AGILE: GIScience Series, 7, 9, 2026. https://doi.org/10.5194/agile-giss-7-9-2026
Proceedings of the 29th AGILE Conference on Geographic Information Science, 16–19 June 2026.
Eds.: Evelyn Uuemaa, Alexander Kmoch, Holger Virro, and Simon Scheider
This contribution underwent peer review based on a full paper submission.

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

Sven Lautenbach [1,2,3], Marcel Maurer [1,2], and Alexander Zipf [1,2,3]

[1]HeiGIT gGmbH, Heidelberg, Germany
[2]GIScience group, Institute of Geography, Heidelberg University, Heidelberg, Germany
[3]Heidelberg Center for the Environment (HCE), Heidelberg University, Heidelberg, Germany

Correspondence: Sven Lautenbach (sven.lautenbach@heigit.org)

Abstract. The COVID-19 pandemic has caused severe public health issues at a global scale. However, effects were not spatially homogeneous. We analyzed how five selects 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) shaped the incidence rate in Germany over the pandemic phase. The analysis was performed at the level of the 400 districts. In addition, we analyzed how spatial autocorrelation changed during the course of the pandemic. As regression residuals were positively spatially autocorrelated we applied a spatial eigenvector approach to derive unbiased estimates for our regression coefficients. We also tested for interactions of the regression coefficients with the selected spatial eigenvectors. With the exception of the votes for the right-winged populist party all predictors showed significant interactions with the eigenvectors. The resulting spatially varying regression coefficients maps indicate that the relationship between incidence rate and the socio-economic indicators might be less straight-forward than commonly assumed and seemed moderated by additional factors. The spatial clusters that emerged during the analysis provide a base for a more detailed analysis of the interplay of regional scale health politics and regional economic geography.

Submission Type. analysis, case study

BoK Concepts. [GS3] Use of geospatial information; [GC2] Spatial simulation modelling; [TA12-7] EO for health surveillance

Keywords. public health, spatial autocorrelation, spatial eigenvector mapping, spatial regression modelling, COVID-19

1 Introduction

The COVID-19 caused severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has lead to severe public health issues at a global scale (Worobey et al., 2020). After its emergence in late 2019 in Hubei province, China, SARS-CoV-2 spread quickly globally. While all countries were effected by the pandemic, the incidence rate differed spatially both between and inside countries (Du et al., 2024). Moderating factors for the incidence rate were the different public health counter measures (Worobey et al., 2020) - including vaccination and social distancing measures - as well as socio-economic factors that moderated how closely inhabitants supported or were able to support the public health interventions (Mishra et al., 2021; Wachtler et al., 2020; Wildman, 2021). We analyzed the role of several socio-economic factors at the example of Germany.

As the spread of COVID-19 as an infectious disease can be assumed to be positively spatially autocorrelated and as socio-economic moderating factors tend to be spatially structured as well, the analysis of spatial patterns of the incidence rate should incorporate spatial regression approaches (Guliyev, 2020). In Germany a few studies have analyzed the role of socio-economic indicators on COVID-19 incidence at the district level. However, these studies have analyzed so far only the first waves of the pandemic. Ehlert (2021) have used spatial lag and spatial error models for this analysis for the first wave till June 2020. These models assume a (spatially structured) gaussian distributed error term which might not be adequate for the count data nature of the incidence data. Rohleder et al. (2022) have used Bayesian spatial modeling - assuming a Poisson distribution - to analyze the effect of economic deprivation, and proportion of non-nationals on COVID-19 incidence, while controlling for age, sex, vaccination coverage, settlement structure. Their analysis was limited till May 2021.

Incidence rate is clearly linked to vaccination rate - socio-economic factors for vaccination uptake can therefore be assumed to be closely linked to incidence rate. Reis et al. (2024) have studied the effect of socio-economic predictors for COVID-19 vaccination rate in Germany using multilevel logistic regression models till February 2022. Our focus here is on incidence rate, treating vaccination rate as a latent outcome of socio-economic factors without distangling direct and indirect effects.

The aim of our analysis was twofold: first, we wanted to quantify the amount of spatial autocorrelation during the different periods of the pandemic till April 2022 (the end of the pandemic phase in Germany). Second, we wanted to quantify the effect of socio-economic predictors over the whole period, putting a special focus on the role of spatially varying regression coefficients.

2 Background: German health system and public health interventions

In Germany, the public health system is organized in a federal way with shared responsibilities between agencies at the national, federal and district level (Tsalampouni, 2022; Kuhlmann and Franzke, 2022). Health insurance is provided - at least at some base level - for all residents, including refugees, asylum seekers, and illegal immigrants (Tsalampouni, 2022). While the same general framework for health interventions was set at the national level, COVID-19 related public health interventions differed with respect to strictness and timing between the federal states (the *Länder*) (Kuhlmann and Franzke, 2022). The German government enacted initial stringent measures in March 2020, leading to the first nationwide lockdown Deutsche Bundesregierung (2022). These measures included the closure of kindergartens, the postponement of academic semesters, the prohibition of visits to nursing homes, and restrictions on the number of individuals allowed at public and private gatherings.

Table 1. COVID-19 periods in Germany, defined according to the RKI (Schilling et al., 2022).

Phase	Time period
Pre-first wave and first wave	01/27/2020 - 05/17/2020
Summer plateau	05/18/2020 - 09/27/2020
Second wave	09/28/2020 - 02/28/2021
Third wave	31/01/2021 - 06/13/2021
Summer plateau	06/14/2021 - 08/01/2021
Fourth wave	08/02/2021 - 12/26/2021
Fifth wave	12/27/2021 - 05/30/2022

3 Methods and data

3.1 Data

We used the COVID-19 case reported by the local health agencies as the response variable. The data was available aggregated by week and by district from the Robert-Koch-Institut (RKI) [1]. The aggregation by week avoids reporting problems as the local health agencies did underreport cases during weekends which were as a result added to cases reported at the following working days. By definition, reported cases show a temporal lag of several days compared to the day of disease occurrence due to the time required for the antibody test and the processing of the reported cases which was especially slow during high incidence periods. The data was available for the 400 districts in Germany. For Berlin information at a subdistrict level (*Bezirke*) was available but had to be aggregated as the socio-economic indicators were only available at the level of the 400 districts. The RKI dataset included as well the number of inhabitants per district which was used as an offset in the count regression models later on. We used the administrative boundaries for 2021 as geometries for the districts (BKG (german federal agency for cartography and geodesy), 2021). Due to a local government reorganization the administrative units for the socio-economic indicators and the incidence data did not match - the districts of Eisenach and Wartburg have been merged at first of July 2021. We recalculated the indicators for the fused district based on the available information.

For the regression analysis, we focused on the sum of COVID-19 cases from the beginning of the pandemic till the end of the fifth wave (cf. Table 1).

Socio-economic indicators were downloaded from the federal institute for research on building, urban affairs and spatial development (BBSR)[2]. We used the following predictors which were available for 2017 at the level of the 400 districts.

