

Reproducibility review of: Analysis of the spatial and temporal pattern of COVID-19 incidence rate across Germany

Eftychia Koukouraki 

2026-06-09



This report is part of the reproducibility review at the AGILE conference. For more information see <https://reproducible-agile.github.io/>. This document is published on OSF at <https://doi.org/10.17605/OSF.IO/DZURB>. To cite the report use

Koukouraki, E. (2026, April). Reproducibility review of: Analysis of the spatial and temporal pattern of COVID-19 incidence rate across Germany. <https://doi.org/10.17605/OSF.IO/DZURB>

Reviewed paper

Lautenbach, S., Maurer, M., and Zipf, A.: Analysis of the spatial and temporal pattern of COVID-19 incidence rate across Germany, AGILE GIScience Ser., 7, 9, <https://doi.org/10.5194/agile-giss-7-9-2026>, 2026.

Summary

The paper investigates the effect of five socio-economic factors (votes for right-winged populist party, share of foreigners, share of highly qualified employees, long-term unemployment rate, share of incoming commuters) on different phases of the COVID-19 pandemic. The study uses data from the Robert Koch Institute analysed with a quasi-Poisson generalized linear model using spatial eigenvector mapping and performed hierarchical clustering on the regression coefficients. The authors shared the scripts via an anonymous GitHub link: <https://anonymous.4open.science/r/covid19germany-5403>. For the purposes of this reproducibility review, we verified the functionality of all scripts and the reported results in Figures 1-7 and in Tables 2 and 3. The reproduced results were mostly in accordance with the reported results and therefore the reproduction of the paper is considered **partially successful**.

Reproducibility reviewer notes

Data and code sharing

The data, code, intermediate results and documented instructions to reproduce this paper were initially shared with the reproducibility committee with the following anonymous GitHub link <https://anonymous.4open.science/r/covid19germany-5403>. The authors have licensed this repository under the GNU General Public License. The materials are packed as a single ZIP file, which contains a README file, an R project file, three R Markdown notebooks and one directory with some raw and intermediate data. The R Markdown files contain the code for pre-processing the data, calculating the spatial autocorrelation time series, performing regression modelling, mapping the spatial eigenvectors and performing hierarchical clustering. Under the *data* directory are data files (in .csv or .Rdata format) which are necessary to run the aforementioned scripts. This repository, as provided by the authors, has been copied to the storage of the reproducibility review repository under *materials shared by the authors* (see <https://osf.io/dzurb/files/osfstorage>).

Computational environment

We followed the requirements mentioned in the README file to recreate the computational environment, including the R package versions, as closely as possible. We used R version 4.4.1 and RStudio 2026.01.1+403. Most packages had to be installed manually, one by one. Some also required additional packages or libraries to be installed on the system for successful installation. Ultimately, we managed to install the following versions of the R packages listed in the requirements:

- rmdformats 1.0.4
- tidyverse 2.0.0
- ggplot2 4.0.2
- GGally 2.4.0
- ggpubr 0.6.3
- spatialreg 1.4.2
- lubridate 1.9.5
- sf 1.1.0
- spdep 1.4.2
- tmap 4.2
- readODS 2.3.2

Moreover we had GDAL 3.4.1, GEOS 3.10.2, PROJ 8.2.1, and sqlite3 3.37.2 installed on our system. The hardware on which the reproduction was run consists of 16.0 GiB of memory and an i7-1185G7 @ 3.00GHz $\times$ 8 processor. The operating system is Ubuntu 22.04.5 LTS.

Runtime

The code was executed chunk by chunk in the three R Markdown notebooks. All code chunks were executed quickly, except from the one that calculates the global Moran's I time series in the SAC.Rmd file (lines 78-114), which returned after around half an hour.

Reproduction efforts

The repository was overall tidy and well organized. The data links in the manuscript were provided in section 3.1, namely *Data*, but for the Robert Koch Institute data, the link was wrong and pointed to 404 (footnote 1), and for the socioeconomic data from inkar.de (footnote 2), the reader would need more instructions beyond the link to find the right dataset.

