Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 27 (2015) 106 – 111

Procedia
Environmental Sciences

Spatial Statistics 2015: Emerging Patterns - Part 2

Spatial diffusion and spatial statistics: revisting Hägerstrand's study of innovation diffusion

Roger Bivand[a,b,*]

[a] Norwegian School of Economics, 5045 Bergen, Norway
[b] Adam Mickiewicz University, 61-680 Poznań, Poland

Abstract

Torsten Hägerstrand's 1953 study of innovation diffusion [1] was pathbreaking in many ways. It was based on an explicit micro-model of information spread, and on Monte Carlo simulation of the hypothesised spatial process. Using the original aggregated data and Hope-type tests of the ability of the simulations to capture the observed adoptions, **(author?)** [2] and **(author?)** [3] and others found problems. This study attempts to examine the extent to which we may be able to "do better" with a range of approaches drawn from spatial statistics, including using a SAR lattice model, geostatistical modelling, Moran eigenvectors, and other approaches.

Keywords: Spatial diffusion, Spatial processes, Spatial statistics

1. Introduction

Torsten Hägerstrand played an important role in the promotion of mathematical geography, both through his pioneering research and through active recruitment of guest scholars, both bringing foreigners to Sweden, and sending his own students abroad. This nicely mirrors his own work on the diffusion of innovations [1, 4], in which he hypothesises that farmers are more likely to adopt innovations if they are in close proximity to earlier adopters. Initial and subsequent adopters were recorded in 125 5km square grid cells around the settlement of Asby for 1929–1932, together with all potential adopters who could be entitled to receive a subsity for pasture improvement, as shown in Figure 1.

1.1. Mean information field

The model of spatial interaction fitted in [1], p. 246, was calibrated from numbers of telephone calls and measured distances for logged telephone calls from each local exchange to destinations up to 50km:

$$\log F_I = 0.7966 - 1.585 \log d \tag{1}$$

where d is distance measured in km [4, 2]. Note that this relationship is isotropic. Using the same coefficient estimates, we can also reconstruct the MIF. We follow [1] by creating a 25×25 grid of one km squares to

*Corresponding author. Tel.: +47-55959355.
Email address: `Roger.Bivand@nhh.no` ()

1878-0296 © 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of Spatial Statistics 2015: Emerging Patterns committee
doi:10.1016/j.proenv.2015.07.103

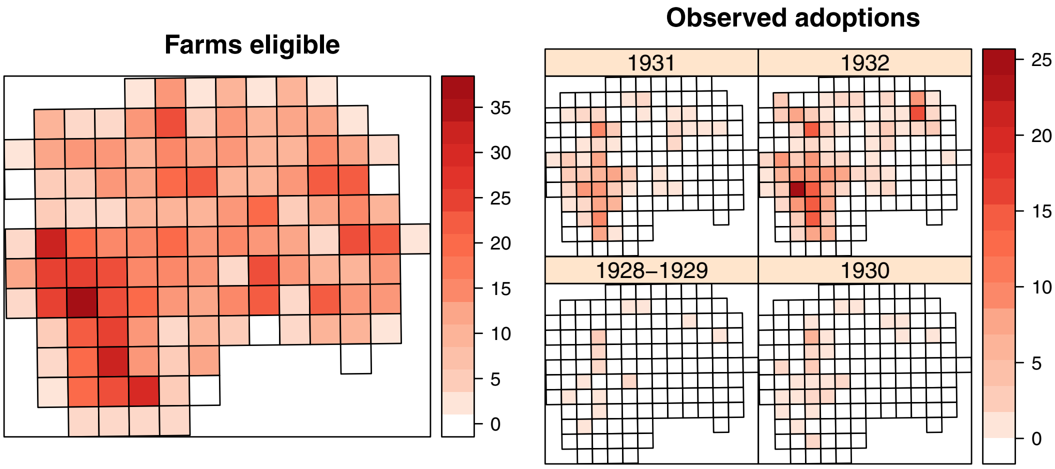
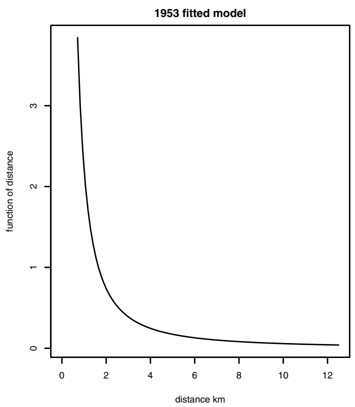