Share of foreigners, normalized by the number of inhabitants: we hypothesized that a larger shares of foreigners indicate a larger share of inhabitants with limited skills in the German language a lower access to health and social distancing related information and that thereby districts with higher shares are associated with higher incidence rates. The effect of the share of foreigners was also tested by Rohleder et al. (2022) for the first three waves of the pandemic in Germany. According to news articles the efforts to spread information in languages other than German - potentially supported by street work activities - differed strongly between the districts. Furthermore, the use of social media and classical media might be limited to specific channels which could be prune to misinformation - e.g. widespread misinformation in

[1] https://github.com/robert-koch-institut/COVID-19_7-Tage-Inzidenz_in_Deutschlandatthe12thDecember2025

[2] https://www.inkar.de/

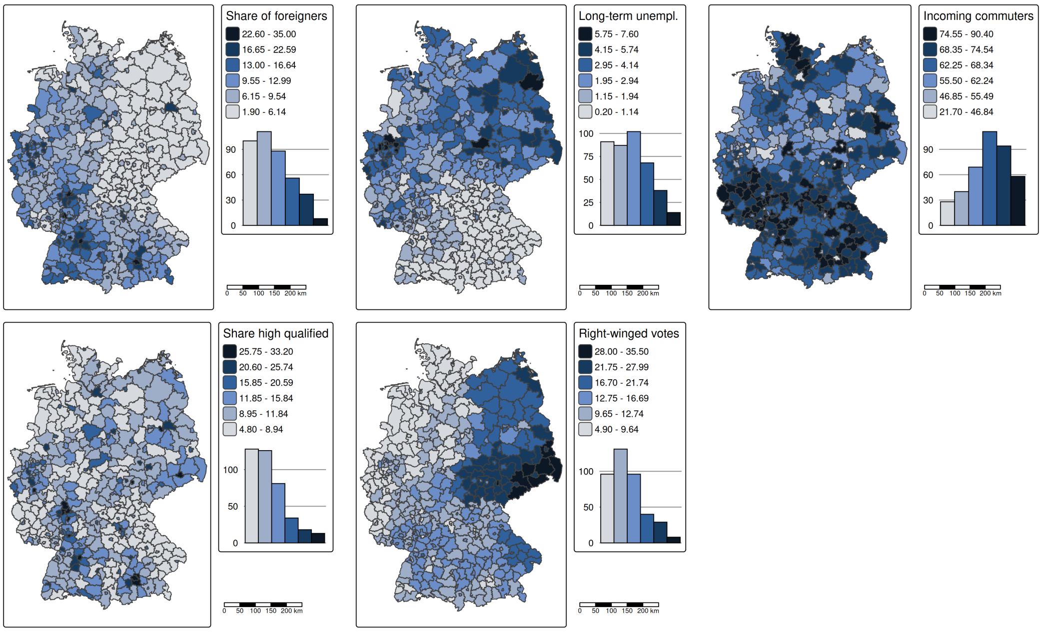

Figure 1. Maps of the predictors used in the regression model: share of foreigners, long-term unemployment rate, share of incoming commuters, share of highly qualified employees (Bachelor or Master degree) and share of votes for the right-winged populist party at the national election 2017.

Russian channels with respect to vaccination campaigns (Kotseva et al., 2023). Clearly, non-national inhabitants form a diverse group ranging from high-income groups – such as IT-specialists – to blue collar and uneducated workers to refugees clustered in refugee shelters. We hypothesized that both type of non-national inhabitants and communication strategies towards that groups tend to cluster geographically, leading to the assumption that the strength and direction of the regression coefficient might spatially vary. The spatial distribution of the share of foreigners shows a clear east-west divide with the lowest share in Eastern Germany (with the exception of Berlin) and the eastern parts of Bavaria (cf. Fig. 1). Higher shares were present in the larger cities, around the Ruhrgebiet, Munich, the Rhein-Main and Rhein-Neckar region and Stuttgart.

Incoming commuters, normalized by the total number of employees subject to social insurance contributions in the district: The values include commuters resident in neighboring countries that work in Germany. We hypothesized for both incoming commuters indicate a spill-over of population between districts and thereby were associated with higher incidence rates. The effect of commuters was also tested by Ehlert (2021) for the first wave of the pandemic in Germany. The share of incoming commuters was low in the bigger cities and high in areas in the neighborhood of these cities (cf. Fig. 1).

Share of highly qualified employees (employees with Bachelor or Master degree), normalized by total number of employees: The hypothesis was that employees with an academic degree more frequently work at home office and districts with a higher share were thereby associated with lower incidence rates. A similar hypothesis was tested by Ehlert (2021) for the first wave of the pandemic in Germany. The share of highly qualified employees also involves another aspect related to income as academics tend to have higher incomes and tend to be less bothered by unemployment. The distribution of the highly qualified employees showed highest values around larger cities and surrounding districts (cf. Fig. 1). Relative high shares of highly qualified employees were present in the Rhein-Main, Rhein-Neckar regions, around Stuttgart and Munich.

Long term unemployment rate: We hypothesizes that long term unemployment rate would be associated with higher incidence rates as it is an indicator for economic conditions in general. Poorer districts might be facing higher incidence rates due to denser living conditions, less access to information and resources. Such a relationship had been observed by Rohleder et al. (2022) for the second and third wave in Germany. The spatial pattern of the long-term unemployment rate shows a north-south divide the lowest rates in the south (Baden-Württemberg and Bavaria) and highest rates in eastern Germany, the Ruhrgebiet area and city districts in the western part of Palatina and in the Saarland (cf. Fig. 1).

Share of right-winged voters (AfD) at the national election in 2017: The hypothesis was that the share of right-winged voters is a proxy for the share of the population skeptical with respect to social distancing, mask wearing and vaccination and might therefor be associated with higher incidence rates at the district level. In contrast to other European countries that were hit harder by the first COVID-19 wave - such as Italy or Portugal - populists in

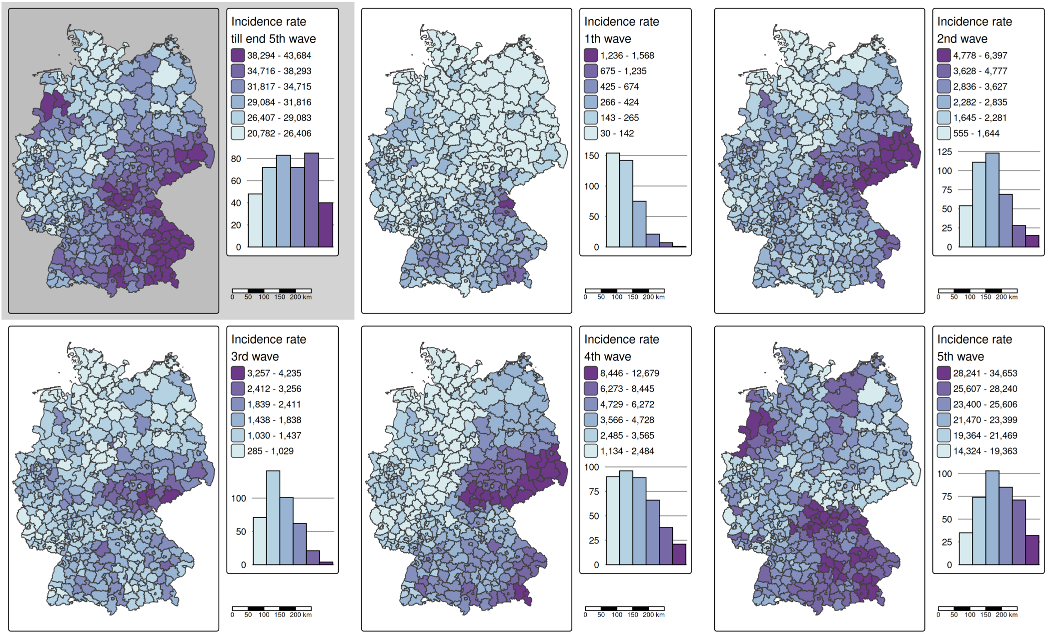

Figure 2. Maps of the COVID-19 incidence rate (cases by 100,000 inhabitants) at the level of the 400 districts in Germany. For the regression analysis only the aggregated incidence rate till the end of the fifth wave was used. Please note that the range of values and the classification differs between the six maps.