We confirmed the functionality of all three .Rmd files. For data_preprocessing.Rmd, the execution was possible using the intermediate results provided by the authors in the anonymised GitHub repository.

The workflow reproduced a similar covidKreise.Rdata file compared to the one provided by the authors, but not identical. One difference that we could quickly spot is that there is an additional “LKZ” column in the reproduced file. In the case of inkarKreisstatistik_preprocessed.Rdata, it was not obvious how to get from the raw to the pre-processed data from data_preprocessing.Rmd alone.

In case the data_analysis.Rmd was executed first, without having executed the data_preprocessing.Rmd before, the shapefile with the federal states will not be in the data directory and this will cause an error (lines 34-36). There was also a typo in line 856 (it should be method=“euclidean”), which we had to correct on our end in order to proceed. Finally, the VIF for the Auslanderanteil was found equal to 2.34 instead of the reported 2.4 (lines 108-110 of data_analysis.Rmd), which we consider minor as a difference.

Reproduced tables

Some minor differences were spotted in the reproduction of Tables 2 and 3, which do not affect the interpretation of the results. In Table 2, the p-value for the “Sporadic cases” was found equal to 0.0104, as opposed to the reported 0.0127. In Table 3, there are some minor rounding differences compared to the manuscript (e.g., Auslaenderanteil found 0.0068, instead of 0.0069).

Name	MoransI	MoransIpval
2 Sporadic cases	0.0267869	0.0104
3 first wave	0.5090324	0.0001
4 Summer plateau 2020	0.3669288	0.0001
5 second wave	0.5850656	0.0001
6 third wave	0.5075857	0.0001
7 Summer plateau 2021	0.3643340	0.0001
8 fourth wave	0.7461868	0.0001
9 fifth wave	0.5420816	0.0001

Figure 1: Global Moran’s I for the aggregated incidence rate of each pandemic period. Corresponds to Table 2 of the reproduced paper. Output of lines 139-158 of SAC.Rmd.

Reproduced figures

All figures were perfectly reproduced, except for the dendrogram (Figure 6 in the manuscript). In this case the output figure had no dimensions and all the labels overlap, so this result could not be verified.

Communication with the author

We communicated these remarks to the authors by email, suggesting improvements to both the repository and the manuscript. The corresponding author responded fast and positively, but we ultimately received no further updates before the reproducibility review had to be completed.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.2102606	0.0350970	-34.483	< 2e-16 ***
Stimmenanteile.AfD	0.0150164	0.0008644	17.371	< 2e-16 ***
Hochqualifizierte	-0.0017282	0.0009278	-1.863	0.063282 .
Langzeitarbeitslosenquote	-0.0461815	0.0035231	-13.108	< 2e-16 ***
Einpendler	-0.0013015	0.0003822	-3.405	0.000732 ***
Auslaenderanteil	0.0068751	0.0011180	6.150	1.97e-09 ***
vec4	-3.9898803	0.7826132	-5.098	5.43e-07 ***
vec6	0.8184963	0.2044539	4.003	7.52e-05 ***
vec13	0.0231083	0.1638101	0.141	0.887891
vec16	-0.3524076	0.0801053	-4.399	1.41e-05 ***
vec20	0.9131945	0.2012869	4.537	7.68e-06 ***
vec21	0.8261690	0.1731283	4.772	2.61e-06 ***
vec36	0.2954645	0.2012154	1.468	0.142825
Auslaenderanteil:vec4	0.0718781	0.0168520	4.265	2.52e-05 ***
Auslaenderanteil:vec13	0.0408693	0.0139815	2.923	0.003674 **
Auslaenderanteil:vec36	-0.0546546	0.0170727	-3.201	0.001484 **
Langzeitarbeitslosenquote:vec4	0.2414797	0.0669906	3.605	0.000354 ***
Langzeitarbeitslosenquote:vec21	-0.1748835	0.0841583	-2.078	0.038379 *
Einpendler:vec4	0.0256983	0.0088330	2.909	0.003835 **
Hochqualifizierte:vec6	-0.0475119	0.0148580	-3.198	0.001502 **
Hochqualifizierte:vec20	-0.0413573	0.0147154	-2.810	0.005204 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 361.4588)

Null deviance: 474388 on 399 degrees of freedom
Residual deviance: 137285 on 379 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 3

Figure 2: Quasi-Poisson generalized linear regression model. Corresponds to Table 3 of the reproduced paper. Output of line 471 of *data_analyses.Rmd*.

