TI 2006-049/3
Tinbergen Institute Discussion Paper

The Use Of Spatial Filtering Techniques

Roberto Patuelli[1]
Daniel A. Griffith[2]
Michael Tiefelsdorf[2]
Peter Nijkamp[1,3]

[1] *Department of Spatial Economics, Free University of Amsterdam, The Netherlands;*
[2] *School of Social Sciences, University of Texas at Dallas, USA;*
[3] *Tinbergen Institute, The Netherlands.*

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to driessen@tinbergen.nl.
Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

THE USE OF SPATIAL FILTERING TECHNIQUES: THE SPATIAL AND SPACE-TIME STRUCTURE OF GERMAN UNEMPLOYMENT DATA

Roberto Patuelli,[1] Daniel A. Griffith,[2] Michael Tiefelsdorf[2] and Peter Nijkamp[1]

[1] Department of Spatial Economics, Free University of Amsterdam, The Netherlands
[2] School of Social Sciences, University of Texas at Dallas, USA

Abstract

Socio-economic interrelationships among regions can be measured in terms of economic flows, migration, or physical geographically-based measures, such as distance or length of shared areal unit boundaries. In general, proximity and openness tend to favour a similar economic performance among adjacent regions. Therefore, proper forecasting of socio-economic variables, such as employment, requires an understanding of spatial (or spatio-temporal) autocorrelation effects associated with a particular geographic configuration of a system of regions. Several spatial econometric techniques have been developed in recent years to identify spatial interaction effects within a parametric framework. Alternatively, newly devised spatial filtering techniques aim to achieve this end as well through the use of a semi-parametric approach. Experiments presented in this paper deal with the analysis of and accounting for spatial autocorrelation by means of spatial filtering techniques for data pertaining to regional unemployment in Germany. The available data set comprises information about the share of unemployed workers in 439 German districts (the NUTS-III regional aggregation level). Results based upon an eigenvector spatial filter model formulation (that is, the use of orthogonal map pattern components), constructed for the 439 German districts, are presented, with an emphasis on their consistency over several years. Insights obtained by applying spatial filtering to the database are also discussed.

1. Introduction

Spatial matters are of critical importance when considering socio-economic (and other) phenomena (see, for example, Bockstael 1996; Weinhold 2002), as well as because of their implications for policymaking (Lacombe 2004). To account for the presence of spatial structures that influence (positively or negatively) observable economic entities, such as unemployment or trade, calls for a rigorous and systematic assessment of their impact and extent. Spatial autocorrelation (SA) represents the correlation, computed among the values of a single georeferenced variable, that is attributable to the geographic proximity of the objects to which the values are attached. Introduction of the SA concept is, of course, a departure from the classical assumption of independence of observations constituting a single variable. SA also complements the concept of temporal autocorrelation, which has been extensively studied and dealt with in time-series econometrics. SA measures are used to quantify the nature and degree of the spatial correlation within a variable, or to test the assumption of independence or randomness. From a statistical analysis viewpoint, spatial correlation patterns are problematic, since they make standard statistics, such as correlation coefficients or ordinary least squares (OLS) estimates, potentially inappropriate.

This paper aims to provide an assessment of how important spatial effects are in explaining unemployment levels in Germany, and, particularly, to show that these (or, more precisely, a subset of these) patterns are consistent over time. The definition of stable and recognizable spatial patterns enables one to observe systematic differences in regional unemployment. Such findings can have implications for policy evaluation and strategic planning. This paper presents analyses carried out by means of a semi-parametric 'spatial filtering' technique, described in Griffith (2003), which is based on the decomposition of geographic weights matrices. In our analysis, these matrices are defined for 439 German districts, according to both topological and distance-based criteria – such as shared boundaries or centroid distance – and economic flows. In this regard, journey-to-work flows are employed as a proxy for economic linkages.

Kosfeld and Dreger (2004) investigate spatial patterns of German regional labour markets, for the period 1992–2000. However, their approach involves computing spatial filters for each year within the framework of a spatial seemingly unrelated regression (SUR) model. Our approach differs from theirs in that we focus on the search for a set of spatial filters that are significant and consistent over time, and therefore can be employed for the entire time period considered (that is, 1996–2002). Also, we employ data at a finer level of disaggregation (439 districts versus 180 regions), which enables a more detailed analysis of the underlying spatial patterns.

2. Spatial Filtering: An Overview

2.1 Preface

A wide array of methods, as well as several dedicated 'spatial' econometric procedures (see, for example, Anselin et al. 2004), for the statistical analysis of georeferenced data are available in the literature. These techniques are useful when analysing regional unemployment data, as in our case study, and, particularly, when the final aim is to develop forecasting models for some regional scale. Among conventional spatial econometric methods, spatial autoregression (see, among others, Anselin 1988; Griffith 1988) is a powerful method commonly employed. Spatial autoregressive techniques take into account spatial effects by means of geographic weights matrices that provide measures of the spatial linkages (dependence) between values of georeferenced variables. Because of bias, statistical efficiency concerns and the normality assumption, OLS should not be carried out with such data. Furthermore, maximum likelihood estimators of spatial regression models are based on restrictive assumptions. An alternative approach to spatial autoregression is the use of spatial filtering techniques, such as the ones described in Griffith (1981), Haining (1991), Getis and Griffith (2002), and Tiefelsdorf and Griffith (2006). The advantage of these filtering procedures is that the variables studied (which, initially, are spatially correlated) are split into spatial and non-spatial components, which can be employed in an OLS modelling framework. Filtering out spatially autocorrelated patterns also enables one to reduce the stochastic noise in the residuals of conventional statistical methods such as OLS. This conversion procedure requires the computation of 'spatial filters.' The approach developed by Griffith (1996; 2000) will be briefly described here. This approach is preferred in our case study to the one by Getis (1990; 1995), which requires variables with a natural origin. This constraint would not allow us to analyse patterns in employment growth rates, which will be studied in the future.

The spatial filtering technique introduced by Griffith is based on the computational formula of Moran's *I* (MI) statistic.[1] This methodology exploits eigenvector decomposition techniques, which extract *orthogonal* and *uncorrelated* numerical components from a *n* x *n* matrix (Tiefelsdorf and Boots 1995).[2] These components can be seen as independent map patterns, and represent the latent SA of a georeferenced variable concerned, according to a given geographic weights matrix. They also can be interpreted as redundant information due to spatial interdependencies, in the framework of standard regression equations.

Formally, these orthogonal components are the computed eigenvectors of the modified geographic weights matrix $(\mathbf{I}-\mathbf{1}\mathbf{1}^{\mathrm{T}}/n)\mathbf{C}(\mathbf{I}-\mathbf{1}\mathbf{1}^{\mathrm{T}}/n)$, where $\mathbf{C}$ is the given geographic weights matrix, $\mathbf{I}$ is an identity matrix of dimension *n* x *n*, and $\mathbf{1}$ is an *n* x 1 vector containing ones. The eigenvectors of the modified matrix are computed, in sequence, to maximize the sequential residual MI values. The first eigenvector, E_1, is, therefore, the one whose numerical values generate the largest MI value among all eigenvectors of the modified matrix. Similarly, the second eigenvector, E_2, is the set of numerical values that, again, maximize the MI value, while being uncorrelated with E_1. The process continues until *n* eigenvectors have been computed. This is the complete set of all possible (mutually) orthogonal and uncorrelated map patterns (Getis and Griffith 2002), and, when employed as regressors, they may function as proxies for missing explanatory variables.