Germany argued strongly critical with respect to social distancing and COVID-19 vaccination (Lewandowsky et al., 2022). A similar was tested by Qamar et al. (2022) for the first waves and in a non-spatial model. Several surveys indicated a strong link between attractiveness of the AfD and radical opposition of COVID-19 control measurements such as vaccination and social distancing (Truffer et al., 2024; Reuband, 2022). We decided to use the last national election before the pandemic as the predictor as the use of election data at the federal state level would have led to hardly comparable effects as these elections are temporally not aligned. The votes for the populist right-winged party at the national election in 2017 showed a clear east-west divide with higher share of votes in the eastern part, especially in Thuringia and Saxony (cf. Fig. 1). The eastern part of Bavaria also showed higher votes for the populist party.

The COVID-19 incidence rate for the period till the end of the fifth wave (cf. Fig. 2) showed hotspots in the south-eastern part of Germany (mainly Bavaria, Thuringia and Saxony) as well as in the small cluster on the north-west (Emsland, Cloppenburg, Vechta). Cold spots were present especially in the northern part (Schleswig-Holstein). The location of the hotspots differed between the five waves of the pandemic: the first wave was characterized by two small clusters in east and south Bavaria. The second and the third wave of the pandemic were characterized by hotspots in Saxony and Thuringia with some local hotspots in eastern Bavaria. The hotspot in Thuringia and Saxony prevailed during the fourth wave of the pandemic, now in combination with higher incidence rates in south-east Bavaria (especially high in the districts Traunstein and Berchtesgardener Land). The fifth wave was characterized by hotspots in Bavaria and parts of Baden-Württemberg and an additional hotspot in the north-west (around Osnabrück and Münster).

3.2 Methods

We first scanned the bivariate relationships between the predictors by Pearson's r to identify problem with collinearity (Dormann et al., 2013). As this indicated only low correlations we continued with all predictors and checked the variance inflation factor (Fox and Monette, 1992) for the regression coefficient estimates.

We defined the spatial weight matrix based on a union of the queen contiguity neighborhood definition and a 70km distance band neighborhood definition using the point on surface of the districts. The spatial weights were down-weighted by an inverse distance approach with a power of 1 and the resulting matrix was finally row-standardized.

We calculated global Moran's I (Moran, 1950) for each week using the z-score of the empirical Bayes modification (Assunção and Reis, 1999) of the number of cases by week as the variable of interest. The empirical Bayes modification applies a Bayesian smoothing to the variable of interest to avoid problems with high crude rates due to low populations. Given the high number of inhabitants per district which step would not have been necessary. Moran's I was calculated for the aggregated incidence rate for each of the periods in Table 1 as well as for each day of the available time series.

For the regression analysis we used a quasi-Poisson generalized linear model with a log link. To account for spatial autocorrelation, we used spatial eigenvector mapping (Griffith et al., 2019). Thereby, the spatial weight matrix is separated into orthogonal spatial eigenvectors which are then tested in a brute-force approach if they improve the likelihood of the existing regression model and decrease global spatial autocorrelation of the residuals (Griffith et al., 2019). To account for the presence of spatial varying regression coefficients, we tested interactions between the selected eigenvectors and the predictors share of foreigners, long-term unemployment rate, share of incoming commuters and share of votes for the right winged party following the approach described in Griffith et al. (2019). Model selection was based on a F-test and followed the selection procedure described in Zuur et al. (2009).

To ease interpretation of the spatially varying regression coefficients we clustered the districts based on their regression coefficients. The regression coefficients were standardized and the euclidean distance was used as the distance measure. We used hierarchical clustering with Ward's minimum variance approach as the criteria, using the original definition by Ward (Murtagh and Legendre, 2014). Our aim was thereby to identify regions with similar functional relationships between the response and predictors.

Finally, we applied the regression approach described above to the individual pandemic phases (five waves and two summer plateaus). We used the sum of the cases for each pandemic phase as the response variable. Due to data availability we used the same predictors for each phase - assuming that these will have change only marginally across the time frame considered.

The analysis was done based on R (R Core Team, 2024) using the packages spatialreg (Pebesma and Bivand, 2023), spdep (Roger Bivand, 2022), tidyverse (Wickham et al., 2019), ggplot2 (Wickham, 2016), tmap (Tennekes, 2018), lubridate (Grolemund and Wickham, 2011), vegan (Oksanen et al., 2025), cluster (Maechler et al., 2023) and sf (Pebesma, 2018).

3.3 Data and Software Availability Section

The data sets are available at the location of the data providers. A pre-processed version of the data is - together with the code for the statistical analysis, the pre-processing and the figure preparation – available at https://github.com/slautenb/covid19germany.

4 Results and Discussion

After the onset of the first COVID-19 wave the incidence rate was clustered for most of the time as indicated by a positive Moran's I (cf. Figure 3). Moran's I was always positive and dominantly significant higher than the expected value (with the exception of 86 days at which Moran's I was not significant at the 5% level). As to be expected for an infectious disease, the spatial clustering was highest during periods of increasing incidence rate. During periods of decreasing incidence rate the spatial clustering was reduced.

For the aggregate cases of the different periods according to Schilling et al. (2022)), the spatial autocorrelation was highest (I=0.75) for the fourth wave which was characterized by a strong increase over most of the period (cf. Table 2). The other waves which were characterized by both increasing and decreasing phases should a lower but still strong spatial clustering (I around 0.5). The incidence rates for the two summer plateaus were even lower clustered (I around 0.36).

Table 2. Global Moran's I for the aggregated incidence rate for each pandemic period as defined by Schilling et al. (2022). The statistic was calculated using the empirical Bayes modification. The p-value was based on a permutation test with 9999 replicates.

Period	Moran's I	p-value
Sporadic cases	0.03	0.0127
First wave	0.51	0.0001
Summer plateau 2020	0.37	0.0001
Second wave	0.59	0.0001
Third wave	0.51	0.0001
Summer plateau 2021	0.36	0.0001
Fourth wave	0.75	0.0001
Fifth wave	0.54	0.0001

Pearson's r identified some moderate associations between the predictors at the district level: The long-term unemployment rate was negatively associated with the share of incoming commuters (r=-0.57). The share of employees with an academic degree was positively associated with the share of foreigners (r=0.53). The share of votes for the right-winged party was negatively associated with the share of foreigners (r=-0.4). The variance inflation factor for regression coefficients for the selected regression models without spatial interactions was too low (maximum VIF of 2.4) to be considered problematic for the analysis. Therefore, we continued using the model with all predictors for the selection of spatial eigenvectors and the testing of interactions between the predictors and the spatial eigenvectors.

The residuals of the non-spatial regression model showed positive spatial autocorrelation (Moran's I= 0.36, p-value $< 2.2 \cdot 10^{-16}$) which was strongly reduced by the inclusion of the spatial eigenvectors (Moran's I= 0.11, p-value= $1.6 \cdot 10^{-7}$). Four of the five predictors showed significant interactions with some of the selected eigenvectors (cf. Table 3).

