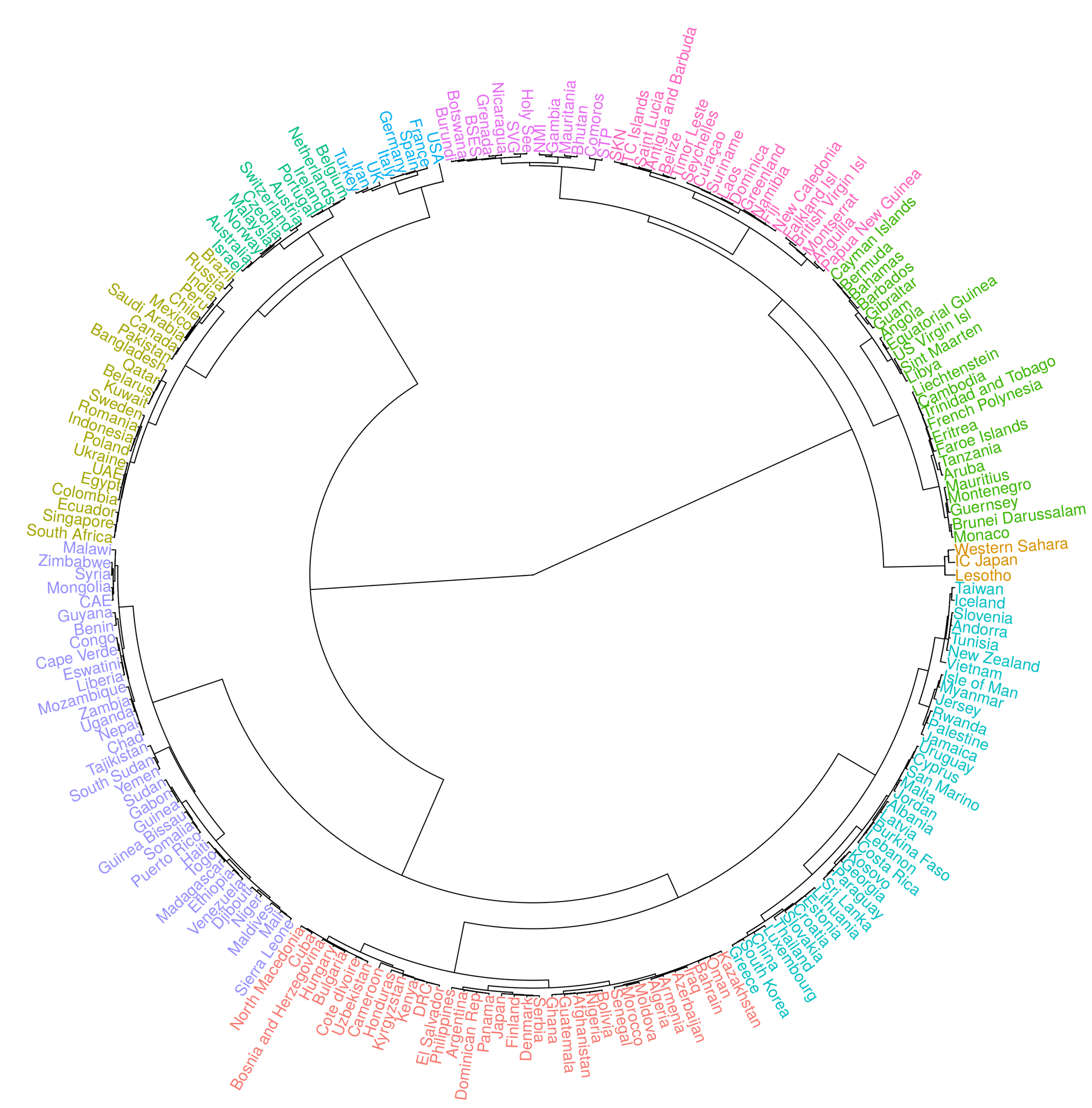
Get more results through our **Dashboard App**: https://prof-thiagooliveira.shinyapps.io/COVIDForecast

Results

- Posterior means of the autoregressive component γ_{it} (solid lines) and associated 95% credible intervals (shaded areas) from 20-Mar-2020 to 19-May-2020.
  - → Direct relationship with the pandemic behaviour over time
  - → Directly proportional to the natural logarithm of the daily number of cases, given what happened in the previous day
  - → Sensitive to changes and can be helpful detecting a possible second wave



- Dendrogram representing the hierarchical clustering of countries based on their estimated autoregressive parameters $\hat{\gamma}_{it}$ from 20-Mar-2020 to 19-May-2020.
  - → Each of 10 clusters is represented with a different colour.
  - → We can see countries have had the most similar recent behaviour of the outbreak.
  - → Study similar or different measures taken by these other countries that may help determine policy.



Final Remarks

- We must be very careful when looking at the forecasted number of cases because these values must not be looked at in isolation.
- It is imperative that the entire context is looked at and that we understand how the data is actually generated.

References

[1] **Kassem, A. M.** (2020). COVID-19: Mitigation or suppression? *Arab. J. Gastroenterol*, **21**, 1 – 2.

[2] **Goodell, J. W.** (2020). COVID-19 and finance: Agendas for future research. *Finance Res. Lett*, DOI: 10.1016/j.frl.2020.101512.

Acknowledgement

We would like to thank Prof. John Hinde for helpful comments when preparing this manuscript. We extend our thanks for the Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289.