Spatial and spatio-temporal studies on pathogens

Egyetemi doktori (PhD) értekezés

Solymosi Norbert

Témavezető:
Dr. Rózsa Lajos DSc.

Debreceni Egyetem
Természettudományi Doktori Tanács
Juhász-Nagy Pál Doktori Iskola

Debrecen, 2009.

Ezen értekezést a Debreceni Egyetem Természettudományi Doktori Tanács Juhász-Nagy Pál Doktori Iskola Kvantitatív és teresztris ökológia programja keretében készítettem a Debreceni Egyetem természettudományi doktori (PhD) fokozatának elnyerése céljából.

Debrecen, 200... .....

Solymosi Norbert

Tanúsítom, hogy Solymosi Norbert doktorjelölt 2009-ben a fent megnevezett Doktori Iskola Kvantitatív és teresztris ökológia programjának keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult. Az értekezés elfogadását javasolom.

Debrecen, 200... .....

Dr. Rózsa Lajos

Spatial and spatio-temporal studies on pathogens

Értekezés a doktori (Ph.D.) fokozat megszerzése érdekében
a környezettudományok tudományágban

Írta: Solymosi Norbert állatorvos

Készült a Debreceni Egyetem Juhász-Nagy Pál doktori iskolája
(Kvantitatív és terresztris ökológia programja) keretében

Témavezető: Dr. Rózsa Lajos DSc.

A doktori szigorlati bizottság:

elnök: Dr. .....

tagok: Dr. .....

Dr. .....

A doktori szigorlat időpontja: 200... ....

Az értekezés bírálói:

Dr. .....

Dr. .....

Dr. .....

A bírálóbizottság:

elnök: Dr. .....

tagok: Dr. .....

Dr. .....

Dr. .....

Dr. .....

Az értekezés védésének időpontja: 200... ....

Contents

List of Acronyms

AD Aujeszky's disease

ADO ActiveX Data Objects

ADV Aujeszky's disease virus

BUGS Bayesian inference Using Gibbs Sampling

BYM Besag, York and Mollié modell

CAR Conditional Autoregressive Modell

CRU TS Gridded monthly meteorological parameter time-series for period 1901-2000

CSV Comma Separated Values file

EOV Uniform National Projection (Hungarian)

ESRI Environmental Systems Research Institute

GCM Global Climate Modell, Global Circulation Modell

GHG GreenHouse Gas

GIS Geographic Information System

GPS Global Positioning System

HOT Highly Optimized Tolerance

IPCC Intergovermental Panel on Climate Change

IPL Inverse Power Law

MC Monte Carlo method

MCMC Markov-Chain Monte Carlo method

MSDE Microsoft Data Engine

OR Odds Ratio

PRUDENCE Prediction of Regional scenarios and Uncertainties for Defining EuropeaN Climate change risks and Effects

RR Relative Risk

SOC Self-Organized Criticality

SRES Special Report on Emissions Scenarios

TYN SC Tyndall climate SCenario dataset

WGS84 World Geodetic System version 84

1 Introduction

For years, my main topic is to study spatial aspects of various phenomens. Mostly I worked on veterinary epidemiology (Földvári et al., 2007; Allepuz et al., 2009; Farkas et al., 2009), but sometimes I was involved into human health related (Galamb et al., 2008a,b; Herszényi et al., 2008; Sipos et al., 2008; Szőke et al., 2008, 2009a,b) or social science (Mészáros et al., 2004; Solymosi et al., 2005; Mészáros et al., 2006, 2007) studies.

In this dissertation I collected some of my works in connection with veterinary spatial, spatio-temporal epidemiology. With our collegues we have been working on the epidemiology of rabies of red foxes in Hungary. From this work one paper was published until now, in which we have analysed the temporal patterns of disease occurence (Harnos et al., 2006b). In spatial epidemiology of rabies we had a few conference presentations, from these some interesting results are included in this dissertation (Solymosi et al., 2002, 2003). We had some interesting experiences using Moran's I method in analyses of spatial clustering of rabies cases. These prompted us to deal with the null distribution of the test statistic (Solymosi et al., 2004b). In spatial epidemiology the usage of the Bayesian methods is increasing. In rabies epidemiology we used Bayesian models as well. To make easier the spatial Bayesian modelling I am developing a tool for preparation of data and maps to help the user of GeoBUGS (Lawson, 2008; Wagner et al., 2009; Allepuz et al., 2009).