The regression model (cf. Table 3) identified as expected a significant positive effect of the share of votes for the right winged party and the incidence rate. This result is

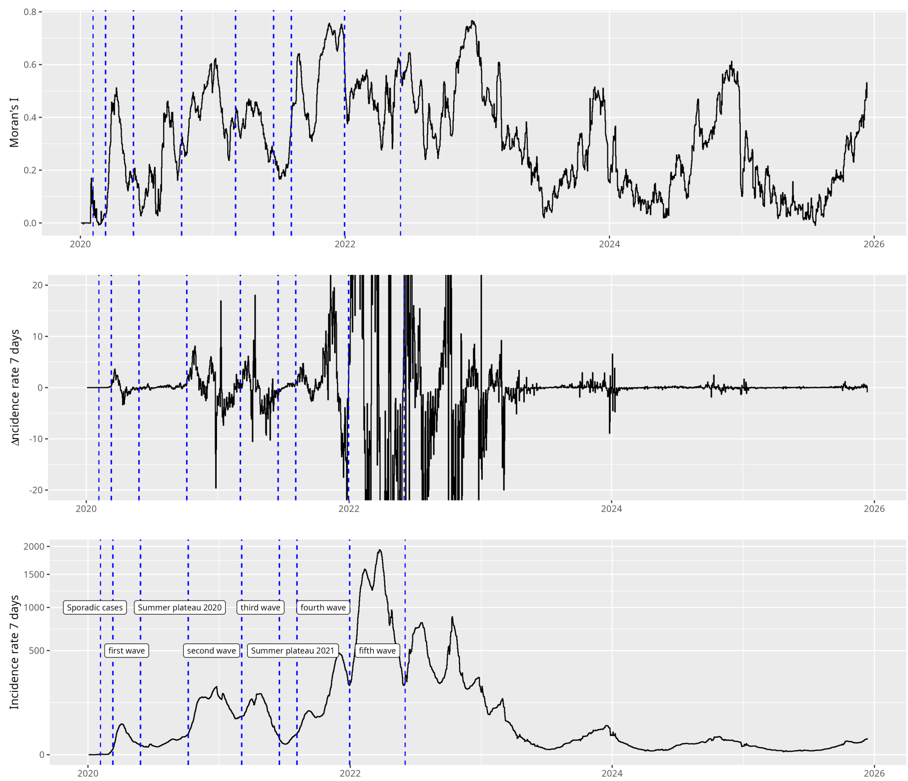

Figure 3. Time series of the COVID-19 incidence rate in Germany, the first order difference of the incidence rate and the global spatial autocorrelation (Moran's I) for the corresponding date. The dashed lines indicate the different periods of the pandemic in Germany. Please not that the y-axis in the lower panel has been square-root transformed to ease the comparison for low incidence cases. Also the middle panel was zoomed in to improve visibility for days with low incidence rates. Data after the end of the pandemic phase (around December 2023) presumably underestimate the incidence rate strongly due to the dominantly asymptomatic infections. After the end of the fifth wave, the RKI did not continue to delineate and to name periods.

in line with the outcome of Qamar et al. (2022) for the first waves of the pandemic in Germany. The regression coefficient did not vary spatially. Districts that had voted stronger for the populist right-winged party AfD showed higher incidence rates for the period till the end of the fifth wave. As the AfD and associated social groups were opposing social distancing and vaccination and organized frequent demonstrations were social distancing rules were violated, the higher incidence rates in this districts can presumably be explained by a higher share of unvaccinated inhabitants and an increased contact rate between - especially unvaccinated - inhabitants of those districts.

The association between incidence rates and the share of highly qualified employees was mixed due to the interaction with two spatial eigenvectors. The hypothesized negative association was present in about 70% of the districts with an especially strong association in an area stretching from the Saarland and western Palatinate till Würzburg and Schweinfurt, including the Rhein-Neckar region. Areas with an opposite direction of the effect showed up in four clusters: i) a region stretching from Koblenz over Bonn and Cologne till the beginning of the Ruhrgebiet, ii) a region between Kassel and Hannover, iii) Berlin and four districts north of it, and iv) a region covering most of Bavaria (with hotspots around Hof, Rosenheim and Munich) and parts of Baden-Württemberg (including Stuttgart). These results imply that the hypothesized relationship of higher qualification, higher share of home office workers and therefor lower incidence rate did not apply for the whole of Germany. It might be necessary to distinguish between the type of employer (industry vs. service) and maybe also the size of the company or organization in addition to the academic degree of the employee.

Unexpectedly, the long-term unemployment rate was always negatively associated with the COVID-19 incidence rate at the district level. The effect showed significant interactions with two spatial eigenvectors (cf. Fig. 4). The strongest negative associations were indicated in Bavaria and at the Ruhrgebiet (western Germany). For Bavaria, the strongest negative association was identified for a cluster around Passau and around Oberallgäu, a secondary cluster showed up around Würzburg and Schweinfurt. The Ruhrgebiet area characterized by a relatively high variability of the long-term unemployment rate across the districts (cf. Fig. 1). The estimated effect of the long-term unemployment rate was lowest in two

Table 3. Quasi-Poisson generalized linear regression model for the number of COVID-19 cases. The number of inhabitants was used to model the incidence rate. All coefficients are reported at the link-scale. The log-link function was used. The selected spatial eigenvectors are abbreviated as EV. Interactions between predictors and spatial eigenvectors are indicated by a ':'. The explained deviance includes the effect of the selected spatial eigenvectors.

	Coefficient	Standard error	p-value
Constant	-1.21	0.035	$< 2 \cdot 10{-}16$
Share votes right-winged	0.015	0.0009	$< 2 \cdot 10{-}16$
Share high qualified employees	-0.002	0.001	0.063
Long-term unemployment rate	-0.046	0.0035	$< 2 \cdot 10{-}16$
Incoming commuters	-0.0013	0.00038	0.00073
Share foreigners	0.0068	0.0011	$1.97 \cdot 10^{-9}$
EV4	-3.99	0.78	$5.43 \cdot 10^{-7}$
EV6	-0.82	0.20	$7.52 \cdot 10^{-5}$
EV13	-0.023	0.16	0.89
EV16	-0.35	0.08	$1.41 \cdot 10^{-5}$
EV20	0.91	0.2	$7.68 \cdot 10^{-6}$
EV21	0.83	0.17	$2.61 \cdot 10^{-6}$
EV36	0.3	0.2	0.14
Share foreigners:EV4	0.072	0.017	$2.52 \cdot 10^{-5}$
Share foreigners:EV13	-0.04	0.014	0.0037
Share foreigners:EV36	-0.055	0.017	0.015
Long-term unemployment rate:EV4	0.24	0.067	0.00035
Long-term unemployment rate :EV21	-0.17	0.084	0.038
Incoming commuters:EV4	0.026	0.0088	0.0038
Share high qualified employees:EV6	-0.048	0.015	0.0015
Share high qualified employees:EV20	-0.041	0.015	0.0052
Explained deviance		0.71	
Dispersion parameter		361.5	

clusters: districts around Pforzheim (southern cluster) and between Wiesbaden and Koblenz (northern cluster).

