Forthcoming in: A. Frenkel, P. Nijkamp and P. McCann (eds)
Societies in Motion: Regional Development, Industrial Innovation and Spatial Mobility (tentative title). Edward Elgar, Cheltenham Northampton

11. Spatial Filtering Methods for Tracing Space-Time Developments in an Open Regional System: Experiments with German Unemployment Data

Roberto Patuelli, Daniel A. Griffith, Michael Tiefelsdorf, Peter Nijkamp

INTRODUCTION

Over the decades, economic location theory and modern economic geography have highlighted the relevance of space-time factors in (socio-)economic development. Spatial matters may be regarded as being of critical importance when investigating socio-economic (and other) phenomena (see, for example, Bockstael 1996; Weinhold 2002), including 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 (SAC) 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 SAC concept is a departure from the classical assumption of independence of observations constituting a single variable. SAC also complements the concept of temporal autocorrelation, which has been extensively studied and dealt with in time-series econometrics. SAC measures are used to quantify the nature and degree of the spatial correlation contained in data for a given variable, or to test the assumption of independence or randomness. From a statistical analysis viewpoint, spatial correlation patterns may become problematic, since they make standard statistics, such as correlation coefficients or ordinary least squares (OLS) estimates, potentially inappropriate.

This chapter 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 may have clear implications for policy evaluation and strategic planning. This chapter presents statistical analyses carried out by means of a semi-parametric 'spatial filtering' technique, described in Griffith (2003), which is based on the decomposition of spatial weights matrices. In our analysis, these matrices are defined for 439 German districts (i.e., Kreise), 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 used as a proxy for economic linkages.

Earlier work on pattern identification in German labour markets was carried out by, amongst others, Kosfeld and Dreger (2006), who investigated Verdoon and Okun's laws for German regional labour markets in the period 1992–2000. Their approach, however, 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, in addition to limiting ourselves to autoregression, 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 more refined level of disaggregation (439 districts versus 180 regions), which enables a more detailed analysis of the underlying spatial patterns.

SPATIAL FILTERING TECHNIQUES

Recent years have witnessed an increasing popularity for the use of spatial-econometric tools in regional research employing georeferenced data (see, e.g., Anselin 1988; Griffith 1988; Anselin et al. 2004; Anselin 2007). Among standard spatial econometric methods, spatial autoregression techniques have become powerful methods, in particular through the use of spatial weight matrices that mirror the intensity of spatial linkages (dependences) between spatially-referenced data. The insight has matured that—due to efficiency and normality problems—OLS may not be carried out with spatially dependent data. Furthermore, maximum likelihood (ML) estimators of spatial regression models are based on restrictive assumptions. These recognitions have led to the search for appropriate statistical estimation techniques. An alternative approach to spatial autoregression is the use of spatial filtering techniques, described inter alia in Griffith (1981), Haining (1991), Getis and Griffith (2002), and Tiefelsdorf and Griffith (2007). The advantage of these filtering procedures is that the variables studied (which, initially, are spatially correlated) can be decomposed into spatial and non-spatial components, which then can be used in an OLS or ML 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, 2004) is adopted in our study. This approach is preferred in our case study to the one by Getis (1990, 1995), which requires variables with a natural origin. This limitation would not allow us to apply the same method to the analysis of other labour market variables, such as employment growth rates.

The spatial filtering technique introduced by Griffith is based on the computational formula of Moran's I (MI) statistic[1]. This methodology utilizes eigenvector decomposition techniques, which extract *orthogonal* and *uncorrelated* numerical components from an n x n matrix (see, for details, Tiefelsdorf and Boots 1995)[2]. We extract the eigenvectors of the modified spatial weights matrix

$$(\mathbf{I}-\mathbf{1}\mathbf{1}^{\mathrm{T}}/n)\mathbf{W}(\mathbf{I}-\mathbf{1}\mathbf{1}^{\mathrm{T}}/n), \tag{11.1}$$

Where **I** is an identity matrix of dimension n x n, and **1** is an n x 1 vector containing ones. Such eigenvectors maximize the sequential residual MI values, so that the numerical values of the first computed eigenvector, E1, generate the largest MI value among all n eigenvectors. All subsequently extracted eigenvectors again maximize the MI values while being orthogonal and uncorrelated with the preceding ones, eventually giving the complete set of all possible (mutually) orthogonal and uncorrelated map patterns posited by the spatial weights matrix used (Getis and Griffith 2002).