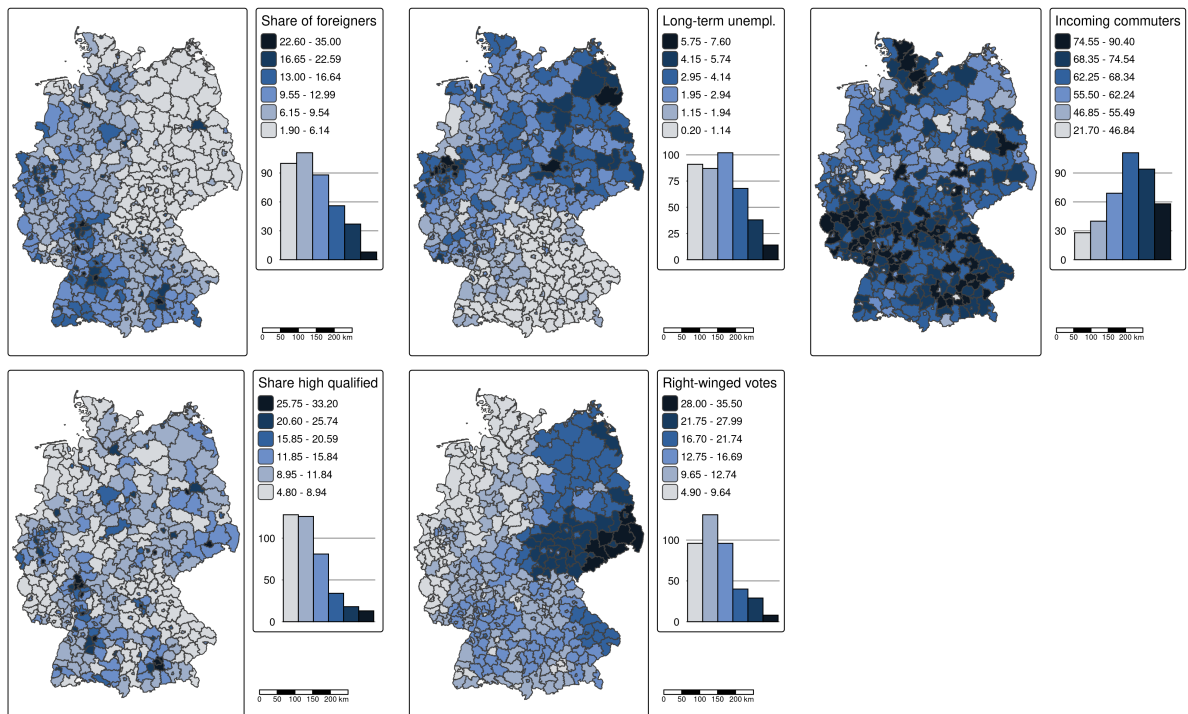


Figure 3: Predictor maps. Corresponds to Figure 1 of the reproduced paper.