The effect of incoming commuters on the COVID-19 incidence rate was negative for most of Germany (cf. Fig. 4) with one exception: the estimates regression coefficients were positive (as hypothesized) for a cluster consisting of the Rhein-Main and the Rhein-Neckar region (including e.g. Frankfurt a.M., Mainz, Mannheim and Heidelberg). This pattern could relate to the high number of commuters in the two regions which also travel relative large distances. The Rhein-Main and the Rhein-Neckar region host large industrial companies (such as the BASF, Höchst, Merck) with high shares of blue collar workers for which home office was mostly no option.

It should be taken into account, that the data reflect the commuting behavior before the pandemic. Due to the increasing possibilities of home office work after the onset of the pandemic the commuting pattern has changed during the pandemic (Ecke et al., 2022): public transport was used somewhat less frequently and economically better of employees tended to commute less. However, the general commuting patterns were unchanged (Ecke et al., 2022). As the analysis aimed for an understanding of socio-economic factor at the baseline situation we decided to use the values for 2017 as the baseline and not to use any updated data.

Presumably, interpretation of the effect of incoming commuters has to be interpreted together with the effect of the share of highly qualified employees. Areas with positive regression coefficients for highly qualified employees tend to tend to have negative coefficients for the incoming commuters. For districts with high shares of of highly qualified incoming commuters a large share of home office workers might have contributed to lower incidence rate while for districts where incoming commuters had less opportunities for home office work commuter might have contributed to an increase in incidence rates. This requires further testing for interactions between the predictors, potentially under involvement of the share of workers in the industry.

The effect of the share of foreigners on the COVID-19 incidence rate was negative for all areas with the exception of a region in south east Bavaria (including Munich, Straubing and Passau) and the Ruhrgebiet till Münster (cf. Fig. 4). The highest positive effect was in the western part of Rhineland-Palatinate and the Saarland. This region spread into the western parts of Baden-Württemberg. These clusters of regression coefficients might relate to different efforts of the local administration to spread information to non-native inhabitants and especially those with limited German speaking capabilities. For the Ruhrgebiet with both high share of non-native inhabitants and a high variability differences in the share of foreigners

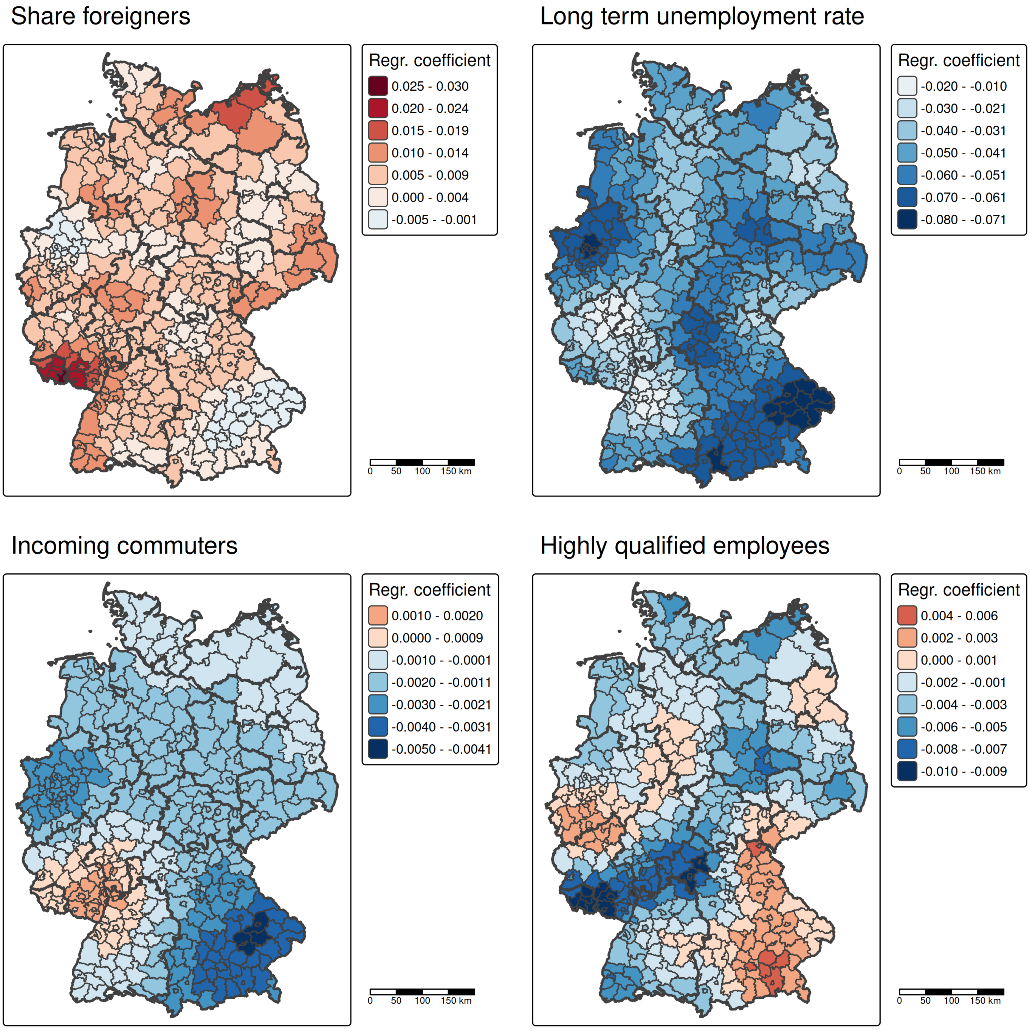

Figure 4. Spatial varying regression coefficients. All maps are at the link scale. The regression coefficients show the main effect plus the effect of the significant interactions with the spatial eigenvectors. The thicker lines indicate the administrative boundaries of the federal states

did not matter with respect to the COVID-19 incidence rate. In contrast for the cluster in western Palatinate and the Saarland the differences in the relatively low share of foreigners matters with respect to the incidence rate.

In addition to the interaction between the spatial eigenvectors and the predictors the model included as well the main effects of the selected eigenvectors. For interpretation, these were multiplied with the regression coefficient and summed up to indicate the spatial pattern (cf. Fig. 5). The model included an effect that increased incidence rates in south-east Bavaria (peak in district of Rosenheim), around Schweinfurt, and in the region between Osnabürck, Münster and the Ruhrgebiet. The effect of the eigenvectors also included a decrease of the predicted incidence rates in a cluster around the Rhein-Neckar region (peak at Ludwigshafen am Rhein and Frankenthal (Pfalz)) which extended till Saarbrücken, Pforzheim and the Frankfurt am Main.

The Ward clustering based on the spatially varying regression coefficients lead to four clusters based on visual inspection of the dendogram (cf. Fig. 6). Based on the dendogram cluster 1 and 2 were more similar to each other. The districts were clustered into two smaller (cluster 3 and 4) and two larger groups (cf. Fig. 7).

Cluster 4 represents the Rhein-Main and Rhein-Neckar region together with the western part of Palatinate and the Saarland (Cluster 4). The cluster was – in contrast to the rest of the country – characterized by a strong negative effect of the share of highly qualified employees on the incidence rate, a positive effect of incoming commuters, a positive effect of the share of foreigners and a weak negative effect of the long-term unemployment rate. The districts in this cluster were in line with the hypothesized effects for the share of foreigners, incoming commuters and the share of highly qualified employees.