When employed as regressors, these eigenvectors may function as proxies for missing explanatory variables. A smaller set of m ($< n$) 'candidate' eigenvectors can be selected for parsimony reasons, on the basis of their MI values. For example, an MI threshold value of 0.25 can be specified for selection screening purposes. The final set of eigenvectors describing the data at hand can be selected (out of the m candidates) by means of stepwise regression. In this regard, the orthogonality of the eigenvectors makes selection easier, because no partial correlations exist between the eigenvectors. The linear combination of the eigenvectors belonging to the final selection can be defined as the 'spatial filter' for the variable examined.

The eigenvector components may be regarded as independent map patterns, and represent the latent SAC of a georeferenced variable, according to a given spatial weights matrix. They also can be interpreted as redundant information due to spatial interdependencies, in the framework of standard regression methods. 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 weights matrix. While points (a) and (b) are discussed subsequently in this chapter, the latter point is briefly addressed now.

The previously discussed spatial filters are computed on the basis of a modified spatial 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 spatial weights matrices can be found in the literature (Tiefelsdorf et al. 1999; Getis and Aldstadt 2004). The main feature 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 2007, with practical details in Chun et al. 2005):

$$\mathbf{V}_{[q]} = \frac{n}{\sum_{i=1}^{n} d_i^{q+1}} \cdot \mathbf{D}^q \cdot \mathbf{B}, \tag{11.2}$$

where **B** is a binary spatial weights matrix, and $\mathbf{D}^q$ is a diagonal matrix that contains d_i^q components $(d_1^q, \ldots, d_n^q)$, belonging to vector $\mathbf{d} = \mathbf{B} \cdot \mathbf{1}$, 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 is commonly 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 is mostly used in autoregressive response (AR) and simultaneous spatial autoregressive (SAR) model specifications, and, contrary to the C-coding scheme, tends to emphasize the weight of objects with small spatial linkages.

Different individual spatial patterns may result from the calculation of the eigenvectors of the different 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 11.1 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.

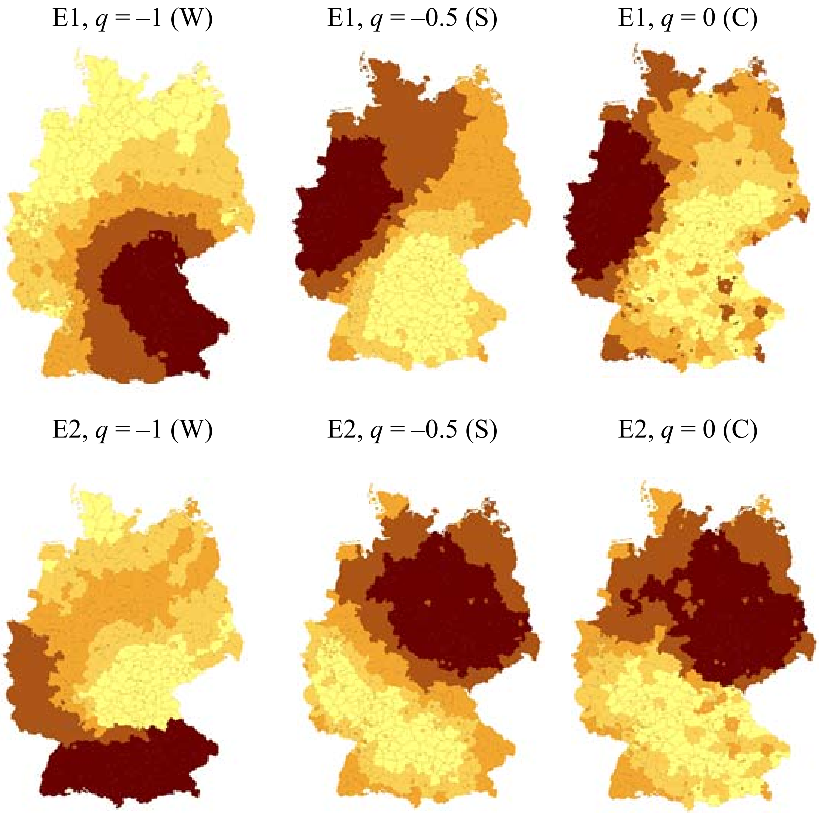


Figure 11.1 Eigenvector variation for different coding schemes, the case of German unemployment

