A HIGHER-ORDER TAYLOR EXPANSION OF THE INITIAL TRAJECTORY OF COVID-19 CASES AND DEATHS VIA BAYESIAN HIERARCHICAL MODELS: A TOY PROBLEM AND POSSIBLE PUBLIC HEALTH INSIGHTS.

Alessandro Maria Selvitella *
Department of Mathematical Sciences
Purdue University Fort Wayne
Fort Wayne, IN - USA
aselvite@pfw.edu

Kathleen Lois Foster
Department of Biology
Ball State University,
Muncie, IN - USA
klfoster@bsu.edu

ABSTRACT

We study the initial evolution of COVID-19 cases and deaths with machine learning methods. Our interest is to understand if a nonlinear time component is present at the beginning of the pandemic. We concentrate on a toy model, a *Besag-York-Mollièe* (BYM) model, and COVID-19 cases and deaths in Ohio ($R_0 > 1$). Our analysis shows the presence of a polynomial time component in both cases and deaths, but questions if this can give any public health insight.

1 INTRODUCTION

A major concern with the spread of any virus such as COVID-19 is how fast it initially spreads. In those stages, the number of infected individuals grows exponentially (Brauer et al., 2019; Diekmann et al., 2013; Hilton & Keeling, 2020; Chowell et al., 2016) if the basic reproduction number R_0 is > 1. In the case $R_0 \simeq 1$, the growth rate might not be exactly exponential anymore. Consider for example, the basic SIR model: $\frac{dS}{dt} = -\frac{\beta SI}{N}$, $\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma, I\frac{dR}{dt} = \gamma I$. Early in the evolution of the disease, the second equation gives a first order approximation: $\ln \frac{I(t)}{I(0)} \simeq [\beta S(0) - \gamma I(0)]t$. Note that in particular when $R_0 \simeq 1$, this expansion is inconclusive, and so higher order terms become fundamental. This perspective is common in singularity theory for ODEs, and Hartman-Grobman Theorem implies that nonlinear terms are essentially different from linear around critical points (Arnold, 1993). Recall that R_0 depends on the contact rate and so space and the socio-economic determinants influence it as well (Foster & Selvitella, Submitted; Selvitella & Foster, 2020; Selvitella et al., Submitted). Understanding the spread of COVID-19 early in the pandemic is important for multiple public health reasons, such as avoiding strain on the healthcare system (Miller et al., 2020; Selvitella & Foster, 2020; Foster & Selvitella, Submitted). Understanding the spatial and temporal characteristics of the transmission of COVID-19 during the first stages of the pandemic can also potentially help predict the dynamics of early stages of subsequent waves or novel mutations.

In this work, we are interested in understanding if higher order terms play a role in the spatio-temporal diffusion of COVID-19. The models we are interested in look like a Taylor-type expansion, which roughly speaking looks like:

$$\ln Y(t,x) = c_0(x) + c_1(x)f_1(t - t_x) + c_2(x)f_2(t - t_x) + \dots.$$

Here Y is a count variable, the c's are functions of the spatial components and can be random, and the f's are deterministic, possibly nonlinear, and capture increasing model complexity. This expansion mimics the separation of variables broadly used in solving linear PDEs (Strauss, 2008; John, 1982). The t_x's are space dependent time translations, which can, for example, account for time-lag between different locations. We will concentrate on a Bayesian Hierarchical model, the

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies. Funding acknowledgements go at the end of the paper.

Besag-York-Mollièe (BYM) model (Besag et al., 1991; Blangiardo & Cameletti, 2015; Bernardinelli et al., 1995), and study the number of cases and deaths in Ohio, as an example of with $R_0 > 1$, for the first three weeks following the first COVID-19 case in each county.

Our analysis suggests that when the linear model catches significant components of the count variables, the optimal nonlinearity exponent p is very high and/or unstable, indicating that the basic structure is linear. This has implications for decision makers as nonlinear terms are a source of irregularity with respect to initial data, and so possibly causing uncertainty even in the short-term evolution of a pandemic, which in turn challenges public health policies.