A smaller set of 'candidate' eigenvectors then can be selected from the *n* eigenvectors, on the basis of their MI values, exceeding some prespecified threshold value. Since the eigenvectors are both orthogonal and uncorrelated, a stepwise linear regression can be used to achieve this end. In this framework, the advantage implied by the orthogonality of the eigenvectors is the absence of partial correlations and, therefore, of multicollinearity issues. Also, residuals obtained with stepwise regression constitute the *spatially filtered* component of the georeferenced variable examined. Each

---

[1] Moran's *I* coefficient is the most common, and oldest, indicator of SA. It is calculated as:

$$I=\frac{n\sum_i\sum_j w_{i,j}(x_i-\bar{x})(x_j-\bar{x})}{(\sum_i\sum_j w_{i,j})\sum_i(x_i-\bar{x})^2},$$

where: n is the number of cases; x_i is the value of variable X at location i; and $w_{i,j}$ is the cell (i, j) of the geographic weights matrix **W** (see Section 2.1). Positive autocorrelation ($I > 0$) implies that geographical proximity tends to produce similar values of the variable examined. This is a phenomenon that is often observed in reality, especially in economics. Negative SA ($I < 0$) is a much rarer phenomenon.

[2] Griffith's spatial filtering techniques are often compared to principal components analysis (CPA), in that both methodologies generate orthogonal and uncorrelated new 'variables' that can be employed in regression analyses. However, the components derived in PCA have an economic interpretation because eigenvectors are used to construct linear combinations of attribute variables, whereas spatial filters are linear combinations of the eigenvectors themselves, and as such should be regarded mostly as patterns of independent spatial dimensions.

eigenvector selected for inclusion is considered to be part of a 'spatial filter' for the dependent variable. The top two eigenvectors computed (E_1 and E_2) often identify map patterns along the cardinal points, that is, major North-South and East-West patterns. Eigenvectors with intermediate values of MI display regional map patterns, whereas eigenvectors with smaller values of MI display local map patterns. A linear combination of the above eigenvectors can be defined as *the* spatial filter for the variable examined.

Also relevant to the use of the eigenvector decomposition process is the choice of the matrix to be used, particularly regarding: (a) the definition of proximity; (b) the variable chosen (if necessary) to indicate proximity; and (c) the coding scheme employed in the calculation of the matrix. While points (a) and (b) will be discussed later in the paper, the latter point will be just briefly addressed in the subsequent section.

2.2 Coding of Geographic Weights Matrices

The spatial filters presented in the previous section are computed on the basis of a modified geographic weights matrix. It is straightforward that the choice of the matrix to be used is critical in defining the set of spatial filters. Many coding techniques for geographic weights matrices can be found in the literature (Tiefelsdorf et al. 1999; Getis and Aldstadt 2004). The main factor that discriminates between the different schemes is the way in which each scheme treats the spatial links between georeferenced objects (like regions).

Generally speaking, we can define a *family* of coding schemes based on the following expression (Tiefelsdorf and Griffith, 2006, with details in Chun et al. 2005):

$$\mathbf{V}_{[q]} = \frac{n}{\sum_{i=1}^{n} d_i^{q+1}} \cdot \mathbf{D}^q \cdot \mathbf{B}, \qquad (1)$$

where **B** is a binary geographic weights matrix, and $\mathbf{D}^q$ is a diagonal matrix that contains d_i^q components ($d_1^q,...,d_n^q$), belonging to vector $\mathbf{d} = \mathbf{B} \cdot 1$, and representing the degree of 'linkage' of spatial object i. Different coding schemes are obtained by varying the q parameter. In particular, the following schemes can be obtained:

- $q = 0$: *C-coding* (globally standardized). This scheme commonly is used in spatial statistics, and tends to emphasize spatial objects with a greater linkage degree. The C-coded matrix is symmetrical;
- $q = -0.5$: *S-coding* (variance stabilized). This scheme tends to even the variation levels of weights assigned to spatial objects; and,
- $q = -1$: *W-coding* (row-sum standardized). This scheme mostly is used in autoregressive response and simultaneous spatial autoregressive model specifications, and, contrary to the C-coding scheme, tends to emphasize the weight of objects with small spatial linkages.

Different spatial patterns may well result from the calculation of the eigenvectors of the above coded matrices. For instance, a W-coded matrix can be expected to show more 'extreme' values along the edges of a study area, while, consequently, a C-coded matrix is expected to present stronger patterns in the inner study area. Figure 1 presents an illustrative example, for the case of German unemployment, of the first two eigenvectors generated from the adjacency matrix coded in the different coding schemes.

FIGURE 1 HERE

The choice of coding scheme, and therefore of the geographic weights matrix, not only determines the set of eigenvectors from which the spatial filters are selected, but also is a factor with a view to the utilization of its results in a spatial econometric or spatial statistics framework. In the empirical application presented in this paper, both W-coding and C-coding are employed (see Section 4). Results of a correlation analysis of the geographic weights matrices used also are presented, in order to compare the different approaches.

3. The Data

The above spatial filtering techniques will now empirically be illustrated. This paper presents results based on German unemployment data. The data set consists of cross-sectional data, collected by the (German) Federal Employment Services (Bundesanstalt für Arbeit, BA), on 439 German districts ('Kreise'). The time period for which the data are available is from 1996 to 2002, while the level of aggregation of the data set is NUTS-3. In particular, the NUTS-3 aggregation level enables a more detailed examination of 'local' unemployment patterns. In fact, data at the NUTS-2 level would have only 41 regions ('Regierungsbezirke'). Alternatively, an intermediate approach is proposed by Kosfeld and Dreger (2004), who carry out a spatial filtering analysis of German regional labour market data, using 180 previously defined regional labour market areas (Eckey 2001). The unemployment rates employed in our analysis are computed as a ratio between the number of unemployed individuals and the active workers population.

A further spatial relationship matrix, German commuting flows, is employed in our analysis. The data consist of, for each couple (i, j) of NUTS-3 origin and destination, the number of employees that live in district i and work in district j. Therefore, we can treat these flows as home-to-work trips. The data used in this paper refer to the year 2002, and are employed in the computation of an 'economic flows' geographic weights matrix (see Section 4). Commuting data for one year only are employed in our case study, since varying commuting data would generate different geographic weights matrices, and, consequently, different sets of eigenvectors. Furthermore, one can assume some spatio-temporal persistence with respect of the local commuting patterns. The daily commuting flows between two districts is transformed to satisfy the statistical symmetry requirement of spatial link matrices. This transformation models the daily to-work and back-to-home flows.

4. The Empirical Application: The Computation and Choice of Spatial Filters for German Unemployment

4.1 Geographic Weights Matrices: The Different Approaches Used

As previously mentioned, the spatial filtering methods employed in this case study are based on the decomposition of a geographic weights matrix. Therefore, it is important to carefully consider, in addition to matrix computation methods (see Section 2.2) the concept of proximity employed, and its consequences.