The choice of coding scheme, and therefore of the spatial weights matrix, although yielding a different set of eigenvectors from which the spatial filters are selected, still produces a linear combination account for SAC in a spatial econometric or spatial statistics context; the differentiating feature tends to be parsimony. In the empirical application presented in this chapter, both W-coding and C-coding are employed (see Section 3). Results of a correlation analysis of the spatial weights matrices used also are presented, in order to compare the various approaches.

Next, the preceding spatial filtering techniques are illustrated empirically using German unemployment data. The dataset consists of cross-sectional data about the number of employed and unemployed individuals, collected by the (German) Federal Employment Services (*Bundesanstalt für Arbeit*, BA), for 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 dataset is NUTS-3. 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, viz. German commuting flows, is employed in our analysis. The data comprise, for each couple (i, j) of NUTS-3 origins and destinations, 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 chapter refer to the year 2002, and are employed in the computation of an 'economic flows' spatial weights matrix (see Section 3). Commuting data for one year only are employed in our case study, because varying commuting data would generate different spatial weights matrices, and, consequently, different sets of eigenvectors (although techniques exist to handle a set of such matrices; e.g., Rogerson and Plane, 1984). Furthermore, one can assume some spatio-temporal persistence with respect to the local commuting patterns. The daily commuting flows between two districts are transformed to satisfy the statistical symmetry requirement of spatial link matrices. This transformation represents the daily to-work and back-to-home flows.

COMPUTATION AND CHOICE OF SPATIAL FILTERS FOR GERMAN UNEMPLOYMENT DATA

Spatial 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 spatial weights matrix. Therefore, in addition to matrix computation methods, carefully considering the concept of proximity employed, and its consequences, is important. In our case study, we present a set of different definitions of the spatial weights matrix:

- *economic flows*: based on patterns of symmetrized commuting flows;
- *shared boundaries*: based on geographical contiguity, which by definition is symmetric; and,
- *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 district centroids define proximity in terms of geographical distance decay relationships.

A total of five spatial weights matrices are employed. 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 above.
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 not 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(SIM)[3]; the variables used for the estimation of the model are district employment data (discussed above), and the distance between the centroids of each district:
   - First, the distance decay exponent of –2.7 is taken from the estimated SIM, and then converted to the W-coding scheme ;
   - 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 SIM used to describe flows and estimate distance decay parameters:

$$F_{ij} = \kappa W_i^{\alpha} J_j^{\beta} e^{-\gamma d_{ij}} + \varepsilon_{ij}, \tag{11.2}$$

where: F_{ij} is the quantity of flows between areal units i and j; W_i is the number of workers residing in origin areal unit i; J_j is the number of jobs located in destination areal unit j; $\kappa, \alpha, \beta \text{ and } \gamma$ are parameters; and ε_{ij} is a random error associated with flows between origin i and destination j. The distance decay parameter, $\hat{\gamma}$, estimated with the log-linear form of Eq. (11.2)—but with a multiplicative rather than an additive error structure—was used to define the W-coding scheme, resulting in the matrix components being defined as:

$$w_{ij} = e^{-\hat{\gamma} d_{ij}} \Big/ \sum_{j=1}^{n} e^{-\hat{\gamma} d_{ij}} \ . \tag{11.3}$$

Next, $\hat{\gamma}$ was incrementally increased until the resulting ($\mathbf{W}^T$ + $\mathbf{W}$)/2 matrix yielded the same number of candidate 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.

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 a spatial 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 about regional unemployment. A minimum MI/max(MI) value of 0.25 has been used in our case study to identify the candidate set to be evaluated for inclusion. The results of this process, carried out for the matrices presented in the preceding section, are presented in Table 11.1.

Table 11.1 Candidate eigenvectors selected and maximum MI values

Spatial weights matrix	No. 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 both 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 logistic regression model, in order to reduce over-correction for SAC. Marginal eigenvectors were excluded as long as their χ^2 values remained non-significant at the 5% level.