The remaining part of this manuscript is organized as follows. Section 2 is dedicated to the methods, Section 3 to our results and a discussion, while in Section 4, we draw our conclusions.

2 METHODS

Our dataset (Center for Disease Control and Prevention) contains the number of cases and deaths associated with COVID-19 for each of the 88 Ohio counties in the first three weeks, after their first case, which might not be synchronous. The analysis utilized the R-INLA statistical package (R-INLA Software package).

Suppose y_{it} represents the count of cases/deaths at time t with $t = 1, \ldots, T$ with $T = 21$. The index $i = 1, \ldots, n$ identifies the county in Ohio with $n = 88$. We will consider $y_{it} \sim Poiss(\lambda_{it})$, or Negative Binomial (McCullagh & Nedler, 1989). The parameter λ_{it} is assumed to be $\lambda_{it} = E_i \rho_{it}$. While ρ_{it} represents the rate of appearance of new cases/deaths at time t in area/county i, the expected number of cases/deaths in area i is defined as follows. Suppose the population under consideration is partitioned into J classes and r_j is the standardized reference rate for class $j = 1, \ldots, J$ and P_{ij} is the population count of class j in county i. Then the expected number of cases/deaths in area i is given by $E_i = \sum_{j=1}^{J} P_{ij} r_j$. We then use the link function $\eta_{it} = \log(\rho_{it})$ to relate the number of cases/deaths to their spatio-temporal covariates as

$$\log(\rho_{it}) = \eta_{it} = b_0 + u_i + v_i + \Phi_{it}.$$

The intercept b_0 represents the average outcome rate in the entire state of Ohio; the covariate v_i is the area-specific effect modeled as exchangeable, while u_i is modeled as an autoregressive stochastic process (Besag, 1974; Blangiardo & Cameletti, 2015). Suppose that $N(i)$ is the set of neighborhoods of area i and that $\mathbf{u}_{-i}$ represents the area-specific effects excluding county i. Then, the conditional distribution $u_i \mid \mathbf{u}_{-i}$ is given by $u_i \mid \mathbf{u}_{-i} \sim \mathcal{N}\left(\psi_i, \theta_i^2\right)$, with $\psi_i = \mu_i + \sum_{k=1}^{n} r_{ik} \left(u_k - \mu_k\right)$ *and* $\theta_i^2 = s_i^2$ where μ_i and $s_i^2 = \sigma_u^2 / N_i$ are the mean and variance of area i respectively, and $N_i = |N(i)|$ is the number of neighborhoods of area i. Here, σ_u^2 represents the amount of variation between the spatially-structured random effects, and r_{ik} is the indicator of spatial proximity between areas i and k and is defined as $r_{ik} = \phi W_{ik}$ with $W_{ik} = a_{ik}/N_i$ and $a_{ik} = 1$ if i and k are neighbours and 0 otherwise. The factor $\phi > 0$ makes the distribution proper and ensures the positivity of the variance-covariance matrix $(I - \phi W)\, S^2$, with I the $n \times n$ identity matrix, $W = \{W_{ik}\}_{i,k=1}^{n}$, and $S^2 = diag\{s_1^2, \ldots, s_n^2\}$ (Cressie, 1993). With these assumptions, the proper *Conditional AutoRegressive* (CAR) $\mathbf{u}$ is multivariate Normal $\mathbf{u} \sim \mathcal{MVN}\left(\mu, (I - \phi W)\, S^2\right)$, with $\mu = \{\mu_1, \ldots, \mu_n\}$ the mean vector. We will further assume $\phi = 1$ and $\sum_{i=1}^{n} u_i = 0$, a specification called *intrinsic Conditional AutoRegressive* which together with the exchangeable random effect gives rise to the so called *Besag-York-Mollièe* model (BYM) (Besag et al., 1991; Blangiardo & Cameletti, 2015). If further $\mu_i = 0$ for every $i = 1, \ldots, n$, then $u_i \mid \mathbf{u}_{-i} \sim \mathcal{N}\left(\frac{1}{N_i} \sum_{k=1}^{n} a_{ik} u_k, s_i^2\right)$. See (Besag et al., 1991; Best et al., 2005; Lawson, 2009; Lee, 2011) for more details. The temporal component Φ_{it} will be a parametric trend of the form $\Phi_{it} = (\beta + \delta_i) t^p$ for $t = 1, \ldots, T$ and $i = 1, \ldots, n$, but with $p = 1$, or $p > 1$ or a sum of both. Here, β represents the main trend, while δ_i, $\sum_{i=1}^{n} \delta_i = 0$, represents the county i differential trend. We assume homogeneity in the population, namely $E_i = 1$ for every $i = 1, \ldots, n$.