Cluster 3 included the northern part of the Ruhrgebiet and the south-eastern part of Bavaria. Districts in this

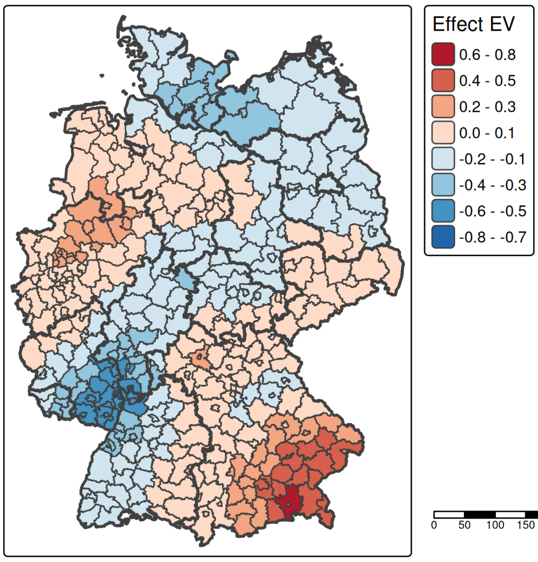

Figure 5. Map of the sum of the main effects for the spatial eigenvectors multiplied with the respective spatial eigenvector. The effect is shown at the link scale. A positive effect indicates that the model increased the prediction of the incidence rate for this district and a negative effect indicates that the incidence rate was reduced for this district. The thicker lines indicate the administrative boundaries of the federal states.

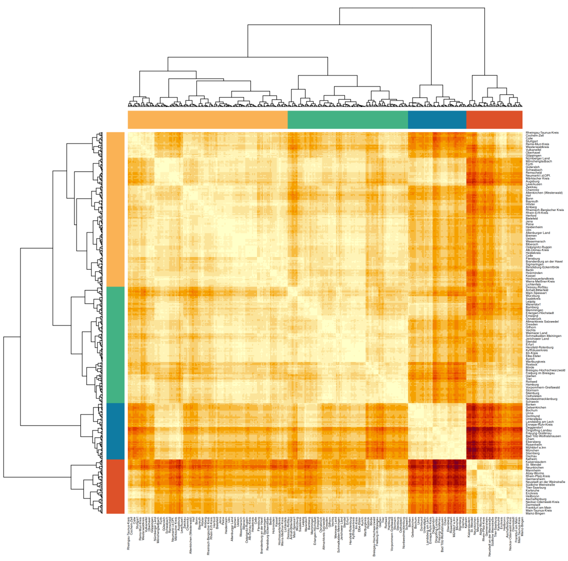

Figure 6. Dendogram of districts based on the euclidian distance of the spatially varying regression coefficients. The clusters are colored according to the colors used in Fig. 7. The rows and columns represent the districts and the colors represent the distance between the districts

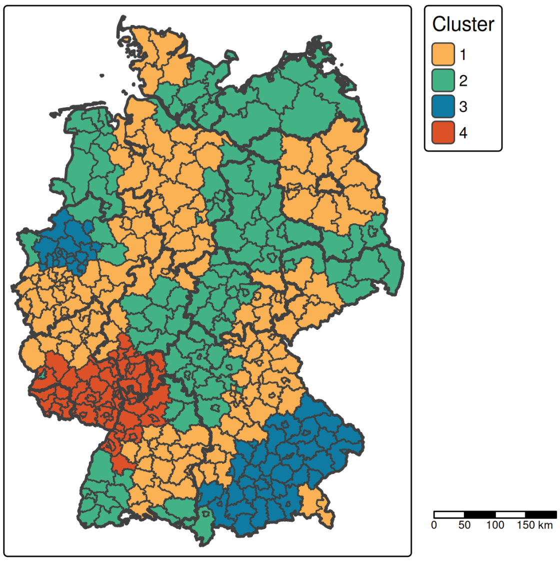

Figure 7. Districts clustered by the spatial varying regression coefficients. The thicker lines indicate the administrative boundaries of the federal states.

cluster were – in contrast to the rest of the country – characterized by a negative effect of the share of foreigners on the incidence rate, strong negative effects of the share of incoming commuters and long-term unemployment rate. The districts of this cluster in Bavaria were also characterized by a positive effect of the share of highly qualified employees on the incidence rate. It is interesting that – beside the effect of the share of highly qualified employees – these two groups showed a similar behavior as the two regions differ strongly e.g. with respect to the regional economy.

Cluster 1 and 2 described larger, spatially not connected regions. Cluster 2 included a band from the south-west to the north east (with interruptions by Berlin-Brandenburg and the Rhein-Main and Rhein-Neckar regions as well as the districts bordering the Ruhrgebiet and the low countries. Cluster 1 and 2 were characterized by relatively similar regression coefficient values. The main difference was present for the regression coefficients for the share of the highly qualified employees: cluster 2 was characterized by a negative relationship between this predictor and the incidence rate while cluster 1 was characterized by a positive relationship.

The four cluster did not differ remarkably with respect to the incidence rate or the predictor values. The districts were assigned to larger spatially connected clusters with only a few exceptions: i) Trier has been assigned to cluster 3 but was surrounded by districts assigned to cluster 4, ii) Traunstein and Berchtesgardener Land were assigned to cluster 1 but were cut of by the districts of cluster 3, iii) Konstanz (cluster 1) separated the neighboring districts of cluster 2 into two spatial units.

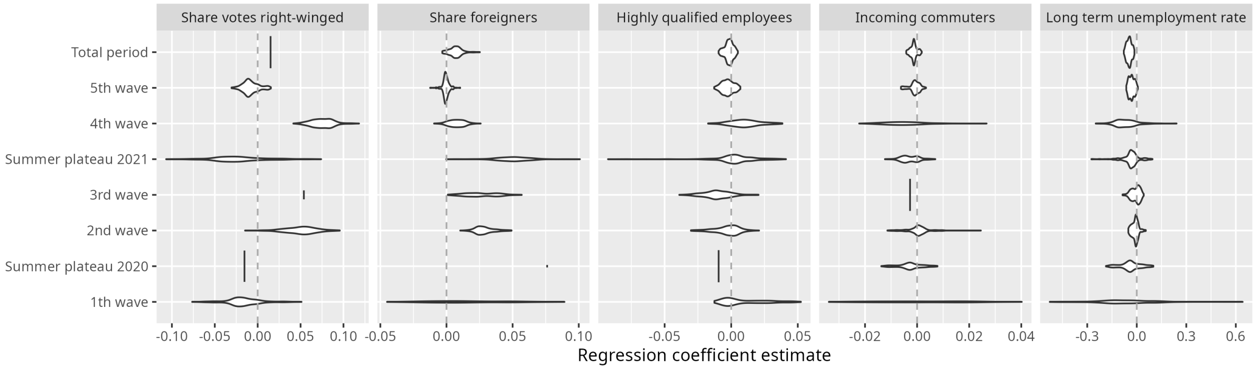

Figure 8. Violin plot of the regression coefficient estimates for the different phases of the pandemic in Germany and for the total period. For some phases some regression coefficients did not interact significantly with the spatial eigenvectors - this is indicated by a vertical line. The dashed line indicates zero.

