

A spatial analysis of the environmental correlates of COVID-19 incidence in the provinces in Spain

Antonio Paez^{*,a}, Fernando A. Lopez^b, Tatiane Menezes^c, Renata Cavalcanti^d,
Maira Galdino da Rocha Pitta^d

^a*School of Geography and Earth Sciences, McMaster University, 1281 Main St W, Hamilton, ON, L8S 4K1, Canada*

^b*Departamento de Metodos Cuantitativos, Ciencias Juridicas, y Lenguas Modernas, Universidad Politecnica de Cartagena, Calle Real Numero 3, 30201, Cartagena, Murcia, Spain*

^c*Departamento de Economia, Universidade Federal de Pernambuco, Av dos Economistas, s/n - Cidade Universitária, Recife - PE, 50670-901, Brasil*

^d*Núcleo de Pesquisa em Inovação Terapêutica NUPIT / UFPE, Av. Prof. Moraes Rego, 1235 - Cidade Universitária, Recife, PE, CEP 50670-901, Brazil*

Abstract

Spreading with astonishing speed, the novel SARS-CoV2 has swept the globe, causing enormous stress to health systems and prompting social distance guidelines and mandates to arrest its progress. While there is encouraging evidence that early public health interventions have slowed the spread of the virus, this has come at a high cost as the global economy is brought to its knees. How and when to ease restrictions to movement hinges in part on the question whether SARS-CoV2 will display seasonality associated with variations in temperature and humidity. In this research, we address this question by means of a spatial analysis of the incidence of COVID-19 in the provinces in Spain. Use of a spatial SUR approach allows us to model the incidence of reported cases of the disease per 100,000 population, as a function of temperature and humidity, while controlling for GDP per capita, population density, percentage of older adults in the population, and presence of mass transit systems. Our results indicate that incidence of the disease is lower at higher temperatures. The evidence with respect to humidity is more mixed, with coefficients for this variable that are significant in only some equations. Our control variables also yield interesting insights, as population density and percentage of older adults display negative associations with incidence of COVID-19, whereas the presence of mass transit systems in the province is associated with a greater incidence of the disease.

*Corresponding Author

Email addresses: paezha@mcmaster.ca (Antonio Paez), fernando.lopez@upct.es (Fernando A. Lopez), tatiane.menezes@ufpe.br (Tatiane Menezes), renata.vcsantos@gmail.com (Renata Cavalcanti), mgrpitta@ufpe.br (Maira Galdino da Rocha Pitta)

Introduction

From a small outbreak linked to a live animal market at the end of 2019 to a global pandemic in a matter of weeks, the SARS-CoV2 virus has threatened to overrun health systems the world over. In efforts to contain the spread, numerous governments in many nations and regions have either recommended or mandated social distancing measures, and as of this writing, millions of people in five continents shelter in place. There are encouraging signs that these measures have arrested the spread of the virus where they have been implemented, and have thus helped to keep a bad situation from becoming even worse (e.g., 2020). However, this has come at a high cost, and the consequences for all spheres of the economy, social, and cultural life have been dire (e.g., Fernandes, 2020; Luo and Tsang, 2020). As a result, there is a sense of urgency to anticipate the progression of the pandemic, in order to plan for progressive lifting of restrictions to movement and social contact (e.g., Kissler et al., 2020). Needless to say, this has become a delicate, and politically charged, balancing act between public health and the economy (Gong et al., 2020).

A salient question in the debate on how and when to ease social distancing measures is whether the virus will display seasonal variations. Earlier, diverse studies have shown the effect of temperature and humidity on the incidence of influenza (e.g., Mäkinen et al., 2009; Jaakkola et al., 2014; Kudo et al., 2019). Jaakkola et al. (2014), for example, found that a decrease of temperature and absolute humidity precedes the onset of symptoms of influenza A and B viruses by 3 days in places where the temperature is low. After the 2002-2004 outbreak of SARS, researchers also began to investigate the relationship between these factors and SARS-CoV. In this way, Casanova et al. (2010) used two surrogates, namely the gastroenteritis (TGEV) and mouse hepatitis viruses (MHV), to find that virus inactivation was more rapid at temperatures of 20C than 4C, and at 40C than 20C; in terms of humidity, these researchers reported that survival of the virus was lower at moderate relative humidity levels. In a similar vein, but working directly with SARS-CoV, Chan et al. (2011) found that viability of the virus was lost at temperatures higher than 38C and relative humidity superior to 95%.