At the end of the nineties by †*István Medveczky* I was invited into the analysis of the results of countrywide Aujeszky's disease eradication program. At county level we published an evironmental association paper from this disease (Solymosi et al., 2004a).

Since 2007 I have been working on studying the possible associations of animal health and climate change (Harnos et al., 2008). In this work we are working on the heat stress modelling of cattle based on observed meteorological (Reiczigel et al., 2008, 2009) data and climate scenarios (Solymosi et al., 2007, 2008b). Based on climate and other environmental (e.g. landcover) datasets researchers can model environmental associations and/or predict spatial distribution of diseases (Solymosi, 2008; Erdélyi et al., 2009). Due to climate change expectedly new vector-borne diseases may appear in temperate zone. For modelling the spatial pattern changes in disease distributions due to global changes we are studying the environmental similarity of dif-

ferent sites of Earth. In this work we try to find sites on the Globe where the climate in the past is close to the future climate in Hungary. Identifying similar areas we could predict for which pathogens will be suitable the climate of Hungary. For this modelling we need climate datasets. Some of that kind of datasets are available, but sometimes it is very difficult to use them. As a first step of this work I have developed a tool (Solymosi et al., 2008a) to make easier the usage of one of climate datasets (Tyndall Centre for Climate Change Research produced one).

1.1 Rabies in Hungary

As a zoonosis and an incurable disease, rabies has always been given due respect both in human and veterinary medicine. In Hungary up to 1954 there were just the urban rabies, spread by dogs. The sylvatic type rabies – spread by red foxes (*Vulpes vulpes*) – appeared in 1954 coming from the North–East. The epidemic propagated at a speed of 50–60 km/year until rabies became an endemic disease in 1970 (Mocsári et al., 1994). To decrease the risk of infection an oral vaccination campaign was started in Hungary in 1992 to immunize the foxes. Figure 1.1(a) shows the location of rabies cases in Hungary in 1992. In the spreading of the infection by foxes there are two natural borders in Hungary: the Danube and the Tisza River. There is low possibility the foxes go through these rivers. By this cogitation these different regions are handled in the immunization as independent areas. First the oral vaccination was started in the region beyond the Danube. The oral vaccination campaign of the red foxes had more steps. The first vaccination campaign was carried out in October 1992 covering an area of 5000 km^2. In April 1993, this area was expanded to 6000 km^2 (Nagy et al., 1995). The vaccinated area was increased in April 1995 to 7000 km^2, in October 1995 up to 10000 km^2, in April 1996 to 15000 km^2 (Kerekes, 1996). Since October 1996 up to 2000 the whole western part of the country (38000 km^2) was involved to the immunization (Kerekes, 2000). Figure 1.1(b) shows rabies cases in 1998, after the immunization in Transdanubia. Since 2000, the oral immunization process is being done in the central part of the country. By this time the systematic immunization in the western part of country was stopped, and it follows that way if there is a case then around that there is vaccination.

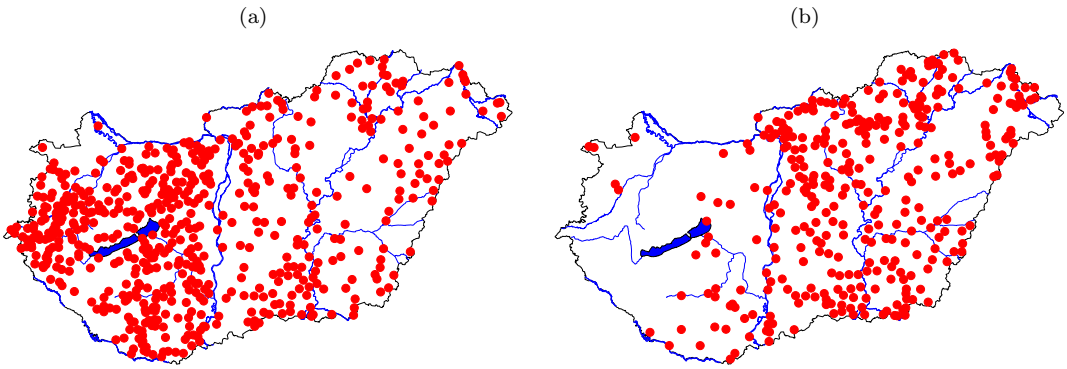


Figure 1.1: The locations of reported rabies cases in Hungary in 1992 (a) and 1998 (b).