Regression coefficient estimates varied strongly between the individual phases of the pandemic in Germany (cf. Fig. 8). The first wave was characterized by widely varying regression coefficients, presumably reflecting the diversity in which political measures were implemented across federal states and districts and the effect of singular events such as outbreaks during carnival. The summer plateau 2020 showed no spatial variation for three predictors, while the same predictors varied strongly during the summer plateau 2021. During the fifth wave variance of the predictors was relatively low compared to the other periods. An interesting pattern was found for the effect of the votes for the populist right-winged party: the positive association with the incidence rate was found for the majority of districts for the second, third and fourth wave, while the association varied between negative and positive for the summer plateau o 2021 and the first and fifth wave. This is in line with the strength of the anti-social-distancing demonstrations which gained momentum over time. For the fifth wave, the specifics of the dominant omicron variant together with the effects on behavior on a larger part of the population – due to less severe health effects of infections by the omicron variant – might have blurred the effects of lower vaccination rates and less strict compliance with social distancing and mask wearing. Changes in direction and strength of the relationship between predictors and the incidence rate have also been reported by Rohleder et al. (2022). We further observed changes in the spatial pattern of relationships between response and predictors across the different phases of the pandemic. Both selected eigenvectors and their interactions with the predictors changed between the phases - requiring an in-depth analysis beyond the scope of the current analysis.

The results indicate that the relationship between socio-economic predictors and incidence rate was less straight forward than initially assumed. This provides new insights as the majority of studies in Germany have not considered spatially varying regression coefficients. Studies on the relationships between socio-economic indicators and COVID-19 incidence rates in other countries which tested for spatially varying regression coefficients have also indicated the usefulness of such a spatial approach (Liu et al., 2021; Mollalo et al., 2020; Chen et al., 2022).

The strength and also the direction of the effect was moderated in space as captured by the spatial eigenvectors. We do not know which real world features the spatial eigenvectors represent. They could represent effects the sub-national politics - the federal system of Germany explicitly incorporates the possibility of modifications of national health policies at the level of the federal states and the districts. The effects of the spatial eigenvectors (cf. Fig. 4 and 5) did not align well with the borders of the federal states. Therefore, the results potentially support the effectiveness of political measures at the district level. As the effects tended to cluster together and showed longer ranging patterns this might as well indicate patterns of economically and or socially interconnected regions with similar behavioral pattern. This needs further investigation in follow-up studies. Of special interest for further investigations are the spatial outliers (Trier, Traunstein and Berchtesgardener Land) and the cluster 3 and 4. Understanding in which respect they differed might reveal important lessons for public health interventions in the future.

It is important to keep the scale of the analysis and the modifiable area unit problem in mind for the interpretation. The current analysis does not allow to draw conclusion at the individual level, only at the level of the 400 districts. Any variation across the individual districts had to be ignored due to data availability issues. This might be especially problematic for the large metropolitan areas such as Berlin or Hamburg. For Berlin COVID-19 case data was available at a finer spatial scale - however, it was not possible to incorporate this information as the socio-economic indicators were only available at the district level.

A further shortcoming of the current analysis is the missing incorporation of information from neighboring countries. Integration of this information would complicated due to differences in the reporting of COVID-19 cases and differences in availability of the socio-economic indicators. Especially the incorporation of votes for populist parties would be complicated as the perception of COVID-19 measurements such as social distancing and vaccination differed between the different countries.

Future analysis will incorporate the effect of additional covariates such as population density, share of employees in industry or the rurality of the districts.

5 Conclusion

There seems to be no simple answer to drivers of the COVID-19 incidence rate during the pandemic phase for Germany. Neighboring districts responded similarly during periods of increasing incidence rate and seem to have responded differently during periods of decreasing incidence rates as indicated by the Moran's I time series.

The effect of four predictors was spatially moderated while the share of votes for the right winged party did not vary spatially. Sticking to the rules of social distancing and getting vaccinated - which is how low votes for the populist party can be interpreted - generally reduced the incidence rate.

For the other four predictors in the final model the effect varied spatially, for three predictors this included even a change in sign. This indicates that the socio-economic variables which have been used in regression models might have been to broad and that their effect was moderated by local factors which presumably include local politics. The moderating role of local politics needs to be further investigated to be allow a learning across districts which might help to prepare for the next pandemic.

Zooming into the individual phases of the pandemic revealed different pattern that require additional investigation. First results indicate that strength and direction of effects changed between the different phase of the pandemic. Spatial pattern have of relationships have thereby been changed, potentially due to changes in politics at different levels together with changes in behavior and changes in virus characteristics - e.g. the differences between the delta and the omicron variant of the SARS-CoV-2 virus.

Declaration of Generative AI in writing

The authors declare that they have not used Generative AI tools in the preparation of this manuscript.

Acknowledgements

The authors acknowledge support by the Klaus-Tschira Stiftung. We would also like the encouraging feedback by two anonymous reviewers and the reproducibility reviewer. The analysis presented builds partly on material that has been compiled for lectures and summer schools, e.g. https://giscience.github.io/global-health-academy/.

References

Assunção, R. M. and Reis, E. A.: A new proposal to adjust Moran's I for population density, Statistics in Medicine, 18, 2147–2162, 1999.

BKG (german federal agency for cartography and geodesy): Administrative boundaries of Germany VG250, Geobasis-DE / BKG 2021 licese: dl-de/by-2-0, http://www.bkg.bund.de, accessed: 2021-08-09, 2021.

Chen, M., Chen, Y., Xu, Y., An, Q., and Min, W.: Population flow based spatial-temporal eigenvector filtering modeling for exploring effects of health risk factors on COVID-19, Sustainable Cities and Society, 87, 104 256, https://doi.org/10.1016/j.scs.2022.104256, 2022.

Deutsche Bundesregierung: Coronavirus: Besprechung Der Bundeskanzlerin mit den Regierungscheffinnen und Regierungschefs der Länder vom 22.03.2022, https://www.Bundesregierung.de/Breg-De/Themen/Coronavirus/Besprechung-Der-Bundeskanzlerin-Mit-Den-Regierungschefinnen-Und-Regierungschefs-Der-Laender-Vom-22-03-2020-1733248, accessed: 2022-12-07, 2022.

Dormann, C., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J., Gruber, B., Lafourcade, B., Leitão, P., Münkemüller, T., Mcclean, C., Osborne, P., Reineking, B., Schröder, B., Skidmore, A., Zurell, D., and Lautenbach, S.: Collinearity: A review of methods to deal with it and a simulation study evaluating their performance, Ecography, 36, 27–46, https://doi.org/10.1111/j.1600-0587.2012.07348.x, 2013.

Du, J., Lang, H.-M., Ma, Y., Chen, A.-W., Qin, Y.-Y., Zhang, X.-P., and Huang, C.-Q.: Global trends in COVID-19 incidence and case fatality rates (2019–2023): a retrospective analysis, Frontiers in Public Health, 12, 1355 097, https://doi.org/10.3389/fpubh.2024.1355097, 2024.

Ecke, L., Magdolen, M., Chlond, B., and Vortisch, P.: How the COVID-19 pandemic changes daily commuting routines – Insights from the German Mobility Panel, Case Studies on Transport Policy, 10, 2175–2182, https://doi.org/10.1016/j.cstp.2022.10.001, 2022.

Ehlert, A.: The socio-economic determinants of COVID-19: A spatial analysis of German county level data, Socio-Economic Planning Sciences, 78, 101 083, https://doi.org/10.1016/j.seps.2021.101083, 2021.

Fox, J. and Monette, G.: Generalized Collinearity Diagnostics, Journal of the American Statistical Association, 87, 178, https://doi.org/10.2307/2290467, 1992.