The *Deviance Information Criterion* (DIC) is a measure of model fitting typically used for Bayesian models (Spiegelhalter. et al., 2002) which includes a trade-off between goodness of fit and model complexity. Given a random variable Y, whose distribution depends on a set of parameters θ, the DIC is given by the following formula: $DIC = \bar{D} + p_D$, where p_D represents the *Effective Number of Parameters*, which is computed as $p_D = \bar{D} - D(\bar{\theta})$ with $\bar{D} := E_{\theta|y}\left[D(\theta)\right]$ *and* $D(\bar{\theta}) :=$

Outcome	Linear		Polynomial		Linear + Polynomial	
	P	NB	P	NB	P	NB
Cases	11670 [1] (1)	8792 [1] (1)	8652 [2] (20)	8663 [0.5] (20)	7964 [3] (20)	8557 [10] (20)
Deaths	1746 [1] (1)	1742 [1] (1)	1744 [0.5] (20)	1742 [1] (20)	1717 [10] (20)	1722 [8] (20)

Table 1: This table reports the (i) DIC value of the optimal model in each group, (ii) p of the optimal model $[p]$, (iii) the number n of models considered (n).

$D\left(E_{\theta|y}\left[\theta\right]\right)$. Here, $D(\theta) := -2\log(p(y|\theta))$ represents the *Deviance* and $\bar{D}$ the posterior expectation of the deviance. The DIC can be equivalently rewritten as $\text{DIC} = D(\bar{\theta}) + 2p_D$, form that resembles the *Akaike Information Criterion* (AIC) (James et al., 2013) that the DIC generalizes.

We consider a total of 164 models: (i) y_{it} as the count of cases/deaths ($2\times$), (ii) y_{it} distributed as Poisson or Negative Binomial ($2\times$); and (iii) $p = 0$ ($1\times$) and $p = [0.5 : 0.5 : 10]$ ($20\times$) and the combination of the two ($20\times$)). It is worth noticing that, in our models, we translate the time index of a county-dependent factor $t_x = t_i$, $i = 1, \ldots, n$.

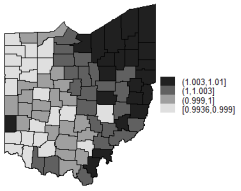

Figure 1: Posterior Mean of the Spatial Main Effect $\zeta_i = \exp(u_i + v_i)$ of the number of deaths and $p = 1$

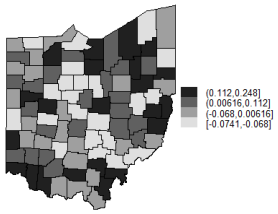

Figure 2: Differential Effect δ_i of the number of deaths and $p = 1$.

3 Results and Discussion

We will report some of the most representative results of our analysis. The first pertains to the linear model of Poisson deaths. Figure 1 and Figure 2 visually represent the distribution of (respectively) spatial and temporal effects on the number of COVID-19 deaths per county in Ohio in the linear case $p = 1$. The spatial effect is given by the posterior mean of the spatial main effect $\zeta_i = \exp(u_i + v_i)$ for $i = 1, \ldots, n$. The temporal effect is given by the differential effect δ_i for $i = 1, \ldots, n$ and represents the extra growth rate of county i with respect to the Ohio growth rate. Darker are those counties where the spatial effect ζ_i and/or the temporal effect δ_i are stronger. The model highlights the presence of spatial correlation between counties. The performances of the linear models are competitive with respect to the other models (Table 1) and are enough to support the presence of a spatio-temporal differential effect, which is the strongest in Northeast Ohio.