The same process was repeated for all years of available data—from 1996 to 2002—and for each spatial 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. Results of the preceding analyses are summarized in Table 11.2, while details about the eigenvectors selected in each context and year are shown in the Appendix (Table A). In Table A, in all cases, the sum-of-squared prediction error (SSPE) divided by the mean squared error (MSE) is roughly 1 (that is, $\sqrt{\mathrm{SSPE}/\mathrm{MSE}} \approx 1$); in other words, a jack-knife 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 11.2 Amount of variance explained by the selected eigenvectors, and the number of common eigenvectors, 1996–2002

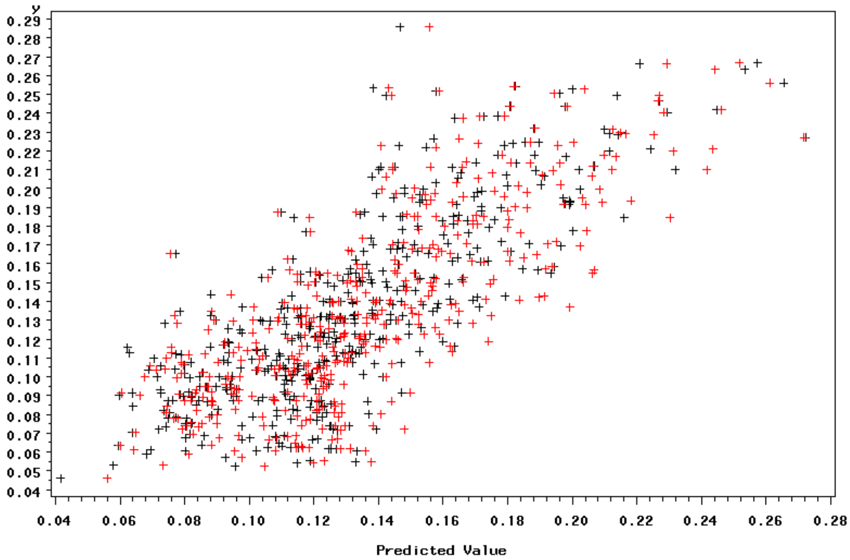
Spatial Weights Matrix	No. of Common Eigenvectors	Adjusted pseudo-R 2						
		1996	1997	1998	1998	2000	2001	2002
Journey-to-Work Flows Matrix	1 4	0.300 4	0.29 11	0.330 5	0.314 2	0.337 9	0.345 3	0.328 5
Rook Matrix (S-coding)	1 7	0.647 7	0.68 21	0.729 3	0.745 3	0.794 5	0.802 2	0.790 9
Rook Matrix (C-coding)	1 5	0.592 9	0.64 25	0.684 6	0.706 8	0.748 3	0.768 3	0.754 9
Distance-based Matrix ($\beta = -2.7$)	6	0.621 5	0.59 68	0.651 9	0.693 0	0.729 6	0.744 8	0.738 2
Distance-based Matrix ($\beta = -6.3$)	1 1	0.623 3	0.60 67	0.650 1	0.681 8	0.724 7	0.744 2	0.733 1

The results summarized in Table 11.2 show that we found sets of eigenvectors (whose linear combinations produce 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 of 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 weights 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 11.2. The 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.

a)



b)

Figure 11.2 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.

As mentioned in the section of Spatial Filtering Techniques, 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 11.3 shows the four eigenvectors with the largest MI values computed for the rook adjacency matrix S-coding scheme, and that are common to all the years examined.