Griffith, D., Chun, Y., and Li, B.: Spatial Regression Analysis Using Eigenvector Spatial Filtering, Academic Press, 2019.

Grolemund, G. and Wickham, H.: Dates and Times Made Easy with lubridate, Journal of Statistical Software, 40, 1–25, https://www.jstatsoft.org/v40/i03/, 2011.

Guliyev, H.: Determining the spatial effects of COVID-19 using the spatial panel data model, Spatial Statistics, 38, 100 443, https://doi.org/10.1016/j.spasta.2020.100443, 2020.

Kotseva, B., Vianini, I., Nikolaidis, N., Faggiani, N., Potapova, K., Gasparro, C., Steiner, Y., Scornavacche, J., Jacquet, G., Dragu, V., Della Rocca, L., Bucci, S., Podavini, A., Verile, M., Macmillan, C., and Linge, J. P.: Trend analysis of COVID-19 mis/disinformation narratives–A 3-year study, PLOS ONE, 18, e0291 423, https://doi.org/10.1371/journal.pone.0291423, 2023.

Kuhlmann, S. and Franzke, J.: Multi-level responses to COVID-19: crisis coordination in Germany from an intergovernmental perspective, Local Government Studies, 48, 312–334, https://doi.org/10.1080/03003930.2021.1904398, 2022.

Lewandowsky, M., Leonhardt, C., and Blätte, A.: Germany: The Alternative for Germany in the COVID-19 pandemic, in: Populists and the Pandemic, pp. 237–249, Routledge, 2022.

Liu, M., Liu, M., Li, Z., Zhu, Y., Liu, Y., Wang, X., Tao, L., and Guo, X.: The spatial clustering analysis of COVID-19 and its associated factors in mainland China at the prefecture level, Science of The Total Environment, 777, 145 992, https://doi.org/10.1016/j.scitotenv.2021.145992, 2021.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K.: cluster: Cluster Analysis Basics and Extensions, https://CRAN.R-project.org/package=cluster, r package version 2.1.6 — For new features, see the 'NEWS' and the 'Changelog' file in the package source), 2023.

Mishra, V., Seyedzenouzi, G., Almohtadi, A., Chowdhury, T., Khashkhusha, A., Axiaq, A., Wong, W. Y. E., and Harky, A.: Health Inequalities During COVID-19 and Their Effects on Morbidity and Mortality, Journal of Healthcare Leadership, Volume 13, 19–26, https://doi.org/10.2147/JHL.S270175, 2021.

Mollalo, A., Vahedi, B., and Rivera, K. M.: GIS-based spatial modeling of COVID-19 incidence rate in the continental United States, Science of The Total Environment, 728, 138 884, https://doi.org/10.1016/j.scitotenv.2020.138884, 2020.

Moran, P. A. P.: Notes on Continuous Stochastic Phenomena, Biometrika, 37, 17–23, 1950.

Murtagh, F. and Legendre, P.: Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?, Journal of classification, 31, 274–295, 2014.

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Borman, T., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H. B. A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M. O., Lahti, L., Martino, C., McGlinn, D., Ouellette, M.-H., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C. J., and Weedon, J.: vegan: Community Ecology Package, https://CRAN.R-project.org/package=vegan, r package version 2.7-1, 2025.

Pebesma, E.: Simple Features for R: Standardized Support for Spatial Vector Data, The R Journal, 10, 439–446, https://doi.org/10.32614/RJ-2018-009, 2018.

Pebesma, E. and Bivand, R.: Spatial Data Science: With applications in R, Chapman and Hall/CRC, https://doi.org/10.1201/9780429459016, 2023.

Qamar, A. I., Gronwald, L., Timmesfeld, N., and Diebner, H. H.: Local socio-structural predictors of COVID-19 incidence in Germany, Frontiers in Public Health, 10, 970 092, https://doi.org/10.3389/fpubh.2022.970092, 2022.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2024.

Reis, M., Michalski, N., Bartig, S., Wulkotte, E., Poethko-Müller, C., Graeber, D., Rosario, A. S., Hövener, C., and Hoebel, J.: Reconsidering inequalities in COVID-19 vaccine uptake in Germany: a spatiotemporal analysis combining individual educational level and area-level socioeconomic deprivation, Scientific Reports, 14, 23 904, https://doi.org/10.1038/s41598-024-75273-9, 2024.

Reuband, K.-H.: AfD-Affinitäten, Corona-bezogene Einstellungen und Proteste gegen die Corona-Maßnahmen: eine empirische Analyse auf Bundesländerebene, Zeitschrift für Parteienwissenschaften (MIP), 28, 67–94, 2022.

Roger Bivand: R Packages for Analyzing Spatial Data: A Comparative Case Study with Areal Data, Geographical Analysis, 54, 488–518, https://doi.org/10.1111/gean.12319, 2022.

Rohleder, S., Costa, D. D., and Bozorgmehr, P. K.: Area-level socioeconomic deprivation, non-national residency, and Covid-19 incidence: A longitudinal spatiotemporal analysis in Germany, eClinicalMedicine, 49, 101 485, https://doi.org/10.1016/j.eclinm.2022.101485, 2022.

Schilling, J., Buda, S., and Tolksdorf, K.: Zweite Aktualisierung der "Retrospektiven Phaseneinteilung der COVID-19-Pandemie in Deutschland, Epidemiologisches Bulletin, https://doi.org/10.25646/9787, publisher: Robert Koch-Institut, 2022.

Tennekes, M.: tmap: Thematic Maps in R, Journal of Statistical Software, 84, 1–39, https://doi.org/10.18637/jss.v084.i06, 2018.

Truffer, J., Frei, N., Schubert, I., Sohre, A., and Nachtwey, O.: The Formation and Consolidation of a Motley Crew: A Mixed-Method Overview of Swiss and German Corona Protests, German Politics and Society, 42, 21–48, 2024.

Tsalampouni, A.: Health systems in the European Union and policy responses to Covid-19: A comparative analysis between Germany, Sweden, and Greece, Journal of Public Health Research, 11, 22799036221129 413, https://doi.org/10.1177/22799036221129413, 2022.

Wachtler, B., Michalski, N., Nowossadeck, E., Diercke, M., Wahrendorf, M., Santos-Hövener, C., Lampert, T., and Hoebel, J.: Socioeconomic inequalities and COVID-19 – A review of the current international literature, Journal of Health Monitoring, https://doi.org/10.25646/7059, publisher: Robert Koch-Institut, 2020.

Wickham, H.: ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York, https://ggplot2.tidyverse.org, 2016.

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., and Yutani, H.: Welcome to the tidyverse, Journal of Open Source Software, 4, 1686, https://doi.org/10.21105/joss.01686, 2019.

Wildman, J.: COVID-19 and income inequality in OECD countries, The European Journal of Health Economics, 22, 455–462, https://doi.org/10.1007/s10198-021-01266-4, 2021.

Worobey, M., Pekar, J., Larsen, B. B., Nelson, M. I., Hill, V., Joy, J. B., Rambaut, A., Suchard, M. A., Wertheim, J. O., and Lemey, P.: The emergence of Sars-CoV-2 in Europe and North America, Science, 370, 564–570, 2020.

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. a., and Smith, G. M.: Mixed effects models and extensions in ecology with R, Springer New York, New York, NY, https://doi.org/10.1007/978-0-387-87458-6, series Title: Statistics for Biology and Health, 2009.