Fig. 1. The left panel shows counts of entitled farms in 5km grid squares; the right panel shows counts of adopters 1928–1932.



Original MIF

0.0096	0.0140	0.0168	0.0140	0.0096
0.0140	0.0301	0.0547	0.0301	0.0140
0.0168	0.0547	0.4431	0.0547	0.0168
0.0140	0.0301	0.0547	0.0301	0.0140
0.0096	0.0140	0.0168	0.0140	0.0096

Generated MIF

0.017	0.029	0.034	0.029	0.017
0.029	0.051	0.066	0.051	0.029
0.034	0.066	0.094	0.066	0.034
0.029	0.051	0.066	0.051	0.029
0.017	0.029	0.034	0.029	0.017

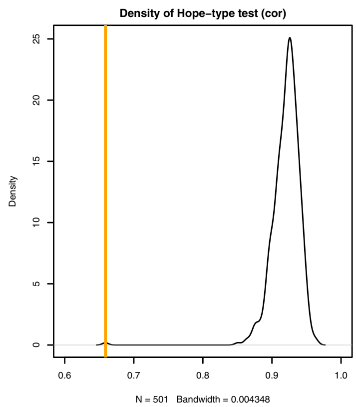


Fig. 2. The left panel shows the calibrated curve relating interaction frequency to distance; the central panel shows the original and reconstructed mean information fields; the right panel shows a density plot of the correlations between 500 MIF-simulated diffusion patterns and the mean simulation; the orange vertical line shows the correlation between the observed 1932 adoptions and the mean simulation.

generate the predicted interactions, which were then summed to 5km squares, and an (unknown) arbitrary value entered in the central cell. From this we can create a MIF summing to unity; Figure 2 shows the calibrated distance function and original and reconstructed mean information fields in the left and central panels.

The mean information field provides a view of the expected covariation between grid cells. Better, it has a clear behavioural motivation in the underlying relationship between contacts generating information spillovers and distance. However, reviews including [2] and [5], extended in [3], suggest that this micro-model is not fully successful when compared with the data.

1.2. Hope-type test

Using a preliminary version of the Hägerstrand simulation model, but simulating up to actual annual adoption counts as proposed by [6] and used in [7], we can conduct a Hope-type test as suggested in [2] and [3] with the Pearson correlation coefficient between the mean map of simulations and observed for Hägerstrand's MIF. As Figure 2 shows, it is very unlikely that the original (or equivalently the reconstructed) MIF could have generated the observed diffusion pattern.

2. Alternatives

2.1. Distance-based measures

Using a field representation of space, we can try to capture the nature of spatial dependence directly by comparing the differences in pairs of residual values for each pair of cell centroids, and plot a summary measure of the squared differences against distance in a variogram. The response is a log-transformed rate; we will model adoptions directly later. Following [8], we fit an exponential variogram model to a log transformed adoption rate for 1932, with only the intercept in the mean model. Figure 3 shows that this approach is hardly better than the micro-based MIF on the basis of a Hope-type test, but that the MIF implied by the variogram model is much less clustered in the central cell.