Understanding the exact growth rate of the number of cases and deaths has crucial implications for public health and can guide government policy decisions. Note that our analysis includes a single family of models and so it cannot be used alone to guide a government decision on such an important issue. However, interesting points of discussion emerge. The models for deaths are generally better than those for cases. This is speculatively related to the higher uncertainty of determining cases vs deaths early in the pandemic. Based on the DIC, the optimal models are those with Poisson distributed counts for both cases and deaths, linear plus polynomial term ($p = 3$ for cases and $p = 10$ for deaths), which gives some support for the inclusion of a polynomial growth rate early in the pandemic. However, this result has to be taken with caution. The fact that the optimal p grows when the linear term is included might indicate that the extra component is actually not modeling the signal. Recall that nonlinear terms vary much quicker than linear and so their presence causes instability. This instability could be of concern for public health decision makers, as policies might be more sensitive to small variations in such situations. Moreover, highly nonlinear components are hard to distinguish from noise. The fit improvement achieved by the inclusion of a polynomial term might still be related to a problem with the model rather than a true effect. Indeed, the DIC is more stable for the Negative Binomial (Figure 3), which in the case of deaths seem to exclude a nonlinear

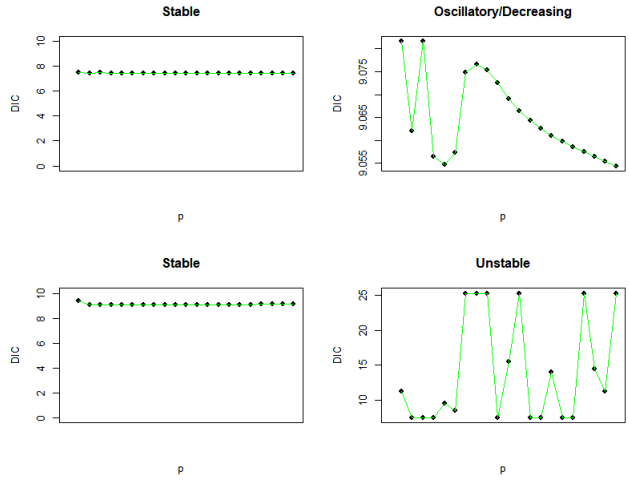

Figure 3: Instability of the optimal p based DIC for our BYM models. Top Left: Poisson Deaths + L and P model; Top Right: NegBin Cases + L and P model; Bottom Left: Poisson Cases + P model; Bottom Right: NegBin Deaths + P model.

effect (Table 1). The models for the number of cases seem more supportive to the inclusion of a nonlinear temporal component, but also this is not fully convincing as the first three weeks of a pandemic might be those with the least reliable data. A highly nonlinear term might detect not a signal, but instead the systematic inaccuracy in the number of the first reported cases.

Another point to make is that the curve $(p, DIC(p))$ varies qualitatively quite a bit. In Figure 3 we show some representative graphs. When the outcome is modeled as Poisson, the search for the optimal p is relatively stable, while when the outcome is Negative Binomial, it is very unstable. Possibly counterintuitively, we interpret this as the Negative Binomial with linear growth fitting our data better, since what remains from a good fit is more noisy and can be mistaken for a very high nonlinearity. The lack of smoothness in p of $DIC(p)$ indicates that including the nonlinear term is possibly not a good idea as the choice of p does not seem univocal.