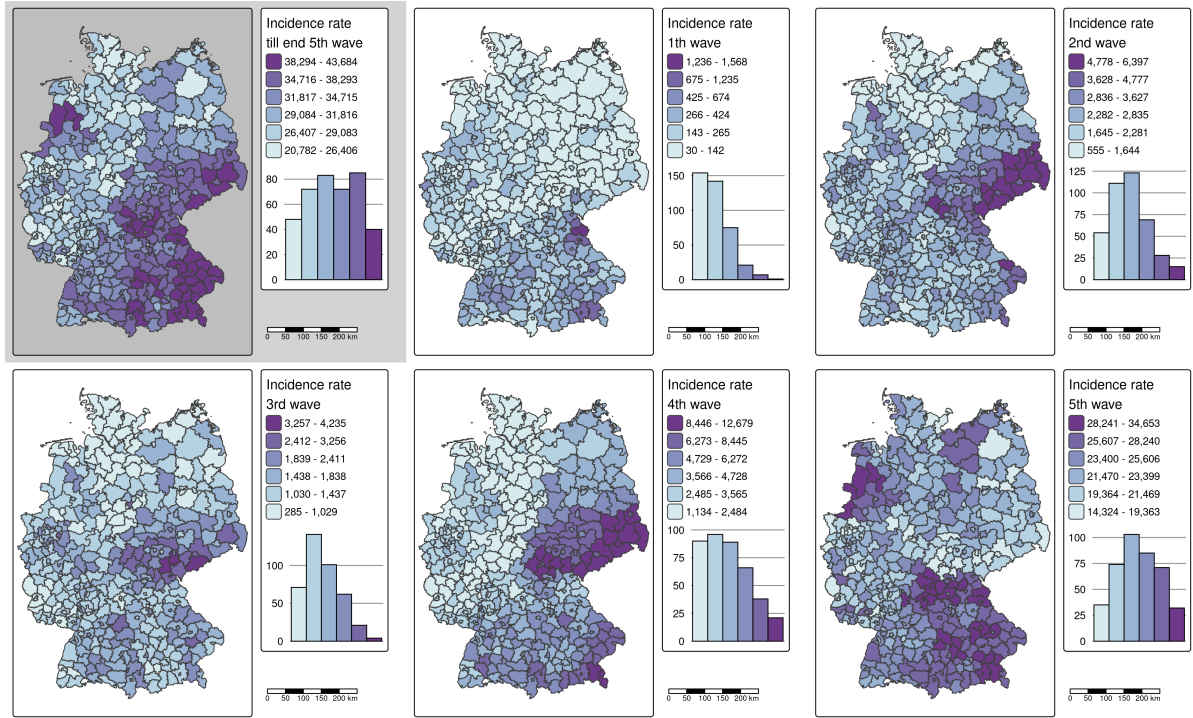


Figure 4: Response maps. Corresponds to Figure 2 of the reproduced paper.