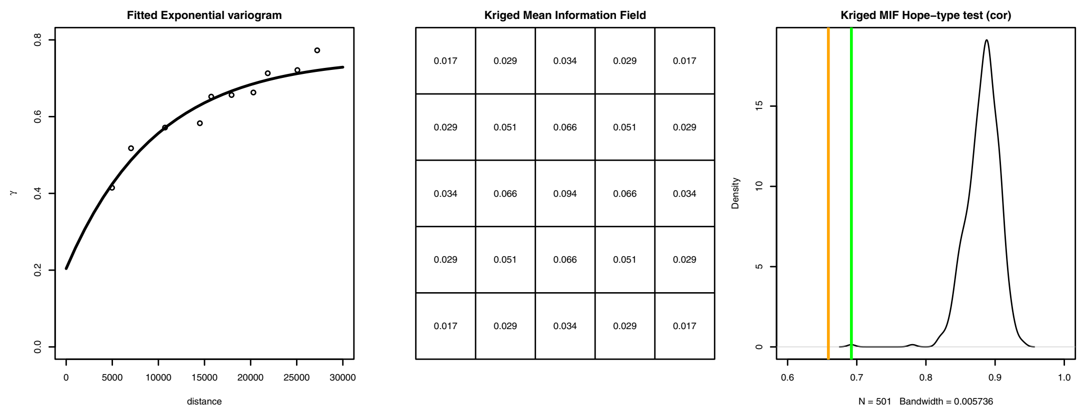


Fig. 3. Left panel shows the fitted Exponential variogram for the log transformed adoption rate for 1932 (support grid square centres); the centre panel shows the MIF derived from the fitted variogram; the right panel shows the Hope-type test of the kriged MIF (original MIF in orange).

2.2. Contiguity-based measures

We can create the simultaneous autoregressive (SAR) covariance structure from the basic objects, ρ and $\mathbf{W}$ (here binary rook contiguity weights). We have not attempted to accommodate heterogenerity through for example case weights, simply inserting σ^2 as fitted. This covariance matrix can the be split into two

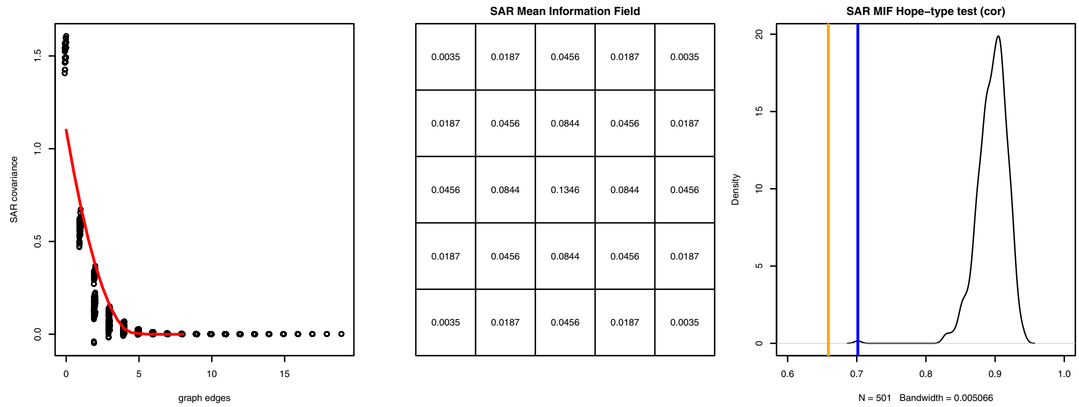


Fig. 4. Left panel shows the fitted Loess curve for SAR covariances for the log transformed adoption rate for 1932 (rook contiguities); the centre panel shows the MIF derived from the fitted SAR covariances; the right panel shows the Hope-type test of the SAR MIF (original MIF in orange).

triangular matrices by Cholesky decomposition to use in inducing autocorrelation; this approach (see [9]) differs from that used by [3].