In our case study, we present a set of different definitions of the geographic weights matrix:

- *economic flows*: based on patterns of commuting flows;

- *shared boundaries*: based on geographical contiguity, which by definition is symmetric;
- *distance*: based on symmetric distances separating district centroids.

The three definitions highlighted here enable one to observe the influence of different operational definitions of proximity on the final results. First, commuting flows are employed as a proxy of the economic intertwining among districts. Second, shared boundaries utilize the topology of administrative boundaries in defining proximity. Third, distance-based matrices calculated using districts centroids define proximity in terms of geographical distance decay relationships.

A total of five geographic weights matrices are employed in this paper. The matrices are computed as follows:

a) A journey-to-work flows matrix is computed according to the q = –1 scaling scheme (W-coding); this matrix is based upon the location-to-location commuting data described in Section 3.
b) Two matrices based upon shared boundaries, constructed by defining contiguity according to the so-called 'rook' rule, and then computed according to the C- and W-coding schemes; results from the application of a 'queen' contiguity rule are considered here, since the two specifications of adjacency differ only by 25 neighbour links.
c) Two distance-based matrices derived from a spatial interaction model[3]; the variables used for the estimation of the model are district employment data (presented in Section 3), and the distance between the centroids of each district:
    a. First, the distance decay exponent of –2.7 is taken from the estimated spatial interaction model (SIM), and then converted to the W-coding scheme ;
    b. Second, this distance decay exponent is increased to –6.3 in order to obtain the same number of candidate eigenvectors as are obtained with the shared boundaries W-coding scheme.

The following unconstrained gravity model is the spatial interaction model used to describe flows and estimate distance decay parameters:

$$F_{ij} = \kappa W_i^{\alpha} J_j^{\beta} e^{-\gamma d_{ij}} + \varepsilon_{ij}, \tag{2}$$

where:

F_{ij} is the quantity of flows between areal units i and j;
W_i is the number of works residing in origin areal unit i;
J_j is the number of jobs located in destination j;
κ, α, β and γ are parameters; and
ε_{ij} is a random error associated with flows between origin i and destination j.

The estimated distance decay parameter, $\hat{\gamma}$, was used to define the W-coding scheme.

$$w_{ij} = \frac{e^{-\hat{\gamma} d_{ij}}}{\sum_{j=1}^{n} e^{-\hat{\gamma} d_{ij}}}. \tag{3}$$

[3] For details about the estimation of spatial interaction models, see, among others, Sen and Smith (1995), and Haynes and Fotheringham (1984).

Next, $\hat{\gamma}$ was incrementally increased until the resulting $(\mathbf{W}^T + \mathbf{W})/2$ matrix yielded the same number of prominent eigenvectors as are obtained with $(\mathbf{W}^T + \mathbf{W})/2$ constructed as the row-standardized version of the topological-based binary 0/1 adjacency matrix **C**. Of note is that the eigenvectors for all W-coding schemes are extracted from $(\mathbf{W}^T + \mathbf{W})/2$ in order to convert the matrix from an asymmetric to a symmetric one.

4.2 Computation and Selection of the Spatial Filters over Time

The first step in the construction of a spatial filter to be applied to the variable of study is the computation of the eigenvectors of the geographic weights matrix, followed by the choice of a set of candidate eigenvectors from which selection is made. Eigenvectors are selected for inclusion on the basis of their MI values and their correlations with the georeferenced data on regional unemployment. A minimum MI/max(MI) value of 0.25 has been used in our case to identify the candidate set. The results of this process, carried out for the matrices presented in the preceding section, are presented in Table 1.

Table 1 – Candidate eigenvectors selected and maximum MI values

Geographic weights matrix	*# of candidate eigenvectors*	*max(MI)*
Journey-to-work flows matrix	78	2.92
Rook matrix (S-coding)	130	1.07
Rook matrix (C-coding)	98	1.24
Distance-based matrix ($\beta = -2.7$)	36	0.97
Distance-based matrix ($\beta = -6.3$)	97	1.02

Once the sets of 'candidate' eigenvectors have been selected, the statistical significance of each, as an explanatory variable for German regional unemployment has to be established. This process was carried out by means of a stepwise logistic regression analysis. The stopping condition employed is a 10% level of significance for inclusion and retention. In addition to the stepwise regression analysis, a further manual backward elimination of regressors was carried out through the sequential estimation of a generalized linear model coupled with a binomial distribution. A marginal eigenvector was excluded as long as its χ^2 value remained non-significant.

The same process was repeated for all years of available data – from 1996 to 2002 – and for each geographic weights matrix. Consequently, seven sets of 'significant' eigenvectors (one set for each year) have been selected, for each of the employed spatial relationship matrices. These are the 'spatial filters' uncovered for each year and matrix.

Next, for each matrix we pinpointed a subset of eigenvectors that is common to the years 1996 to 2002. The results of the analyses described above are summarized in Table 2. Details about the eigenvectors selected in each context and year are shown in the Appendix (Table A). Of particular note, in Table A, is that the sum-of-squared prediction error (SSPE) divided by the mean squared error (MSE) in all cases is roughly 1 (that is, $\sqrt{\text{SSPE}/\text{MSE}}$); in other words, a jackknife type of cross-validation assessment of the selected eigenvectors yields prediction error that is almost identical to the OLS error minimization results, validating the constructed spatial filters.

Table 2 – Amount of variance explained by the selected eigenvectors, and the number of common eigenvectors, 1996–2002

Geographic weights matrix	# of common eigenvecs	Adj. pseudo R^2 1996	Adj. pseudo R^2 1997	Adj. pseudo R^2 1998	Adj. pseudo R^2 1999	Adj. pseudo R^2 2000	Adj. pseudo R^2 2001	Adj. pseudo R^2 2002
Journey-to-work flows matrix	14	0.3004	0.2911	0.3305	0.3142	0.3379	0.3453	0.3285
Rook matrix (S-coding)	17	0.6477	0.6821	0.7293	0.7453	0.7945	0.8022	0.7909
Rook matrix (C-coding)	15	0.5929	0.6425	0.6846	0.7068	0.7483	0.7683	0.7549
Distance-based matrix ($\beta = -2.7$)	6	0.6215	0.5968	0.6519	0.6930	0.7296	0.7448	0.7382
Distance-based matrix ($\beta = -6.3$)	11	0.6233	0.6067	0.6501	0.6818	0.7247	0.7442	0.7331

The results summarized in Table 2 show that we found sets of eigenvectors (spatial filters) that are significant, as explanatory variables of regional unemployment, over the entire time period considered. Of note here is that all contexts (i.e., economic flows, shared boundaries, and distance) enable us to define sets of common spatial filters.

In terms of statistical relevance, the amount of variance explained by the spatial filtering regressors is fairly consistent over the years (reasonably, unemployment patterns do not change much from year to year), and at comparable levels, for all the *geographic* contexts (that is, shared boundaries and distance). The adjusted pseudo-R^2 values found for these analyses are around 0.60–0.80, with the S-coded rook matrix approach being the most significant. The results obtained for the commuting flows matrix approach are not as encouraging. The amount of variance explained by the model, in this case, is only in the 0.29–0.35 range.

A plot of the real and estimated unemployment values is shown in Figure 2. These plots refer to the rook adjacency matrix S-coding scheme and to the years 1996 and 2002, and show a fairly good fit, though a tendency toward underestimation can be observed, particularly for the year 2002, which exhibits more 'extreme' unemployment percentages.