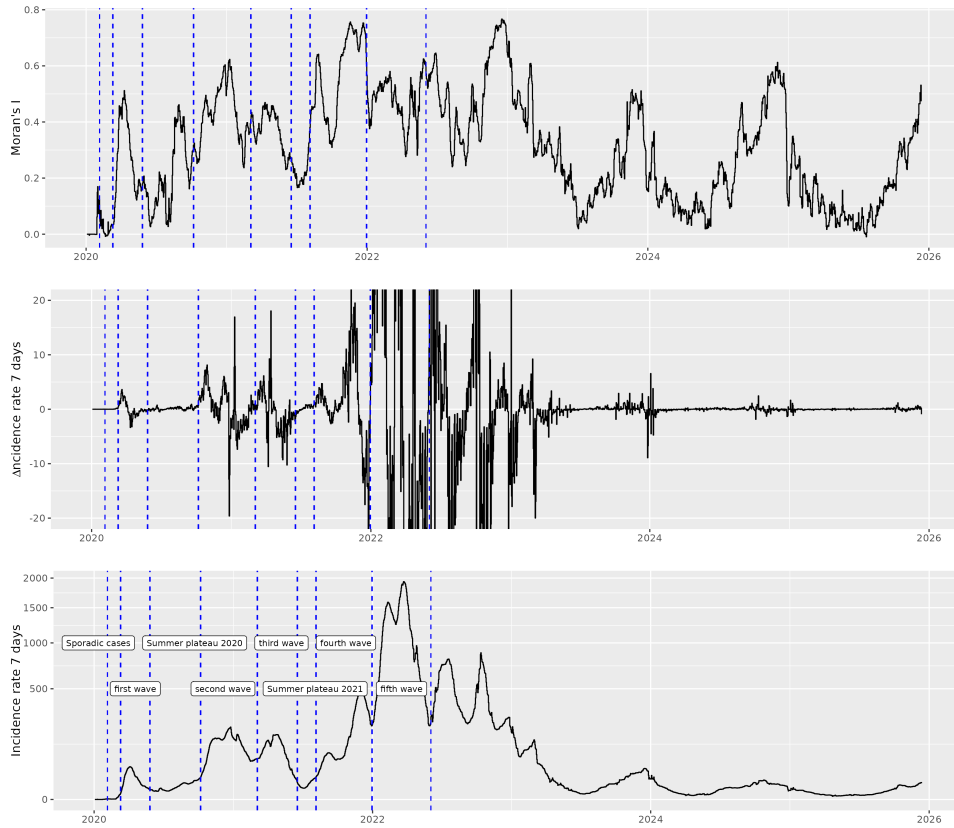


Figure 5: Moran's I time series. Corresponds to Figure 3 of the reproduced paper.

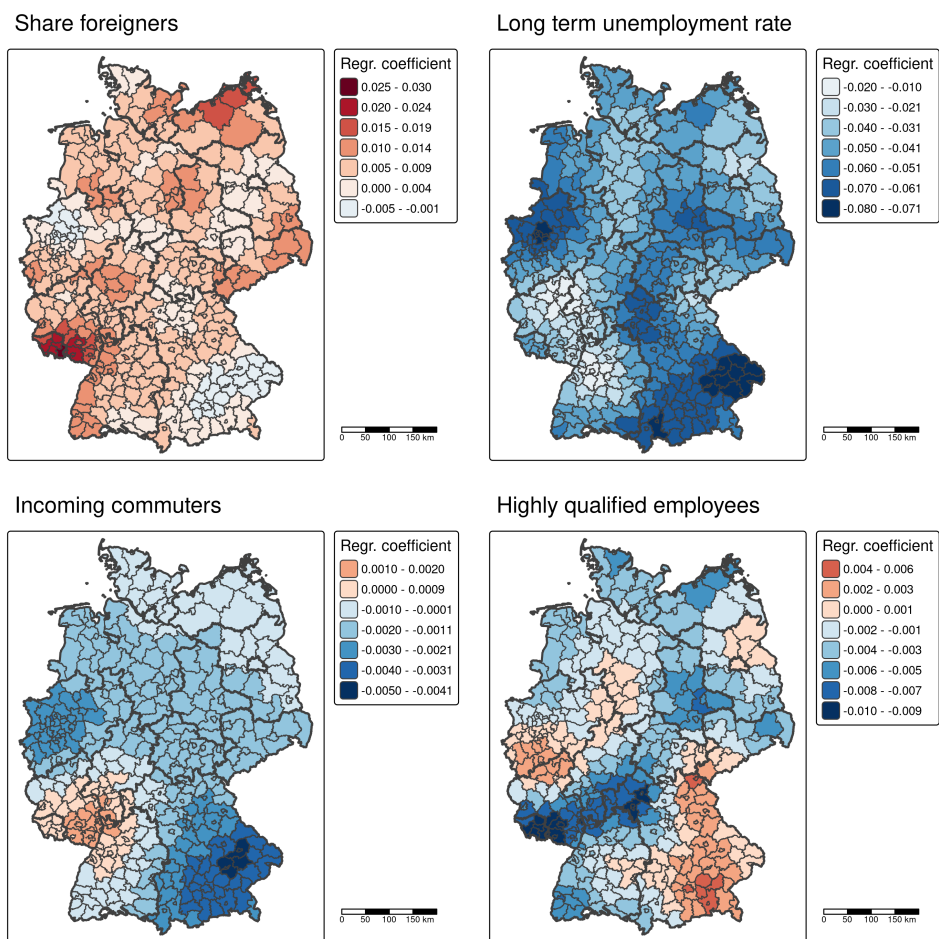


Figure 6: Regression coefficient maps. Corresponds to Figure 4 of the reproduced paper.