While existing research on similar pathogens suggests that SARS-CoV is more stable and potentially easier to transmit in conditions of low temperature and low humidity, it is far from certain that this will also be the case with the novel SARS-CoV2. If such is the case, there is the possibility of easing restrictions to social contact as the weather warms; however, weeks or possibly months of costly measures could become undone if not, and the restrictions are lifted prematurely. Not surprisingly, given the stakes involved, this issue has already triggered a lively debate.

Some of what is thought about the possible seasonality of COVID-19 is based on analogies to the patterns of other known respiratory viruses. However, de Ángel Solá et al. (2020) note that “not all seasonal respiratory viruses experience the same spatiotemporal patterns” (section 4). This urges caution when extrapolating from known viruses, and indeed, the evidence in this respect

is inconclusive. At a global scale, whereas de Ángel Solá et al. (2020) see less risk in the Caribbean Basin, Coelho et al. (Coelho et al., 2020) warn that at least during the exponential phase, expansion of the virus is not driven by climate. Similarly, whereas Araujo and Naimi (2020) argue that spread of SARS-CoV2 will likely be constrained by climate, Harbert et al. (2020) remain unconvinced that spatial modelling can currently discriminate the distribution of the disease on the basis of climate, at least in the United States. To further complicate matters, much of the relevant work has yet to be peer-reviewed and therefore is open to change (see for example the challenge of Harbert et al. (2020) to Araujo and Naimi (2020)). In the United States, the National Academy of Sciences, Engineering, and Medicine was engaged by the Office of the Executive for guidance on this matter (see National Academies of Sciences, Engineering and Medicine, 2020). Their conclusion summarizes the situation well (see p. 6): “Some limited data support a potential waning of cases in warmer and more humid seasons, yet none are without major limitations... Additional studies as the SARS-CoV-2 pandemic unfolds could shed more light on the effects of climate on transmission.”

With the above considerations in mind, our objective with this paper is to contribute to the knowledge basis regarding the spread of COVID-19 and the influence of environmental factors, particularly temperature and humidity. Hence, here we report results from a spatial analysis of incidence of COVID-19 in fifty provinces in Spain. Spain is one of the countries hardest hit by the virus, and enacted lockdown measures on March 16, 2020, in response to a rapidly growing outbreak of COVID-19. We combine data on reported cases of the disease with metereological information, to create a spatio-temporal dataset covering a period of 30 days. We then join this dataset with provincial-level economic and demographic information to act as controls to our key environmental variables. These data are analyzed using a spatial SUR approach, which allows us to account for residual spatial autocorrelation. The results provide evidence of the effect of temperature on the incidence of COVID-19, as **NOTE ABOUT THE MAGNITUDE OF THE EFFECT.** The evidence concerning humidity is more mixed: while the direction of the effect is negative, as anticipated, the coefficients for this variable are only significant in some of equations in the system. Our control variables also provide some intriguing insights. The results of this analysis provide support to the hypothesis of seasonality of the novel SARS-CoV2, and should be of interest to public health officials and policy makers wrestling with the question of the trajectory of the pandemic.

Please note that this paper is prepared as a reproducible research document. The source R Markdown document, as well as all data and code needed to reproduce/review/extend the analysis can be obtained from the following repository:

Include link to repository

Context and Data

Covid-19 in Spain

When was the first case reported? When did the country close international arrivals? When did initial measures of social distancing were implemented? The lockdown?

Selection of Variables

Explain the rationale for selecting the variables.

For example: The literature about COVID-19 suggested that population density is the one of the most important proliferate cause of these viscous, however this ill spread with different intensity at big cities of the world. Controlling for socioeconomic characteristics the objective of this paper is observe the effect of climate on COVID-19 proliferation.

Some questions are important to point out. Is not evident present a diary association between the temperature (resp. humidity) with declarate case. There is evidences (ref) that a time lag is necessary between the day that the individual infect and the case is diagnosticate

Sources of Data

Our dataset includes information about the daily number of cases of COVID-19 reported in Spain ¹ at the provincial level (NUTIII in eurostat terminology) from March 13th to April 11th (30 days and 50 spatial units). The Spanish National Government publishes periodic updates at the regional level (NUTII) and the information is also released at the provincial level as part of a collaborative project ². This information is compiled from several sources, mainly the official web pages of the Spanish regional governments. In addition, we consider two sets of explanatory variables. The first one, and the focus of this research, is set of two environmental variables, namely temperature and humidity. The second set provides some relevant controls for multivariate analysis, and refers to economic and demographic attributes of the province. Table 1 shows the descriptive statistics and the provenance of the data.