FIGURE 2 HERE

As mentioned in Section 2, the constructed spatial filters can be interpreted not only as potential explanatory variables substituting for missing ones, but also as map patterns. A graphical visualization of the spatial filters uncovered by our analysis provides an example of the map features embedded in the eigenvectors' values. Figure 3 shows the top four spatial filters with the largest MI values computed for the rook adjacency matrix S-coding scheme, and that are common to all the years examined.

FIGURE 3 HERE

As noted previously, the first two eigenvectors for adjacency matrices usually show East-West and North-South patterns. Spatial filter (a) (E_2) in Figure 3 seems, in fact, to be characterized by a North-South pattern. When we observe the subsequent spatial filter components (b, c and d), the geographic patterns mapped relate to characteristics of smaller geographical scale, showing patterns that can be categorized first as 'regional', then as 'local'. Athough they may contain some common map patterns (for example, North-South and East-West patterns), spatial filters computed with different geographic weights matrices will vary to some degree. Meanwhile, an assessment of the statistical significance of the spatial filters (shown in Table 2) enables us to assess the utility of the different proximity approaches employed.

4.3 Concluding Remarks: The Results of Different Proximity Approaches

The preceding section reveals that all of the definitions employed in this paper in order to operationalize proximity have been found to generate sets of eigenvectors (whose linear combinations are spatial filters) that are significantly correlated with the dependent variable, regional unemployment, and for all the years examined. Consequently, our focus is on similarities and differences in the statistical performance of the different definitions used.

In order to understand the descriptive performance associated with different geographic weights matrices, we need to compare the matrices themselves. Therefore, a correlation analysis of the matrices employed in our paper has been carried out. Results of this analysis appear in Table 3 (for details on the computation of matrix correlation, see Oden 1984, and Tiefelsdorf 2000).

Table 3 – Correlations of geographic weights matrices

	Journey-to-work flows matrix	*Rook matrix (S-coding)*	*Rook matrix (C-coding)*	*Distance-based matrix ($\beta = -2.7$)*	*Distance-based matrix ($\beta = -6.3$)*
Journey-to-work flows matrix	1.0000	0.5641	0.5102	0.4919	0.5949
Rook matrix (S-coding)	0.5641	1.0000	0.9152	0.6892	0.7923
Rook matrix (C-coding)	0.5102	0.9152	1.0000	0.6533	0.6879
Distance-based matrix ($\beta = -2.7$)	0.4919	0.6892	0.6533	1.0000	0.8775
Distance-based matrix ($\beta = -6.3$)	0.5949	0.7923	0.6879	0.8775	1.0000

Several features of Table 3 are noteworthy. The most conspicuous result pertains to the correlations between the journey-to-work flows matrix and the remaining matrices (that is, shared boundaries and distance-based). The low correlation values found are plausible and, to a certain degree, to be expected. The flows matrix differs from the other matrices in that it is not based on topology, but is a proxy for economic links between the districts. These links are, in fact, not fully limited by geographic contiguity, embracing hierarchical components of the geographic landscape, as well. With regard to the remaining matrices, they all seem to have fairly high correlations, which would be consistent with similarities in the statistical performance of their computed eigenvectors (see Table 2).

Also of note is that:

- matrices based on more similar definitions tend to be more strongly correlated with each other than with those based on less similar definitions;
- the correlation between the two rook adjacency-based matrices is higher than between the two distance-based matrices, in spite of the different coding schemes employed; and,
- both distance-based matrices, which have been constructed with the W-coding scheme, seem to be more strongly correlated with the S-coded than with the C-coded rook matrix.

These findings call for a more in-depth analysis of the issues related to the choice of a coding scheme, particularly in view of the type of data patterns that a spatial analyst wants to emphasize. The discussion of such problems goes beyond the scope of this paper; an interesting treatment can be found in Tiefelsdorf et al. (1999).

5. Conclusions

In this paper we present an analysis of German regional unemployment by means of 'spatial filtering' techniques. The analysis enables us to uncover spatial structures underlying the georeferenced unemployment data by selecting sets of 'spatial filters' that significantly explain geographic variations in the data. In addition, we have observed subsets of spatial filters that (partially) define the spatial structures of the data over time. The spatial filters selected in this case are the ones that were common to the analyses carried out for each year in the 1996–2002 period. Several definitions have been employed, in order to operationalize spatial linkages according to geographic and non-geographic criteria. All of these definitions have yielded sets of time-stable spatial filters, though at different levels of statistical significance.

Initial sets of eigenvectors have been selected on the basis of the SA they accounted for (that is, by decreasing MI values), only to be later reduced by means of stepwise regression followed by manual backward elimination. The final subsets of eigenvectors used to construct spatial filters render fairly satisfactory statistical descriptions. In the shared boundaries- and distance-based approaches, the spatial filters explain 60 to 80 per cent of the total variance when employed as the sole regressors of unemployment in a generalized linear regression model. But the 'economic flows' approach, based on a journey-to-work flows matrix, fail to produce the same encouraging results. This finding might be due to the artificial nature of the data used (logical connections between districts) and the lack of a more proper measure of regional economic linkages.

A correlation analysis (see Section 4.3) of the geographic weights matrices employed in our analysis shows that matrices computed on the basis of the same proximity measure tend to be highly correlated, regardless of the coding scheme applied in their standardization. Also, the journey-to-work matrix seems to be much less correlated with the topological-based matrices. This result is consistent with the varying statistical performance of the spatial filters computed.

If shown as graphical visualizations, the spatial filters found in our analyses provide hints on the geographical distribution of unemployment trends. Using Figure 3 as an example, map (a) can be interpreted as the visualization of a North-South divide, while map (b) seems to show lower Bavaria as opposed to the rest of the country. Both maps (c) and (d) suggest differences between East and West Germany. Additional eigenvectors (not shown here) show smaller scale patterns of regional/local spatial dependency structure.

The results obtained in this paper have illustrated the spatial structure underlying the georeferenced unemployment data. However, future research along these lines is needed. On the empirical side, a better proxy of spatial economic linkages should be employed, one that differs from commuting flows. Also, the analysis of unemployment levels has its counterpart in that of employment growth rates. Future investigations need to address this issue. On the methodological side, a comparison of the performance of the spatial autoregressive approach with other conventional spatial econometrics methods, as well as with non-linear approaches, such as neural networks, is desirable. Mixed neural networks/spatial filtering approaches also should be tested. From a policy perspective, examination of the spatially-filtered residuals resulting from the analysis should be carried out, in order to fully grasp the benefits of the methodology applied.

Finally, the identification of sets of statistically significant and time-stable spatial filters also enables us to employ them in a dynamic setting. This task will be undertaken in detail in future research.

5.1 A Preliminary Spatial Autoregressive Dynamic Model for German Unemployment

Future research will deal with the utilization of the spatial filters computed for German regional unemployment in a dynamic framework. As a preliminary step, a generalized linear mixed model was estimated, for the case of the rook geographic weights matrix using the C-coding scheme. The 15 common selected eigenvectors (see Section 4.2) were entered as regressors in a generalized linear model with a binomial distribution for the response variable (a graphical visualization of the linear combination of these eigenvectors appears in Figure 4), together with a normally distributed random effects intercept variable to handle the temporal correlation.