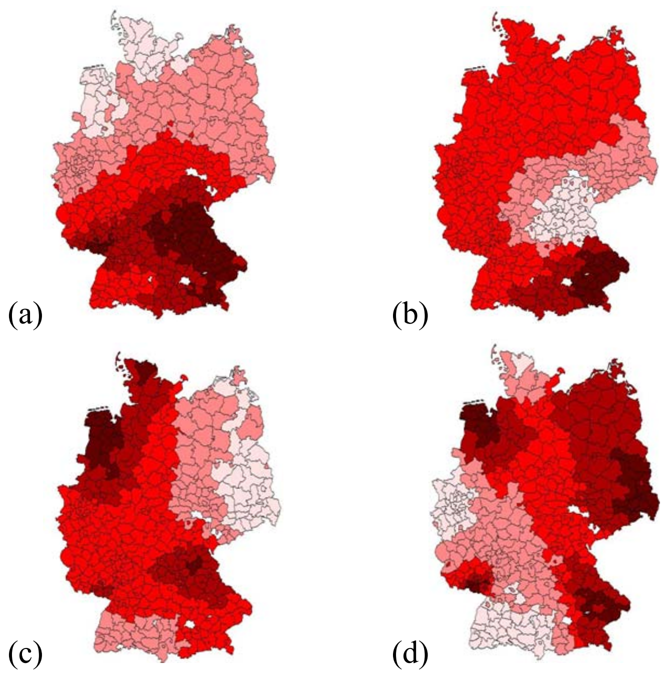


Figure 11.3 Eigenvectors computed for the rook matrix (S-coding): a) E2; b) E3; c) E5; and d) E6 (see Appendix, Table A)

As noted previously, the first two eigenvectors for adjacency matrices usually show East-West and North-South patterns. Spatial filter (a) (E2) in Figure 11.3 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', and then as 'local'. Although they may contain some common map patterns (for example, North-South and East-West patterns), spatial filters computed with different spatial weights matrices will vary to some degree. Meanwhile, an assessment of the statistical significance of the spatial filters (shown in Table 11.2) enables us to assess the utility of the different proximity approaches employed.

Synthesis: Results for Different Proximity Approaches

The preceding section reveals that all of the definitions employed in this chapter 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 spatial weights matrices, we need to compare the matrices themselves. Therefore, a correlation analysis of the matrices employed in our chapter has been carried out. Results of this analysis appear in Table 11.3 (for details about the computation of correlation between matrices, see Oden 1984, and Tiefelsdorf 2000).

Table 11.3 Correlations of Spatial 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 ($\beta = -6.3$)	0.5949	0.7923	0.6879	0.8775	1.0000

Several features of Table 11.3 are noteworthy. The most conspicuous result pertains to the low correlations between the journey-to-work flows matrix and the remaining matrices (that is, shared boundaries and distance-

based). The 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 closeness, 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 the similarities seen in the statistical performance of their computed eigenvectors (see Table 11.2).
Finally, we also note 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 the one 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 chapter; an interesting treatment can be found in Tiefelsdorf et al. (1999).

CONCLUSIONS

Our study aimed to map out German regional unemployment patterns by means of spatial filtering techniques, so as to uncover spatial structures underlying the georeferenced unemployment data. Several definitions of proximity have been employed, in order to operationalize spatial linkages according to geographic and non-geographic criteria. Each of these definitions has yielded a set of time-stable spatial filters, though at different levels of statistical significance.

Initial sets of eigenvectors have been selected on the basis of the SAC they accounted for (that is, by decreasing MI values), only later to be reduced in size by means of stepwise regressions 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 displayed by unemployment when utilized as the sole regressors in a logistic regression model. But the 'economic flows' approach, based on a journey-to-work flows matrix, fails to produce the same encouraging results. This finding might be due to the artificial nature of the data used (logical connections between districts), and to the lack of a more proper measure of regional economic linkages.

A correlation analysis of the spatial weights matrices employed in our analysis (see Synthesis: Results for Different Proximity Approaches) 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.

Future research will start from this preliminary estimation in order to carry out more detailed experiments about the dynamics of unemployment patterns. However, the joint employment of spatial filters and other explanatory variables involves further attention to spatial filtering, since eigenvectors that are significant both to the explained and (an) explanatory variable(s) imply filtering also of the latter. This research challenge may be addressed in the framework of new dynamic modelling experiments.