Methods: Spatial SUR

The econometric model

The baseline model proposed in this paper is a classical SUR model without spatial effects (from here, SUR-SIM). The classical expression for this model in stacked form is,

¹cases tested

²https://code.monterea34.com:4443/numeroteca/covid19/-/blob/master/README_providencialdata19.md

Table 1: Descriptive statistics

Variable	Note	Min	Mean	Max	SD	Source
SARS-19 Incidence	Incidence in reported cases of SARS-19 per 100,000 people	0.38	153.61	1149.36	186.23	Datalista
GDPpc	GDP per capita in €1.000s	16.67	22.51	36.00	4.77	INE
Older	Percentage of people aged 65 and older in the province	15.16	21.03	31.36	3.95	INE
Population Density	Population density in the province in people per sq.km	8.60	140.04	829.76	181.25	INE
Mean Temperature	Mean temperature in province by date, in C	1.00	12.18	23.20	3.67	AEMET
Humidity	Relative humidity in province by date	2.00	77.82	100.00	10.37	MAPA

Note:

Datalista:

INE (Instituto Nacional de Estadística): <https://www.ine.es/>

AEMET (Agencia Estatal de Meteorología): <http://eportal.mapa.gob.es>

MAPA (Ministerio de Agricultura, Pesca y Alimentación): <http://eportal.mapa.gob.es>

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_T \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_1 \\ \vdots \\ \beta_T \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{bmatrix} \quad (1)$$

where $y_t = (y_{1t}, \dots, y_{Nt})$ is a $N \times 1$ vector, and in our case y_{st} is the incidence ratio in the province s ($s = 1, \dots, N$) the day t ($t = 1, \dots, T$); $X_t = (X^1, \dots, X^{k_t})$ is a $N \times k_t$ matrix of the k_t independent variables, $X_i = (X_{st}^i)$; $\beta_t = (\beta_{1t}, \dots, \beta_{Nt})$ is a vector of coefficients and $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{Nt})$ is the error vector.

the more relevant characteristic in SUR model, is the dependence among error vector,

$$E[\epsilon_t \epsilon'_{t'}] = \sigma_{tt'} \quad (2)$$

Note that this specification is highly flexible, allowing changes in the coefficients β_{it} in order to modulate the effect of X_{it}^i on y_t . This flexibility can be reduced and it is possible to impose restrictions considering time constant some β coefficients. In this case, we can reformulate the coefficients expression of $\beta_t = (\beta_1, \dots, \beta_r, \beta_{r+1}, \dots, \beta_{Nt})$ considering the first r coefficients constant.

The equation (1) can be rewrite in compact form,

$$y = X\beta + \epsilon \quad (3)$$

where....

Like in case of cross-section, it is possible identify spatial autocorrelation in the residuals of (1) and several Lagrange Multipliers tests has been develop to test the null of spatial independence (see López et al. (2014)). In case of reject the null alternative several specifications has been propose to include spatial effects (Anselin (1988), see also Anselin (2016)). In this paper we consider the

specific spatial SUR model tahr incorpore a spatial lag of the dependent variable as an explanatory factor. The stack expresion for the spatial lag SUR model (SUR-SLM),

$$\mathbf{Ay} = \mathbf{X}\beta + \epsilon\epsilon = \mathbf{N}(\mathbf{0}, \boldsymbol{\Omega}) \quad (4)$$

where $\mathbf{A} = I_{TN} - \boldsymbol{\Lambda} \otimes \mathbf{W}$ with $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_T)$.

This specification assumes that incidence in a province (y_{st}) at time t is partially determined by the weighted average (Wy_{st}) of incidence in neighbouring provinces. Parameters of spatial dependence are estimated for each time period λ_t and identifies the intensity and the sign of the impacts of neighbourhood. It is possible test the null hypothesis of identtical levels of spatial dependence ($\lambda_i = \lambda_j, \forall i, j$). The correspond Wald test is available in the R package **spsur**

The SUR-SLM model can be estimated by maximum likelihood (López et al. (2014)) or instrumental variables (Mínguez et al. (2019)). We considerer this methodology to estimate the model.