Table 4 presents summary results regarding the spatial autocorrelation accounted for by this model.

Table 4 – Spatial autocorrelation measures for German unemployment, based upon the rook (C-coding) geographic weights matrix

Variable	*Raw data*			*Spatial filter residuals*	
	Moran's I	z_I	*Gearyratio*	*Moran's* I	*Gearyratio*
1996 unemployment	0.6651	21.9	0.3213	0.2107	0.6161
1997 unemployment	0.7320	24.1	0.3268	0.2004	0.6627
1998 unemployment	0.7596	25.0	0.2869	0.1999	0.6389
1999 unemployment	0.7854	25.8	0.2492	0.2057	0.6128
2000 unemployment	0.8324	27.4	0.2222	0.2454	0.5862
2001 unemployment	0.8537	28.1	0.2088	0.2653	0.5701
2002 unemployment	0.8500	28.0	0.2140	0.2713	0.5632
Spatial filter	1.1358	–	0.1459		

Note: z_I denotes the *z*-score for Moran's *I*.

The statistical results presented in Table 4 show that the spatial filter (linear combination of the common set of eigenvectors employed) accounts for a large share of SA, though not all of it (a perfectly random map pattern, free of SA, has a Moran's I of -0.0023 and a Geary ratio of 1). In terms of goodness-of-fit, the model has an adjusted-pseudo-R^2 of 0.9425, and all of the eigenvectors employed are significant. Figure 5 provides plots of the observed and predicted unemployment values. The model is shown to fit the data fairly well. In particular, better fitting is found as information on the previous years is fed into the model (Figure 5b).

FIGURE 5 HERE

Future research will start from this preliminary estimation in order to carry out more detailed experiments on the dynamics of unemployment patterns. The utilization, in the model, of additional explanatory variables should also be attempted. The joint employment of spatial filters and other explanatory variables involves further attention to spatial filtering. Eigenvectors that are significant both to the explained and an explanatory variable(s) imply filtering also of the latter. This issue will be addressed in the framework of the dynamic modelling.

Acknowledgments

The authors wish to thank Prof. Uwe Blien (IAB, Nuremberg, Germany) for providing the data on the labour market, as well as Prof. Günter Haag (STASA, Stuttgart, Germany) for the data set on commuting flows.

References

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers

Anselin, L., R.J.G.M. Florax and S.J. Rey (eds) (2004) *Advances in Spatial Econometrics*. Advances in Spatial Science. Berlin Heidelberg New York, Springer

Bockstael, N.E. (1996) Economics and Ecological Modeling: The Importance of a Spatial Perspective. *American Journal of Agricultural Economics* 78 (5), pp. 1168-80

Chun, Y., R. Bivand and M. Tiefelsdorf (2005) Using Open Source Data Analysis Environments for Prototyping Modelling Implementations for Spatial Data: Weights in R. Paper presented at the Geo-Computational Meeting, Ann Arbor, MI, August 1-3

Eckey, H.-F. (2001) Der Wirtschafliche Entwicklungsstand in den Regionen des Vereinigten Deuschland

Getis, A. (1990) Screening for Spatial Dependence in Regression Analysis. *Papers of the Regional Science Association* 69, pp. 69-81

Getis, A. (1995) Spatial Filtering in a Regression Framework: Examples Using Data on Urban Crime, Regional Inequality, and Government Expenditures. In: *New Directions in Spatial Econometrics*, L. Anselin and R.J.G.M. Florax (eds). Heidelberg, Springer, pp. 172-85

Getis, A. and D.A. Griffith (2002) Comparative Spatial Filtering in Regression Analysis. *Geographical Analysis* 34 (2), pp. 130-40

Getis, A. and J. Aldstadt (2004) Constructing the Spatial Weights Matrix Using a Local Statistic. *Geographical Analysis* 36 (2), pp. 90-104

Griffith, D. (1981) Towards a Theory of Spatial Statistics: A Rejoinder. *Geographical Analysis* 13, pp. 91-93

Griffith, D. (1988) *Advanced Spatial Statistics*. Dordrecht, Kluwer Academic

Griffith, D.A. (1996) Spatial Autocorrelation and Eigenfunctions of the Geographic Weights Matrix Accompanying Geo-Referenced Data. *The Canadian Geographer* 40, pp. 351-67

Griffith, D.A. (2000) A Linear Regression Solution to the Spatial Autocorrelation Problem. *Journal of Geographical Systems* 2, pp. 141-56

Griffith, D.A. (2003) *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*. Berlin, New York, Springer

Haining, R. (1991) Bivariate Correlation and Spatial Data. *Geographical Analysis* 23, pp. 210-27
Haynes, K.E. and A.S. Fotheringham (1984) *Gravity and Spatial Interaction Models*. Beverly Hills, Sage Publications
Kosfeld, R. and C. Dreger (2004) Thresholds for Employment and Unemployment. A Spatial Analysis of German Regional Labour Markets 1999-2000, Discussion Papers in Economics 52/04, University of Kassel, Institute of Economics
Lacombe, D.J. (2004) Does Econometric Methodology Matter? An Analysis of Public Policy Using Spatial Econometric Techniques. *Geographical Analysis* 36 (2), pp. 105-18
Oden, N.L. (1984) Assessing the Significance of a Spatial Correlogram. *Geographical Analysis* 16, pp. 1-16
Sen, A. and T.E. Smith (1995) *Gravity Models of Spatial Interaction Behavior*. Heidelberg and New York, Springer
Tiefelsdorf, M. and B. Boots (1995) The Exact Distribution of Moran's *I*. *Environment and Planning A* 27, pp. 985-99
Tiefelsdorf, M., D.A. Griffith and B.N. Boots (1999) A Variance Stabilizing Coding Scheme for Spatial Link Matrices. *Environment and Planning A* 31, pp. 165-80
Tiefelsdorf, M. (2000) *Modelling Spatial Processes – The Identification of Spatial Relationships in Regression Residuals by Means of Moran's* I. Berlin: Springer-Verlag
Tiefelsdorf, M. and D.A. Griffith (2006) Semi-parametric Filtering of Spatial Autocorrelation: The Eigenvector Approach. *Environment and Planning A* (forthcoming)
Weinhold, D. (2002) The Importance of Trade and Geography in the Pattern of Spatial Dependence of Growth Rates. *Review of Development Economics* 6 (3), pp. 369-82

Annex

Table A – Common and year-specific eigenvectors selected, years 1996–2002

Year	*# of eigenvecs*	*Year-specific eigenvectors*			*Common eigenvectors*			*Scale*	*Adj. pseudo R^2*	$\sqrt{SSPE/SSE}$
		Global	*Regional*	*Local*	*Global*	*Regional*	*Local*			
Eigenvectors extracted from the journey-to-work flows matrix (78 candidate eigenvectors)										
1996	20		E10	E35, E52, E62, E63, E69	E1	E4, E5, E7	E13, E18, E19, E25, E38, E44, E48, E50, E54, E77	26.6747	0.3004	1.0438
1997	18			E31, E62, E63, E78				31.3204	0.2911	1.0407
1998	22		E10	E31, E62, E63, E68, E70, E71, E78				31.0519	0.3305	1.0550
1999	20		E10	E31, E62, E63, E71, E78				32.1972	0.3142	1.0442
2000	21		E10	E31, E62, E63, E69, E71, E78				35.2523	0.3379	1.0487
2001	21		E10	E31, E62, E63, E69, E70, E78				37.9095	0.3453	1.0507
2002	19		E10	E31, E68, E71, E78				37.9052	0.3285	1.0411