According to our goal to understand the epidemiological properties of rabies in Hungary, to analyse the effectiveness of eradication program we performed analyses on the temporal, spatial and spatio-temporal properties of reported rabies cases in Hungary (Bánhidy et al., 2002; Harnos et al., 2002; Solymosi et al., 2002, 2003; Harnos et al., 2006a,b).

In one of our studies (Harnos et al., 2006b) the goal was to analyse the fox rabies cases time series from 1967 to 2001 and to identify its main characteristics. We were particularly interested in the changes due to the oral immunization campaign and motivated to show that it not only affects the average number of cases but also the way the number of cases fluctuate. We showed that the distribution of extremely large fluctuations develops an inverse power law (IPL) tail, a characteristic feature in recent epidemiological models. IPL has been proposed by (Rhodes & Anderson, 1996a,b; Rhodes et al., 1997, 1998) to describe time series from small and large vaccinated human populations. Trottier & Philippe (2005) applied this model for measles, rubella, pertussis and mumps outbreaks in Canada. IPL is a function with no characteristic scale ad self-similar upon rescaling (scale invariance), and it may fit to the extreme values of the distribution (power law tail). This can explain the co-existence of small and very large epidemics. Large outbreaks are expected from this type of distributions (Philippe, 2000). The goal of the study on temporal properties was: (1) identification of the long time trends and the seasonal variations in the time series; (2) analysis and comparison of the extreme fluctuations in the periods before and after immunization; (3) explanation of the findings in terms of Highly Optimized Tolerance (HOT) introduced by Doyle & Carlson (2000).

Beside the temporal trends we were interested in spatial pattern of rabies occurences. In epidemiological pattern analysis one of the main investigation areas is the cluster analysis in space and/or time. In epidemiology cluster has more definitions, one of them as it is used in this dissertation: *„cluster is an unusual aggregation, real or perceived, of health events that are grouped together in time and/or space„* (Centers for Disease Control and Prevention, 1990).

Numerous statistical methods were developed to analyse the clustering in epidemiology. The methods may be categorized according to different viewpoints. According to dimension we may distinguish *spatial*, *temporal* and *spatio-temporal*. According to spatial reference unit they can be *point*, *line* or *area* based methods. With *global* spatial clustering methods we can answer the question: Is there clustering in the observed disease occurence over the studied region or not. Global methods can't identify the position of clusters. For this purpose one can use *local* spatial clustering methods. A special case of local methods is *focused* clustering, when aggregation of occurences are examined in the environment of supposed sources of danger.

Some clustering methods applied more frequently in epidemiology.

Spatial:

Global: *BB, BW* (Gebhardt, 1998), *Besag & Newell* (Besag & Newell, 1991), *Cuzick & Edward's* (Cuzick & Edwards, 1990); *Geary's c* (Geary, 1954); *Moran's I* (Moran, 1950; Oden, 1995); *Ohno's method* (Ohno et al., 1979; Ohno & Aoki, 1981); *Potthoff-Whittinghill's statistic* (Potthoff & Whittinghill, 1966a,b); *Ripley's K-function* (Ripley, 1976, 1977, 1981); *Rogerson's CuSum* (Rogerson, 1997); *Tango's method* (Tango, 1995).

Local: *Openshaw's Geographical Analysis Machine* (Openshaw et al., 1987); *Turnbull's Cluster Evaluation Permutation Procedure* (Turnbull et al., 1990); *Kulldorff's Spatial Scan statistic* (Kulldorff & Nagarwall, 1995); *Local Moran's I* (Anselin, 1992); *Local Geary's c*

Temporal: *cusum* (Ewan & Kemp, 1960; Healy, 1968; Bjerkdal & Bakketeig, 1975; Kennett & Pollack, 1983); *Ederer-Myers-Mantel* (Ederer et al., 1964); *negative binomial* (Hill et al., 1968, 1976); *Poisson* (Flynt, 1974); *scan* (Naus, 1965, 1966; Weinstock, 1981; Wallenstein,

1980; Wallenstein & Neff, 1987; Turnbull et al., 1990); *sets* (Chen, 1979, 1986); *two-stage* (Hardy et al., 1990).

Spatio-temporal: *Barton* (Barton et al., 1965); *kth nearest neighbour test* (Jacquez, 1996); *Knox* (Knox, 1964); *Mantel* (Mantel, 1967); *space-time k-function* (Diggle et al., 1995).