Analysis and Results

Exploratory Data Analysis

We begin with the exploratory analysis of the data.

Figure 1 shows the geographical variation in the incidence of the disease, as well as its temporal progression in weekly averages. Our analysis begins on March 13, which is the first date when every province had reported at least one case of COVID-19. It can be seen that the highest incidence at this early date was in the provice of Álava, in the North of Spain. While not the most populous province, with a population of only 331,549, Álava has the highest GDP per capita of all provinces. Vitoria, its main city, is the capital of the Basque country and has been the focus of efforts to develop a “Global Basque City” (Meijers et al., 2008), along with San Sebastian and Bilbao. The other early focus of the pandemic in Spain was Madrid, which is the most populous province in the country and has the second highest GDP per capita after Álava. The early outbreaks in these two provinces can be traced throughout the progression of the pandemic over time, although by the end of the period under consideration, other provinces had matched and even surpassed their incidence rates, including Segovia and Soria to the north of Madrid, and Ciudad Real and Albacete to the south.

To visualize the distribution of temperature and humidity we aggregate the provinces by Autonomous Community. In Figure ?? the communities have been sorted by latitude, so that Principado de Asturias is the northernmost community, and Canarias the southernmost. There is a relatively wide range of values both within and between provinces over the 30 day period examined. The top panel of the figure shows the distribution of mean temperature. The lowest mean temperature for a community on any given day is 2.8C, and the highest is 22.4C for a range of approximately 20 degrees. Likewise, there is a great deal of variability in humidity, as seen in the bottom panel of the figure, where the