4 Conclusions

We have discussed the possible inclusion of nonlinear terms in the evolution of COVID-19 early in the pandemic using as a toy model: a Bayesian Hierarchical BYM model and Ohio's case and death counts. The linear model seems preferable (eg.the results for Poisson deaths with a substantial spatial effect). Although in the cases where $R_0 \sim 1$ higher order terms can play an important role in understanding the dynamics of COVID-19, we did not find strong evidence that for our setting, which has $R_0 > 1$, nonlinear terms play an important role in the initial phases of a pandemic. Our analysis is ongoing and we are verifying our results with more complex models and more general situations (eg. all US counties). Since higher order terms in the expansion of the trajectory of COVID-19 cases/deaths can rule the spreading of the disease, especially in critical cases ($R_0 \simeq 1$), their understanding can provide valuable information to policymakers and ultimately lead to positive public health consequences.

Data and Code availability

Data is publicly available and code is available upon request.

Acknowledgments

AMS and KLF would like to thank their families for their constant support.

REFERENCES

V. I. Arnold. *Geometrical Methods in the Theory of Ordinary Differential Equations*, volume 250. Springer Verlag, 1993.

L. Bernardinelli, D. Clayton, C. Pascutto, C. Montomoli, M. Ghislandi, and M. Songini. Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*, 14:2433–2443, 1995.

J. Besag. Spatial interaction and the statistical analysis of the lattice systems (with discussion). *Journal of Royal Statistical Society, Series B*, 36:192–236, 1974.

J. Besag, A. York, and J. Mollie. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–59, 1991.

N. Best, S. Richardson, and A. Thomson. A comparison of bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, 14:35–59, 2005.

M. Blangiardo and M. Cameletti. Bayesian regression and hierarchical models. In *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley and Sons, 2015.

F. Brauer, C. Castillo-Chavez, and Z. Feng. *Mathematical Models in Epidemiology*. Springer, 2019.

Center for Disease Control and Prevention, 2020. URL `https://www.cdc.gov/`.

G. Chowell, C. Viboud, L. Simonsen, and S. Moghadas. Characterizing the reproduction number of epidemics with early sub-exponential growth dynamics. *Journal of the Royal Society Interface*, 13:20160659, 2016.

N. Cressie. *Statistics for Spatial Data*. Wiley, 1993.

O. Diekmann, H. Heesterbeek, and T. Britton. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press, 2013.

K. L. Foster and A. M. Selvitella. On the relationship between covid-19 reported fatalities early in the pandemic and national socio-economic status. Submitted.

J. Hilton and M. J. Keeling. Estimation of country-level basic reproductive ratios for novel coronavirus (covid-19) using synthetic contact matrices. *medRxiv*, pp. 1–7, 2020.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *Introduction to Statistical Learning*. Springer, 2013.

F. John. *Partial Differential Equations*. Springer Verlag, 1982.

A. Lawson. *Bayesian Disease Mapping. Hierarchical Modeling in Spatial Epidemiology*. Chapman and Hall/CRC, 2009.

D. Lee. A comparison of conditional autoregressive models used in bayesian disease mapping. *Spatial and Spatio-Temporal Epidemiology*, 2:79–89, 2011.

P. McCullagh and J. A. Nedler. *Generalized Linear Models*. Chapman and Hall/CRC, 1989.

I. F. Miller, A. D. Becker, B.T. Grenfell, and J.E. Metcalf. Disease and healthcare burden of covid-19 in the united states. *Nature Medicine*, 26:1212–1217, 2020.

R-INLA Software package. https://www.r-inla.org/home.

A. M. Selvitella and K. L. Foster. Societal and economic factors associated with covid-19 indicate that developing countries could suffer the most. *Technium Social Sciences Journal*, 10:637–644, 2020.

A. M. Selvitella, L. Carolan, J. Smethers, C. Hernandez, and K. L. Foster. A spatio-temporal investigation of the growth rate of covid-19 incidents in ohio early in the pandemic. Submitted.

D. J. Spiegelhalter., N. G. Best, B. P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical society, Series B*, 64:583–639, 2002.

W. A. Strauss. *Partial Differential Equations : An Introduction*. John Wiley and Sons, 2008.