Year	# of eigenvecs	Year-specific eigenvectors			Common eigenvectors			Scale	Adj. pseudo R^2	$\sqrt{\text{SSPE/SSE}}$
		Global	Regional	Local	Global	Regional	Local			
			Eigenvectors extracted from the rook matrix (S-coding) (130 candidate eigenvectors)							
1996	23	E1	E24, E25, E60	E113, E124	E2,	E15,	E130	20.8775	0.647662	1.0226
1997	24		E17, E25, E28, E70, E82, E97	E113	E3, E5,	E16, E22,		23.3176	0.682067	1.0395
1998	20		E14, E25, E28, E36, E60, E70, E82	E113, E129	E6, E7,	E39, E41,		22.0736	0.729332	1.0476
1999	26		E14, E23, E36, E38, E70, E82	E113, E115, E129	E8, E9,	E52, E71		22.0632	0.745331	1.0438
2000	31		E14, E25, E28, E33, E36, E38, E40, E50, E70, E82, E85	E113, E115, E129	E10, E11			22.0026	0.794492	1.0702
2001	28		E14, E18, E23, E32, E36, E38, E40, E82	E110, E115, E129				23.8654	0.802178	1.0489
2002	25		E14, E23, E36, E38, E40, E82	E115, E129				24.8114	0.790917	1.0387

Year	# of eigenvecs	Year-specific eigenvectors			Common eigenvectors			Scale	Adj. pseudo R^2	$\sqrt{\text{SSPE}/\text{SSE}}$
		Global	Regional	Local	Global	Regional	Local			
Eigenvectors extracted from the rook matrix (C-coding) (98 candidate eigenvectors)										
1996	24		E9, E16, E21, E25, E41, E52, E53, E64	E89	E2, E3, E4, E5	E6, E7, E8, E11, E18, E24, E28, E30, E39, E60	E74	21.9823	0.5929	1.0232
1997	23	E1	E15, E19, E21, E34, E38, E64	E93				24.3798	0.6425	1.0412
1998	27		E13, E15, E16, E19, E21, E34, E38, E42, E52, E66	E68, E93				23.5231	0.6846	1.0438
1999	27		E9, E13, E15, E16, E19, E21, E34, E38, E42, E52, E66	E93				23.2457	0.7068	1.0364
2000	30		E9, E13, E15, E16, E19, E21, E25, E34, E38, E42, E51, E52, E66	E93, E97				23.8338	0.7483	1.0507
2001	30		E9, E12, E13, E15, E16, E19, E34, E42, E52, E56, E65, E66	E68, E93, E97				25.1826	0.7683	1.0489
2002	29	E1	E9, E12, E13, E15, E16, E19, E20, E25, E38, E42, E52, E65, E66					26.0816	0.7549	1.0459

Year	*# of eigenvecs*	*Year-specific eigenvectors*			*Common eigenvectors*			*Scale*	*Adj. pseudo* R^2	$\sqrt{\text{SSPE}/\text{SSE}}$
		Global	*Regional*	*Local*	*Global*	*Regional*	*Local*			
Eigenvectors extracted from the distance-based matrix (β = –2.7) (36 candidate eigenvectors)										
1996	18		E7, E11, E12, E14	E26, E29, E30, E32, E34	E1, E2, E3	E5, E6, E16, E23		21.5904	0.6215	1.0018
1997	13		E7, E12, E14, E17	E31, E32				25.6508	0.5968	1.0063
1998	14		E11, E12, E17, E21	E26, E31, E32				24.8788	0.6519	1.0079
1999	14		E11, E14, E17, E21	E26, E30, E31				24.4060	0.6930	1.0040
2000	15		E11, E14, E17, E21	E26, E30, E31, E32				25.5416	0.7296	1.0092
2001	14		E11, E17, E20, E21	E26, E30, E31				26.6811	0.7448	1.0087
2002	13		E11, E17, E20, E21	E26, E30				26.8710	0.7382	1.0101

Year	*# of eigenvecs*	*Year-specific eigenvectors*			*Common eigenvectors*			*Scale*	*Adj. pseudo R^2*	$\sqrt{\text{SSPE}/\text{SSE}}$
		Global	*Regional*	*Local*	*Global*	*Regional*	*Local*			
		Eigenvectors extracted from the distance-based matrix ($\beta = -6.3$) (97 candidate eigenvectors)								
1996	24		E13, E17, E26, E27, E29, E31, E34, E35, E39, E52	E90, E96	E1, E2, E3, E5, E6, E8	E15, E32, E55, E64	E91	20.7515	0.6233	1.0460
1997	20	E7	E13, E17, E20, E23, E34, E35	E79, E96				24.6605	0.6067	1.0334
1998	21		E13, E17, E20, E23, E24, E26, E31, E34	E79, E96				24.2862	0.6501	1.0424
1999	19		E13, E20, E23, E26, E39, E63	E79, E96				24.0949	0.6818	1.0248
2000	24		E13, E17, E20, E23, E24, E26, E29, E39, E40, E63	E71, E79, E96				24.6917	0.7247	1.0371
2001	23		E13, E17, E20, E23, E24, E25, E26, E39, E40, E63	E79, E96				26.0291	0.7442	1.0392
2002	21		E17, E20, E23, E25, E26, E27, E39, E40, E63	E79				26.9843	0.7331	1.1589