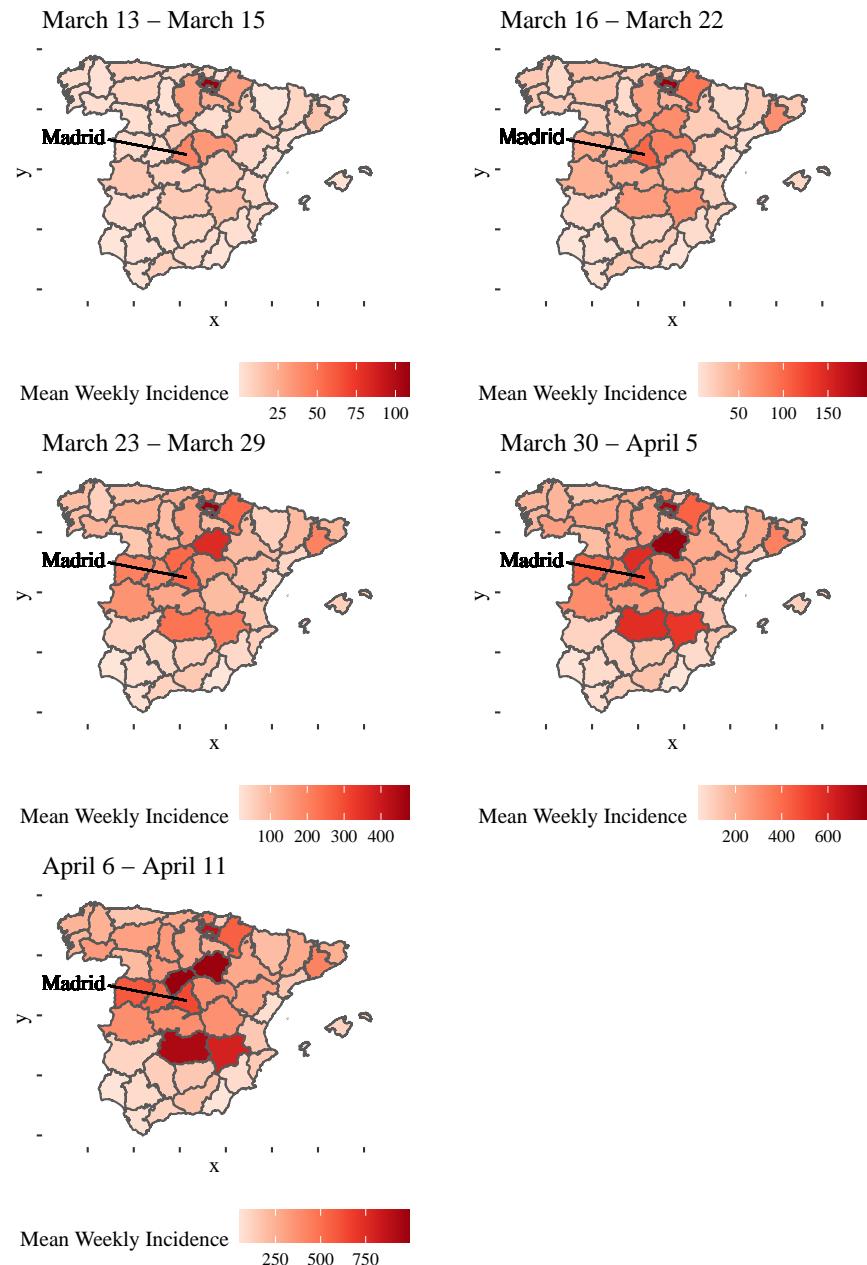


Figure 1: Mean weekly incidence of COVID-19 by province, in reported cases by 100,000 people (map omits Islas Canarias)

lowest mean humidity for any community is 48.3 and the highest is 99.6. The actual values for the provinces display somewhat more variability even.

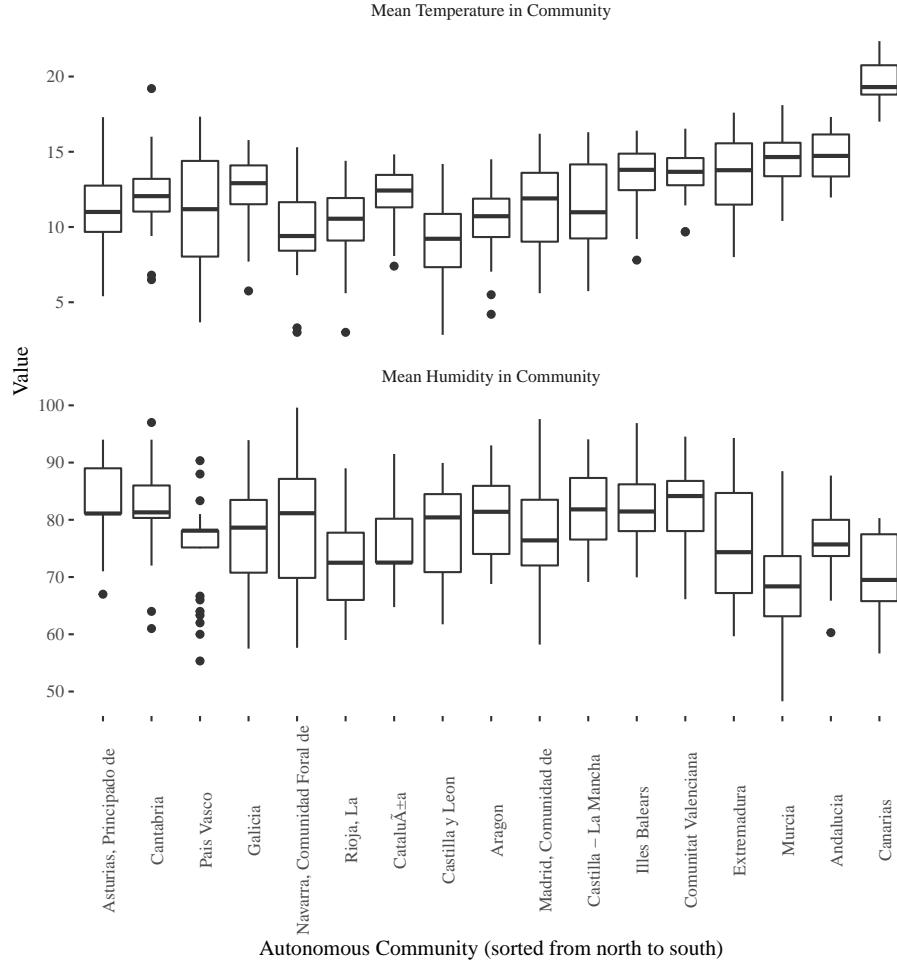


Figure 3 shows the distribution of daily correlations of the independent variables with incidence of COVID-19, after log-transforming all variables. It can be seen there that the correlation of GDPpc and temperature (in its three definitions) have the strongest positive and negative correlations with incidence, respectively. Percentage of older adults displays somewhat weaker negative correlations with incidence, as does density. It can be seen that the humidity variable, in its three forms, tends to be positively correlated with incidence of COVID-19.