APPENDICES DETAILED EIGENVECTOR SELECTIONS, BY YEAR

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

Year	No. of eigenvecs	Year-specific eigenvectors			Common eigenvectors			Scale	Adj. pseudo R^2	$\sqrt{SSPE/S}$
		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	1.055
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

Eigenvectors extracted from the rook matrix (S-coding) (130 candidate eigenvectors)										
1996	23	E1	E24, E25, E60	E113, E124	E2, E3, E5, E6, E7, E8, E9, E10, E11	E15, E16, E22, E39, E41, E52, E71	E130	20.8775	0.64766	1.0226
1997	24		E17, E25, E28, E70, E82, E97	E113				23.3176	0.68207	1.0395
1998	20		E14, E25, E28, E36, E60, E70, E82	E113, E129				22.0736	0.72933	1.0476
1999	26		E14, E23, E36, E38, E70, E82	E113, E115, E129				22.0632	0.74533	1.0438
2000	31		E14, E25, E28, E33, E36, E38, E40, E50, E70, E82, E85	E113, E115, E129				22.0026	0.79449	1.0702
2001	28		E14, E18, E23, E32, E36, E38, E40, E82	E110, E115, E129				23.8654	0.80218	1.0489
2002	25		E14, E23, E36, E38, E40, E82	E115, E129				24.8114	0.79092	1.0387

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

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			24.406	0.693	1.004
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			26.871	0.7382	1.0101

Eigenvectors extracted from the distance-based matrix (β = –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	1.046
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

NOTES

1. Moran's I coefficient is defined 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 $\mathbf{W}$. Positive values $[I > -1/(n-1)]$ imply that geographical proximity tends to produce similar values of the variable examined.
2. Griffith's spatial filtering techniques may be compared to principal components analysis (CPA), since 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.
3. For details about the estimation of SIMs, see, among others, Sen and Smith (1995), and Haynes and Fotheringham (1984).

REFERENCES

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

Anselin L. (2007) Spatial Econometrics in RSUE: Retrospect and Prospect. *Regional Science and Urban Economics* 37 (4), 450-6

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

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

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

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

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

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

Griffith D.A. (1981) Towards a Theory of Spatial Statistics: A Rejoinder. *Geographical Analysis* 13, 91-3

Griffith D.A. (1988) *Advanced Spatial Statistics*. Kluwer Academic Publishers: Dordrecht

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

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

Griffith D.A. (2004) A Spatial Filtering Specification for the Autologistic Model. *Environment and Planning A* 36 (10), 1791-811

Haining R. (1991) Bivariate Correlation and Spatial Data. *Geographical Analysis* 23, 210-27

Haynes K.E. and A.S. Fotheringham (1984) *Gravity and Spatial Interaction Models*. Sage Publications: Beverly Hills

Kosfeld R. and C. Dreger (2006) Thresholds for Employment and Unemployment. A Spatial Analysis of German Regional Labour Markets 1999-2000. *Papers in Regional Science* 85 (4), 523-42

Lacombe D.J. (2004) Does Econometric Methodology Matter? An Analysis of Public Policy Using Spatial Econometric Techniques. *Geographical Analysis* 36 (2), 105-18

Oden N.L. (1984) Assessing the Significance of a Spatial Correlogram. *Geographical Analysis* 16, 1-16

Sen A. and T.E. Smith (1995) *Gravity Models of Spatial Interaction Behavior*. Springer: Heidelberg and New York

Tiefelsdorf M. (2000) *Modelling Spatial Processes - The Identification of Spatial Relationships in Regression Residuals by Means of Moran's I*. Springer-Verlag: Berlin

Tiefelsdorf M. and B. Boots (1995) The Exact Distribution of Moran's *I*. *Environment and Planning A* 27, 985-99

Tiefelsdorf M. and D.A. Griffith (2007) Semiparametric Filtering of Spatial Autocorrelation: The Eigenvector Approach. *Environment and Planning A* 39 (5), 1193-221

Tiefelsdorf M., D.A. Griffith and B.N. Boots (1999) A Variance Stabilizing Coding Scheme for Spatial Link Matrices. *Environment and Planning A* 31, 165-80

Weinhold D. (2002) The Importance of Trade and Geography in the Pattern of Spatial Dependence of Growth Rates. *Review of Development Economics* 6 (3), 369-82

ACKNOWLEDGMENTS

The authors wish to thank Uwe Blien (IAB, Nuremberg, Germany) for providing the labour market dataset, as well as Günter Haag (STASA, Stuttgart, Germany) for the commuting flows dataset. The authors are grateful to participants of the 52nd Annual North American Meetings of the Regional Science Association International (Las Vegas, NV) and the International Workshop on Spatial Econometrics and Statistics (Rome), and to an anonymous referee, for useful comments on earlier versions of the chapter.