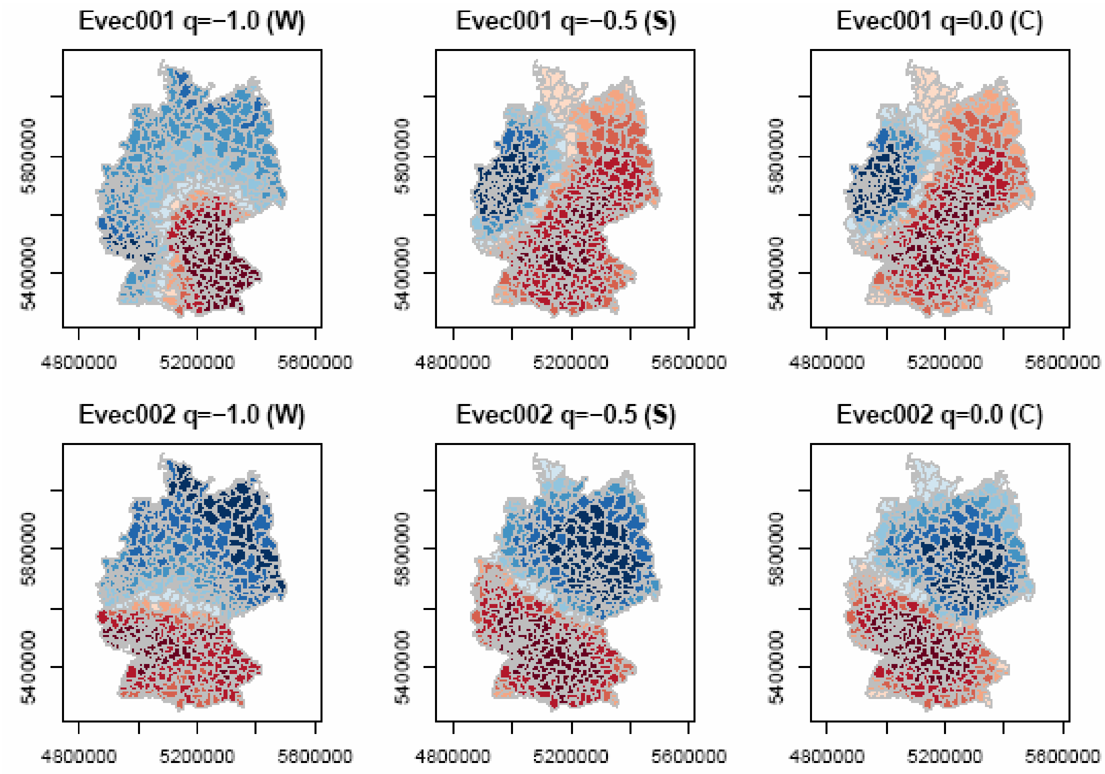
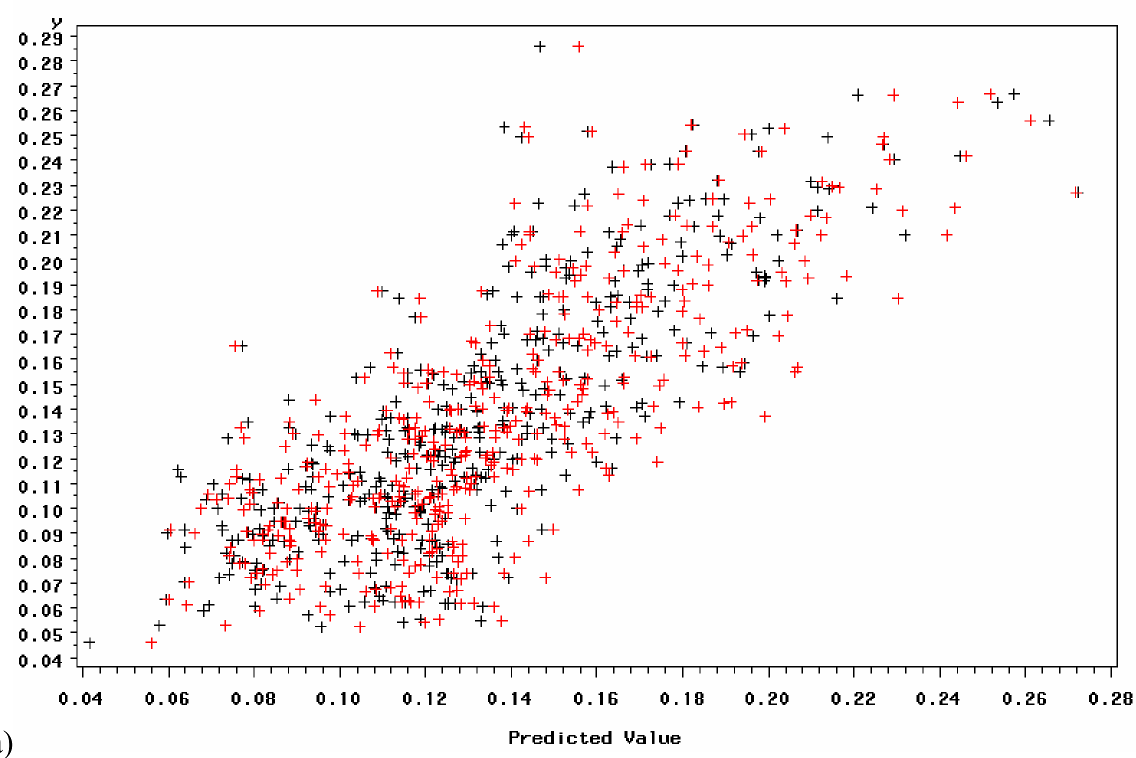
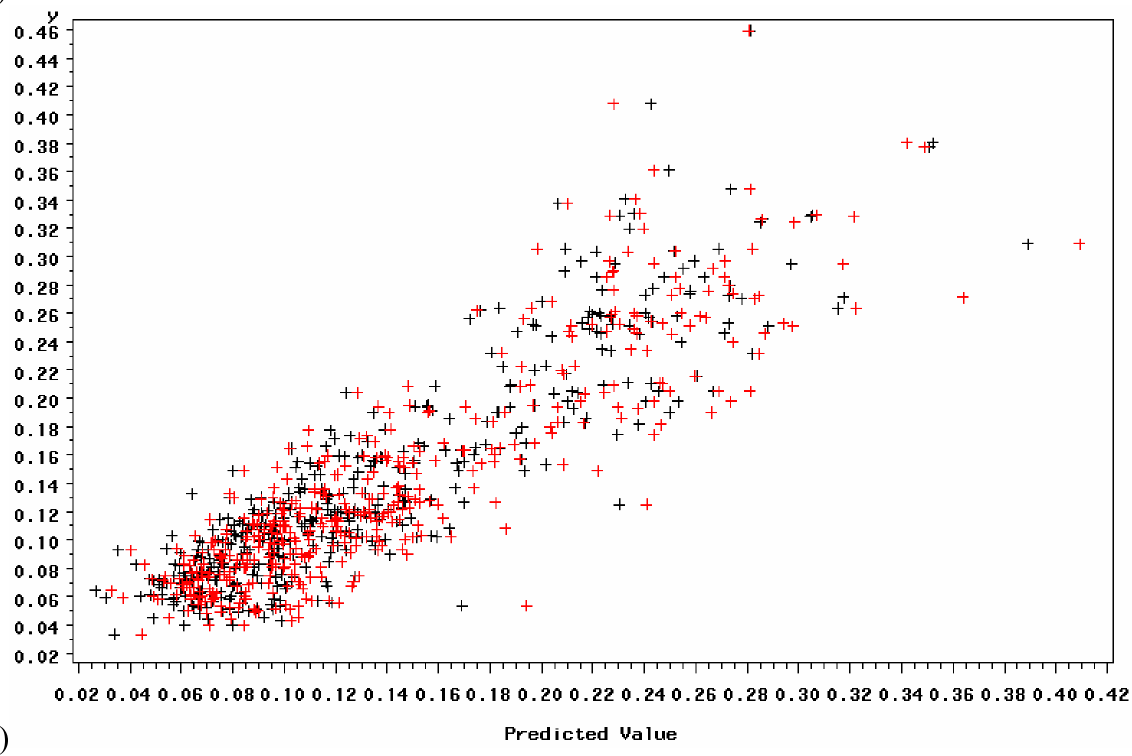


Figure 1 – Eigenvector variation for different coding schemes, the case of German unemployment



a)



b)

Figure 2 – Predicted and observed unemployment values: rook adjacency matrix (S-coding scheme), years 1996 (a) and 2002 (b): black crosses denote model-predicted values, and red crosses denote cross-validation-predicted values.