SUR Systems

The goodness of fit of the three systems of equations is shown in Figure 4.

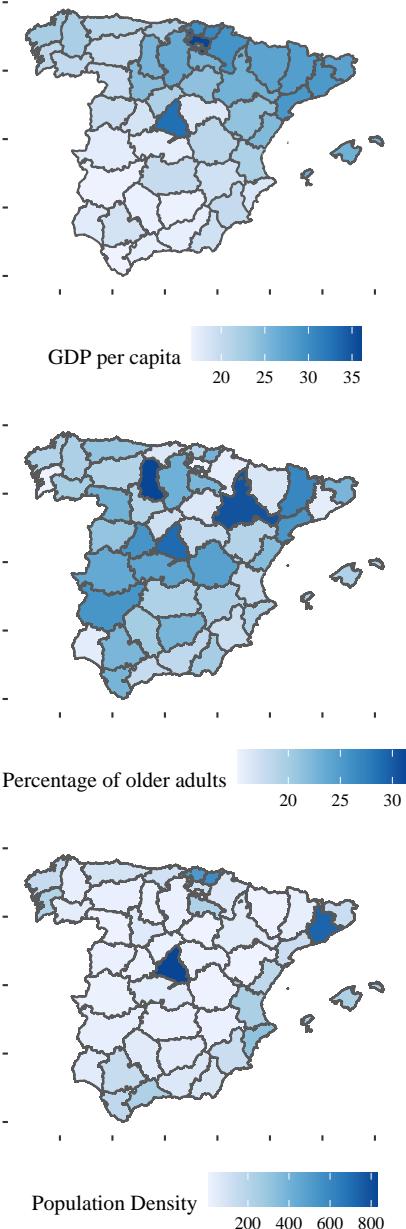


Figure 2: Spatial distribution of control variables by province (maps omits Islas Canarias)

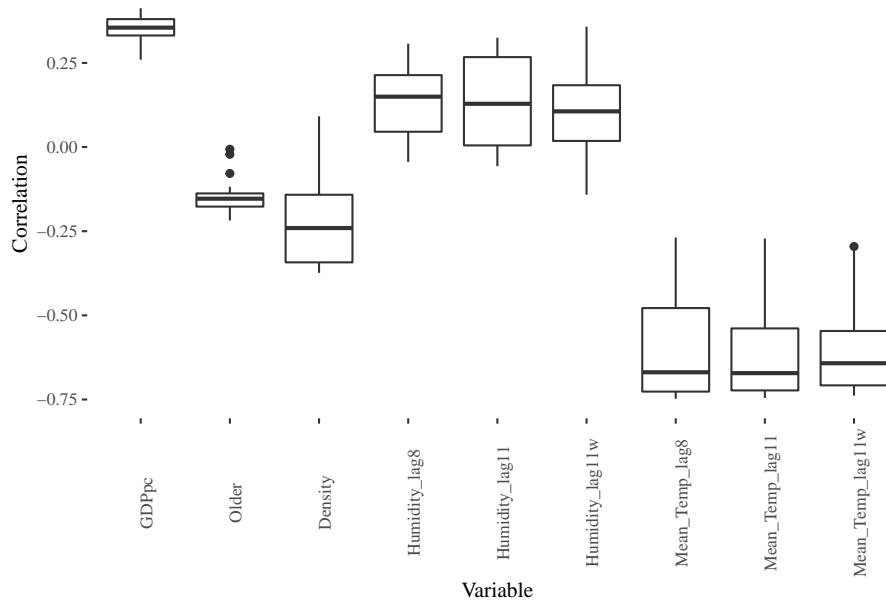


Figure 3: Distribution of daily correlations of the independent variables with daily incidence of COVID-19 (all variables have been log-transformed)