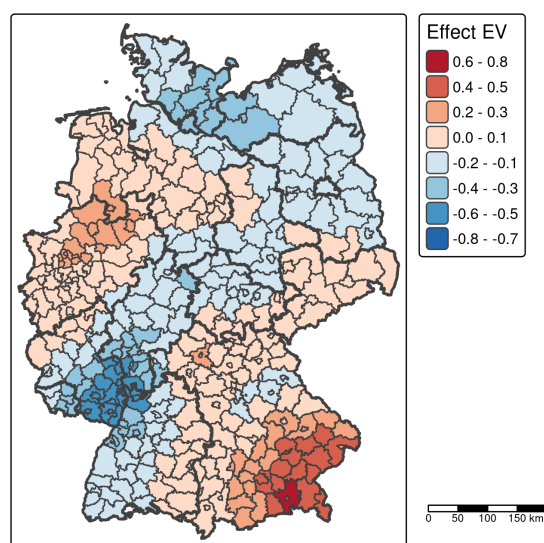


Figure 7: Eigenvector effect map shown at link scale. Corresponds to Figure 5 of the reproduced paper.

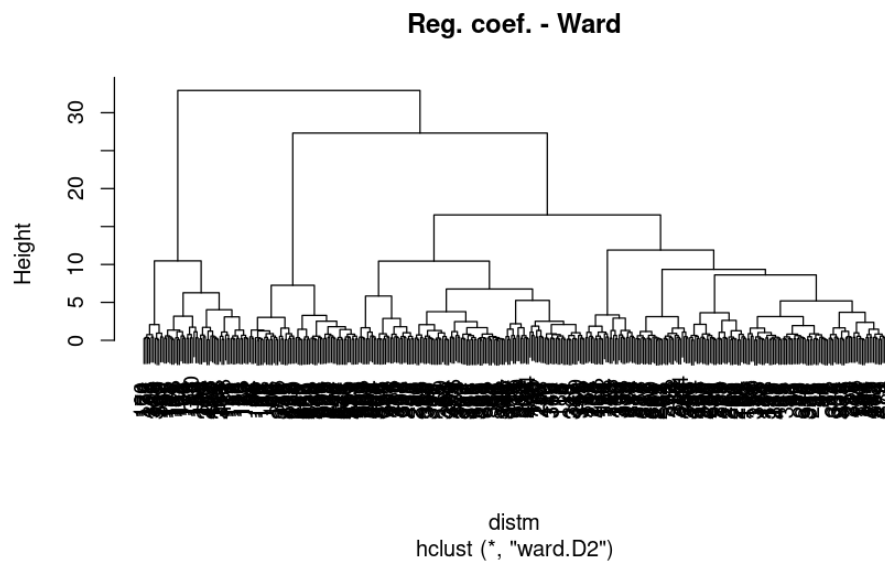


Figure 8: Failed reproduction of the dendrogram. Corresponds to Figure 6 of the reproduced paper.

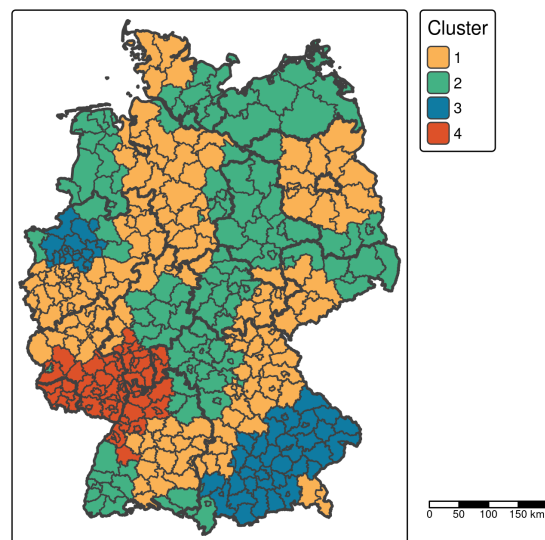


Figure 9: Districts clustered by the spatial varying regression coefficients. Corresponds to Figure 7 of the reproduced paper.