Based on the variogram of the log-transformed adoption rate, the shape of the implied spatial process is fairly intuitive, with strong mutual dependency at short distances falling at longer distances. Wall [8] questioned whether the shape of the implied process in lattice representations was as intuitive. We use distance measured in numbers of graph edges to be traversed between cells, up to a maximum order of 19, and plot the corresponding values of the Cholesky decomposition of the SAR covariance matrix. We use a Loess curve to show the shape of the process relationship, and predict using this fitted curve to extract the MIF. Again, Figure 4 shows a much "flatter" MIF for the SAR model, but still a performance that is little better than the original case.

2.3. *Poisson contiguity-based measures*

A major weakness of the empirical approach used here the use of the fitted variogram and the spatial weights matrix to retreive MIF spatial process operators is the use of a transformed rate. We should perhaps actually be modelling the rate directly, using for example generalised linear models and their spatial extensions. Let us try out a Spatial Filtering GLM, and using the INLA "slm" model, to fit Poisson models with intercept-only mean models and spatial components.

Using Spatial Filtering (Moran Eigenvectors, ME) as a heuristic, we augment the very simple aspatial Poisson rates model (log link) with the subset of the eigenvectors of the doubly-centred matrix derived from the mean model and the spatial weights matrix that remove residual autocorrelation; the chosen eigenvector maps are shown in Figure 5. When we update the model by including the selected eigenvectors, the fit improved markedly.

Integrated Nested Laplace Approximation permits the use of Bayesian inference on marginal posterior distributions in a flexible framework, and may be estimated rapidly as MCMC sampling is not required for models implemented in the R INLA package. Figure 6 shows the predictions made by the approaches used compared with observed 1932 adoption counts.

3. Conclusions

Already in the early 1970s, doubts had been raised about a specific micro-model used to predict spatial interaction. Both variograms and parsimonious spatial weights matrices can be used to attempt to retreive spatial processes from data. These rely on relevant matches in support between the variables involved —

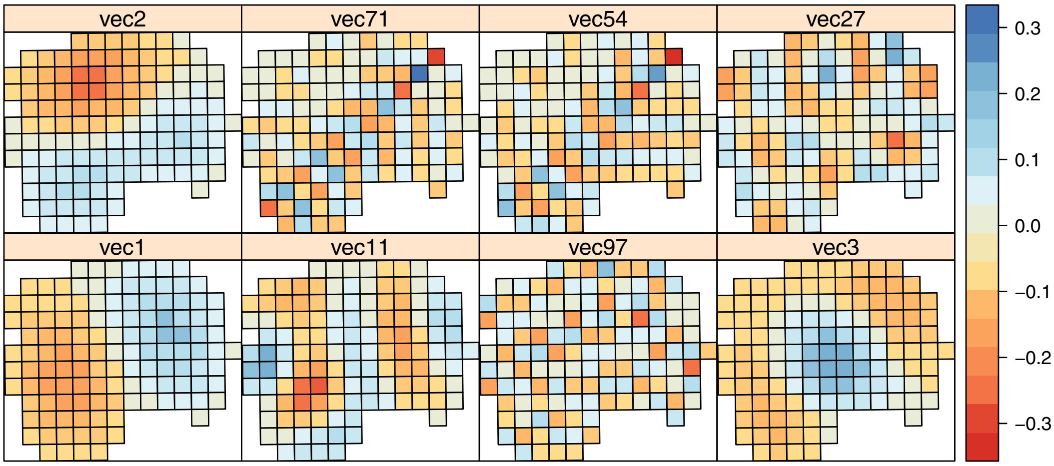


Fig. 5. The eight Moran eigenvectors selected to represent the spatial diffusion process for adoptions in 1932 given the entitled farms.