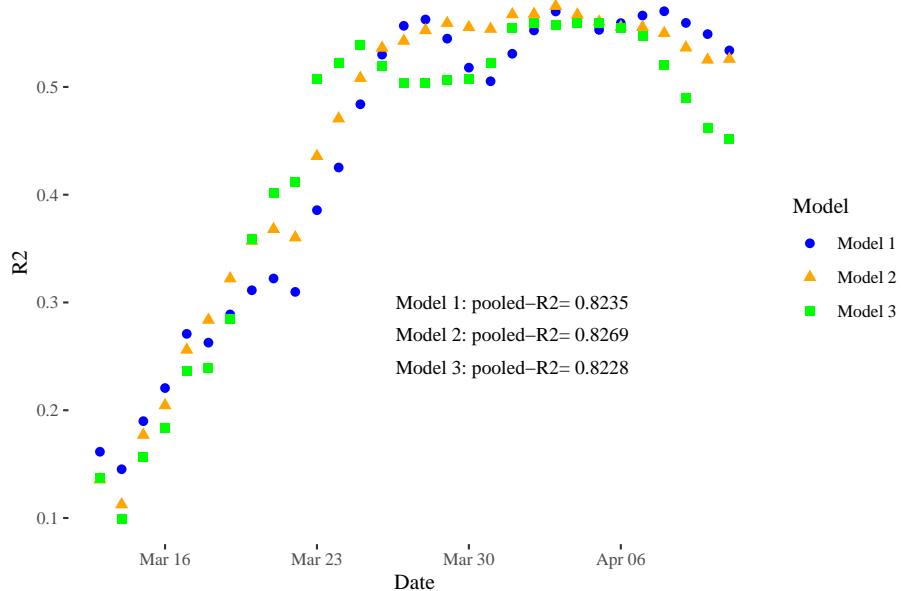


Figure 4: Goodness of fit of the SUR systems: by date and pooled

Discussion

Possibly do some simulations with the model.

Concluding Remarks

More words go here.

Acknowledgments

Add acknowledgments as appropriate in final draft.

The following R packages were used in the course of this investigation and the authors wish to acknowledge their developers: `aemet` [], `ggthemes` (Arnold, 2019), `gridExtra` (Auguie, 2017), `kableExtra` (Zhu, 2019), `knitr` (Xie, 2015, 2014), `lubridate` (Grolemund and Wickham, 2011), `plm` (Millo, 2017), `rticles` (Allaire et al., 2020), `sf` (Pebesma, 2018), `spdep` (Bivand et al., 2013), `spsur` (Angulo et al., 2020) `tidyverse` (Wickham et al., 2019), `units` (Pebesma et al., 2016).

References

- Allaire, J., Xie, Y., R Foundation, Wickham, H., Journal of Statistical Software, Vaidyanathan, R., Association for Computing Machinery, Boettiger, C., Elsevier, Broman, K., Mueller, K., Quast, B., Pruij, R., Marwick, B., Wickham, C., Keyes, O., Yu, M., Emaasit, D., Onkelinx, T., Gasparini, A., Desautels, M.-A., Leutnant, D., MDPI, Taylor and Francis, Ögreden, O., Hance, D., Nüst, D., Uvesten, P., Campitelli, E., Muschelli, J., Kamvar, Z.N., Ross, N., Cannoodt, R., Luguern, D., Kaplan, D.M., 2020. Rtitles: Article formats for r markdown.
- Angulo, A., Lopez, F.A., Minguez, R., Mur, J., 2020. Spsur: Spatial seemingly unrelated regression models.
- Anselin, L., 1988. Spatial econometrics: Methods and models, Studies in operational regional science. Kluwer Academic Publishers, Dordrecht.
- Anselin, L., 2016. Estimation and testing in the spatial seemingly unrelated regression (sur). Geoda Center for Geospatial Analysis; Computation, Arizona State University. Working Paper 2016-01.
- Araujo, M.B., Naimi, B., 2020. Spread of sars-cov-2 coronavirus likely to be constrained by climate. medRxiv.
- Arnold, J.B., 2019. Ggthemes: Extra themes, scales and geoms for 'ggplot2'.
- Auguie, B., 2017. GridExtra: Miscellaneous functions for "grid" graphics.
- Ángel Solá, D.E. de, Wang, L., Vázquez, M., Lázaro, P.A.M., 2020. Weathering the pandemic: How the caribbean basin can use viral and environmental patterns to predict, prepare and respond to covid-19. Journal of Medical Virology.
- Bivand, R.S., Pebesma, E., Gomez-Rubio, V., 2013. Applied spatial data analysis with R, second edition. Springer, NY.
- Casanova, L.M., Jeon, S., Rutala, W.A., Weber, D.J., Sobsey, M.D., 2010. Effects of air temperature and relative humidity on coronavirus survival on surfaces. Appl. Environ. Microbiol. 76, 2712–2717.
- Chan, K., Peiris, J., Lam, S., Poon, L., Yuen, K., Seto, W., 2011. The effects of temperature and relative humidity on the viability of the sars coronavirus. Advances in virology 2011.
- Coelho, M.T.P., Rodrigues, J.F.M., Medina, A.M., Scalco, P., Terribile, L.C., Vilela, B., Diniz-Filho, J.A.F., Dobrovolski, R., 2020. Exponential phase of covid19 expansion is not driven by climate at global scale. medRxiv.
- Fernandes, N., 2020. Economic effects of coronavirus outbreak (covid-19) on the world economy. Available at SSRN 3557504.
- Gong, B., Zhang, S., Yuan, L., Chen, K.Z., 2020. A balance act: Minimizing economic loss while controlling novel coronavirus pneumonia. Journal of Chinese Governance 1–20.
- Grolemund, G., Wickham, H., 2011. Dates and times made easy with lubridate. Journal of Statistical Software 40, 1–25.
- Harbert, R.S., Cunningham, S.W., Tessler, M., 2020. Spatial modeling cannot currently differentiate sars-cov-2 coronavirus and human distributions on the basis of climate in the united states. medRxiv.