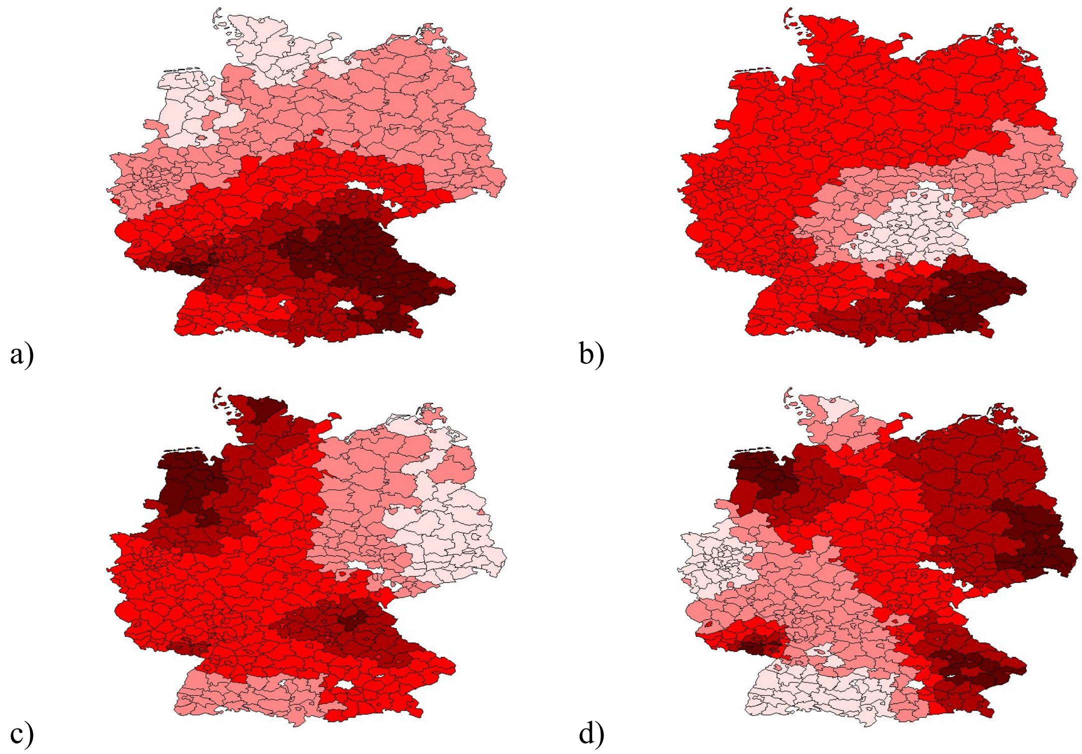


Figure 3 – Spatial filters computed for the rook matrix (S-coding): a) = E_2; b) = E_3; c) = E_5; and d) = E_6 (see Annex, Table A)

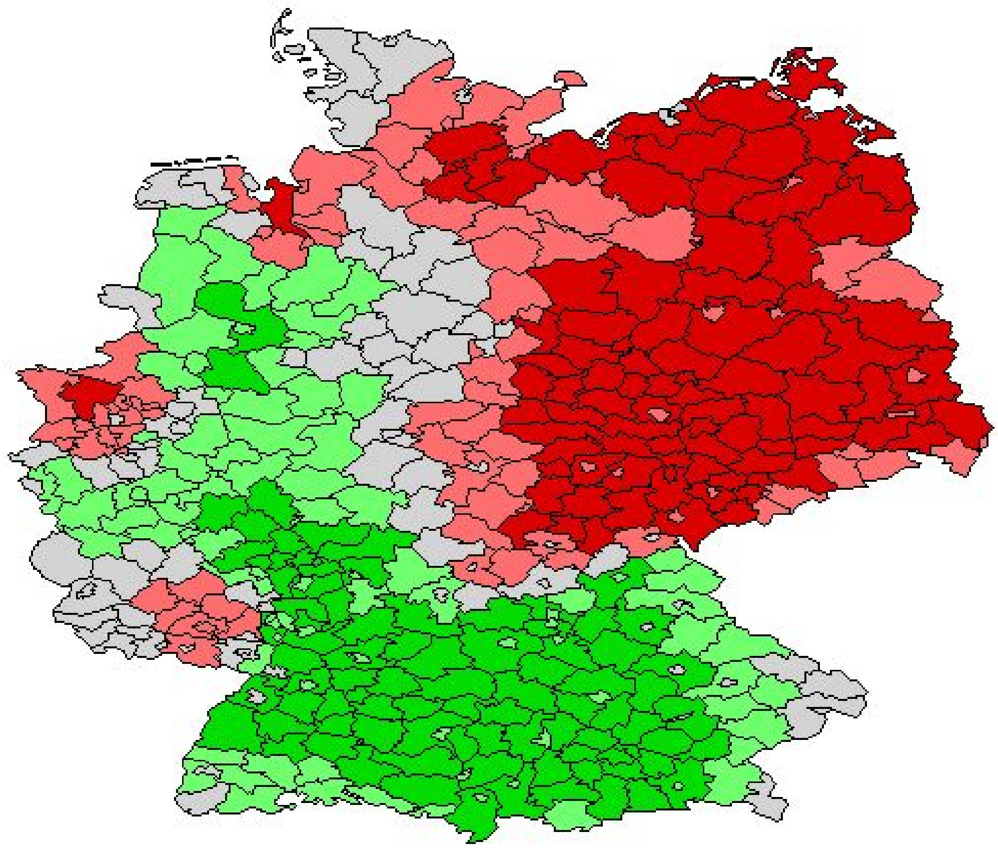
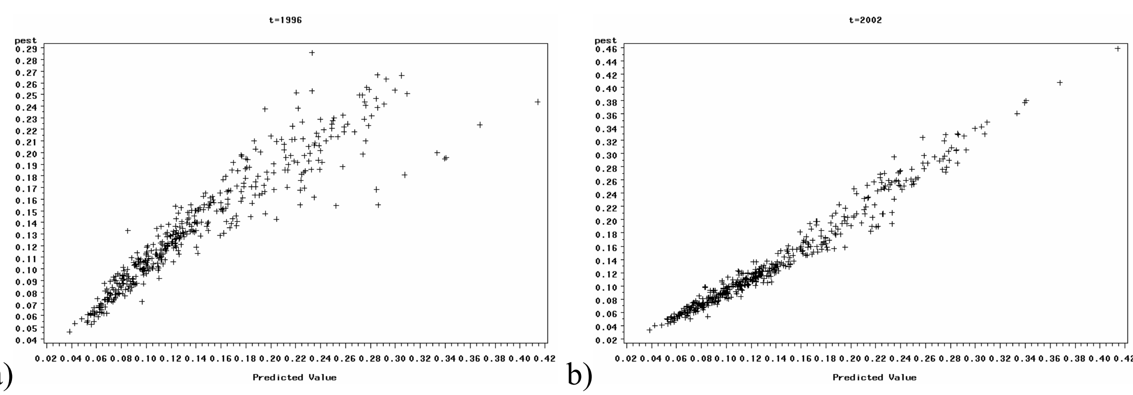


Figure 4 – Graphical visualization of the spatial filter obtained in the case of the rook geographic weights matrix (C-coding)



a) b)

Figure 5 – Real and predicted unemployment values (generalized linear mixed model): rook matrix (C-coding), years 1996 (a) and 2002 (b)