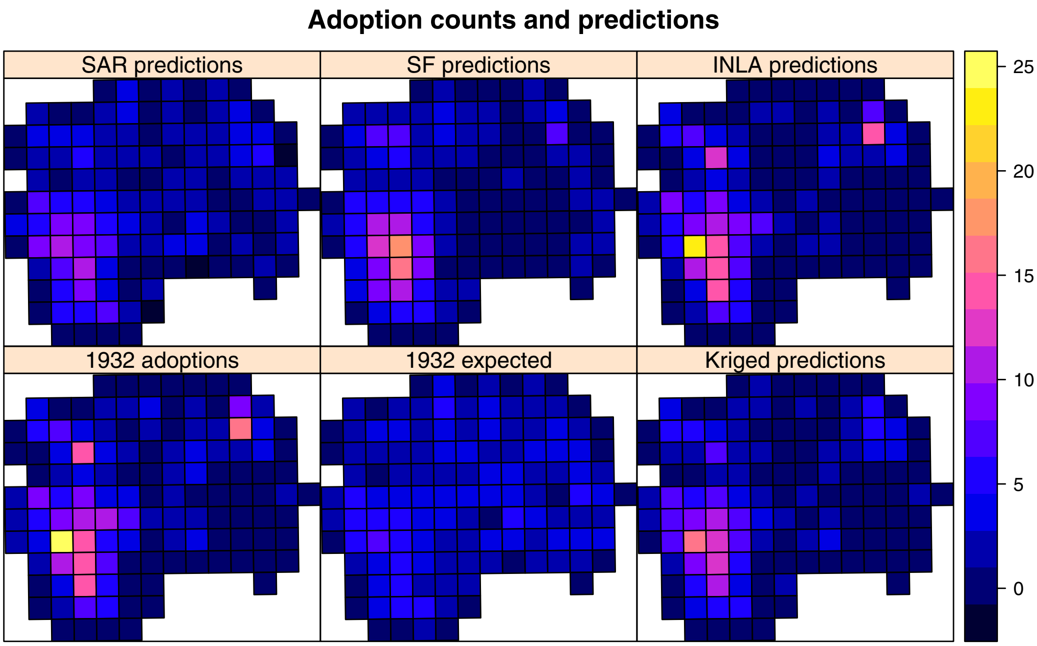


Fig. 6. 1932 adoptions, expected adoptions given only entitled farm counts, and predictions using kriging with fitted Exponential variogram and fitted spatial autoregresive model (backtransformed), and spatial Poisson regression using Spatial Filtering, and INLA.

contiguities between inappropriate regions are obviously of little value. Had we had access to the point locations of the entitled farms and adopters, more progress may be possible.

References

[1] T. Hägerstrand, Innovationsforloppet ur Korologisk Synspunkt, Gleerup, Lund, 1953.
[2] A. D. Cliff, Computing the spatial correspondence between geographical patterns, Transactions of the Institute of British Geographers 50 (1970) 143–154.
[3] A. D. Cliff, J. K. Ord, Spatial Autocorrelation, Pion, London, 1973.
[4] T. Hägerstrand, On monte carlo simulation of diffusion, in: W. L. Garrison (Ed.), Quantitative Geography: Economic and Cultural Topics, no. 13 in Studies in Geography, Northwestern University, 1967, pp. 1–32.
[5] R. Tinline, Linear operators in diffusion research, in: M. Chisholm, A. E. Frey, P. Haggett (Eds.), Regional Forecasting: Proceedings of the Twenty-Second Symposium of the Colston Research Society, Butterworth, London, 1971, pp. 71–81.
[6] R. S. Bivand, Nowe podejście do problemu przekazu informacji w modelach dyfuzji, Sprawozdania PTPN, Wydział Matematyczno-Przyrodniczy 96 (1980) 20–23.
[7] Z. Kamiński, Räumliche simulation in der diffusionsforschung, Erdkunde 42 (1988) 225–233.
[8] M. M. Wall, A close look at the spatial structure implied by the CAR and SAR models, Journal of Statistical Planning and Inference 121 (2004) 311–324.
[9] R. P. Haining, Spatial data analysis in the social and environmental sciences, Cambridge University Press, Cambridge, 1990.