- Jaakkola, K., Saukkoriipi, A., Jokelainen, J., Juvonen, R., Kauppila, J., Vainio, O., Ziegler, T., Rönkkö, E., Jaakkola, J.J., Ikäheimo, T.M., 2014. Decline in temperature and humidity increases the occurrence of influenza in cold climate. *Environmental Health* 13, 22.
- Kissler, S.M., Tedijanto, C., Goldstein, E., Grad, Y.H., Lipsitch, M., 2020. Projecting the transmission dynamics of sars-cov-2 through the postpandemic period. *Science* eabb5793. doi:10.1126/science.abb5793
- Kudo, E., Song, E., Yockey, L.J., Rakib, T., Wong, P.W., Homer, R.J., Iwasaki, A., 2019. Low ambient humidity impairs barrier function and innate resistance against influenza infection. *Proceedings of the National Academy of Sciences* 116, 10905–10910.
- Lancastle, N.M., 2020. Is the impact of social distancing on coronavirus growth rates effective across different settings? A non-parametric and local regression approach to test and compare the growth rate. *medRxiv*.
- López, F.A., Mur, J., Angulo, A., 2014. Spatial model selection strategies in a sur framework. The case of regional productivity in eu. *The Annals of Regional Science* 53, 197–220.
- Luo, S., Tsang, K.P., 2020. How much of china and world gdp has the coronavirus reduced? Available at SSRN 3543760.
- Mäkinen, T.M., Juvonen, R., Jokelainen, J., Harju, T.H., Peitso, A., Bloigu, A., Silvennoinen-Kassinen, S., Leinonen, M., Hassi, J., 2009. Cold temperature and low humidity are associated with increased occurrence of respiratory tract infections. *Respiratory medicine* 103, 456–462.
- Meijers, E., Hoekstra, J., Aguado, R., 2008. Strategic planning for city networks: The emergence of a basque global city? *International Planning Studies* 13, 239–259. doi:10.1080/13563470802521440
- Millo, G., 2017. Robust standard error estimators for panel models: A unifying approach. *Journal of Statistical Software* 82, 1–27. doi:10.18637/jss.v082.i03
- Mínguez, R., López, F., Mur, J., 2019. ML versus iv estimates of spatial sur models: Evidence from the case of airbnb in madrid urban area. *The Annals of Regional Science* 1–35.
- National Academies of Sciences, Engineering and Medicine, 2020. Rapid expert consultation on sars-cov-2 survival in relation to temperature and humidity and potential for seasonality for the covid-19 pandemic (april 7, 2020). The National Academies Press, Washington, DC. doi:doi:10.17226/25771
- Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* 10, 439–446. doi:10.32614/RJ-2018-009
- Pebesma, E., Mailund, T., Hiebert, J., 2016. Measurement units in R. *R Journal* 8, 486–494. doi:10.32614/RJ-2016-061
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4, 1686. doi:10.21105/joss.01686

Xie, Y., 2014. Knitr: A comprehensive tool for reproducible research in R, in: Stodden, V., Leisch, F., Peng, R.D. (Eds.), Implementing Reproducible Computational Research. Chapman; Hall/CRC.

Xie, Y., 2015. Dynamic documents with R and knitr, 2nd ed. Chapman; Hall/CRC, Boca Raton, Florida.

Zhu, H., 2019. KableExtra: Construct complex table with 'kable' and pipe syntax.