In the Transtisza region between 1990-2001 there was no vaccination program for wild foxes. We have analysed the spatial clustering of yearly occurences there during that interval with different global, area based culstering methods (Solymosi et al., 2003). This study was a starting point of our analysis of spatial patterns of rabies cases in Hungary. In epidemiology it is not well published which conditions should come true to use the certain method. Besides the spatial clustering of rabies we were interested in the properties of applied methods (Black-black, Geary's c, Moran's I) as well. In that study we also examined the associations of statistic of used clustering methods, case number and number of affected areas. The results of this study initiated the examination of the null distribution of Moran's I.

Spatio-temporal clusters were also investigated using Knox method. In this study we were interested in how change clustering depending on case number and region of Hungary.

The changes of overall and local relative risk (RR) of occurence of rabies were also analysed. Time trend of relative risk was also studied in spatial context. In these analyses Bayesian spatial, spatio-temporal models were used.

1.2 Null distribution of Moran's I based on stratified MC

Moran's I is a frequently used statistic for quantifying spatial clustering. Moran's I measures the spatial autocorrelation of the variable of interest with respect to the neighbourhood structure of the spatial regions to be analyzed. It serves as a basis of local as well as global tests (Anselin, 1992; Carpenter, 2001; Dale et al., 2002).

The variable of interest may vary from observed frequencies (Zhang & Selinus, 1997; McGrath & Zharig, 2003; Overmars et al., 2003; Shaoxiang et al., 2003; Weishampel et al., 2003) to rates (Oden et al., 1996; Rosenberg et al., 1999; Pfeiffer, 2000; Wakefield et al., 2000; Carpenter, 2001).

The null hypothesis of Moran's test is lack of autocorrelation. Traditionally the null distribution of I is determined on the basis of distributional assumptions and asymptotic normal theory. Another method to determine the null distribution, which is regarded more reliable in practice, is based on a permutation argument relying on the assumption of exchangeability under the null hypothesis. It simply means that under the null hypothesis permuting the values of the variable of interest does not affect the distribution of I. Under this assumption, the null distribution can be determined either theoretically or by Monte Carlo simulation, generating a large number of random permutations of data. If exchangeability does not hold, the p-value obtained from the test may be invalid, i.e., the test may falsely detect significant autocorrelation or clustering (the actual Type I error rate may be higher than the nominal one). It is often quite clear how exchangeability depends on the choice of the variable to analyse. If incidence is known to be proportional to population size and population size varies from region to region, then exchangeability requires the use of rate (incidence per 10000 inhabitants) rather than raw incidence counts. This is the case for many chronic diseases. For a contagious disease, however, the assumption of proportionality may not hold. Dependence of incidence on population density may not be linear, that is, a two times bigger population density may result in an incidence more than twice as much. In this case, the use of rate does not ensure exchangeability.

Researchers typically do not search for spatial clustering due to known factors like population size or age distribution. They would rather eliminate the effects of those explanatory variables to avoid results which trivially mirror the pattern of a known factor, like e.g. the clustering of urban-rural areas in the country. So the natural null hypothesis sounds like lack of autocorrelation after controlling for the uninteresting or known explanatory variables. Sometimes it is rather puzzling how to eliminate the effects of some factors or how to adjust or standardize observed values to ensure exchangeability. For example, if the distribution of traffic accidents is analysed, the basis for such a standardization may be population size, number of cars in use, total length of roads, total length travelled by all cars in use, etc. It is hard to decide theoretically which is the right one, or which is the best one with respect to the exchangeability assumption.

Here we propose a method for the case in which the variable of interest is in a presumably monotone but unknown relationship to an observed covariate. (If we know the type of relation-

ship, e.g. quadratic, logarithmic, etc., then the best to do is to standardize the values by this function.)

In our example, the size of the fox population is unknown but there is presumably a monotone increasing relationship between the size of region and the size of fox population. Moreover, even if the population size were known, the linear relationship between population size and case count were not easily justifiable because rabies is a contagious disease, and the number of contacts may depend nonlinearly on the population size.

1.3 Environmental association studies

Spatial risk assessment of herd sero-status of Aujeszky's disease

The Aujeszky's disease virus (ADV) belongs to the family of Herpesviridae which establish a lifelong persistent infection with intermittent shedding of the virus. The infection can cause clinical disease with neurological symptoms in suckling and weaned pigs and respiratory or reproductive disorders in grower-finisher and adult animals. Pigs are the only natural host of ADV. Other animals found on farms (such as rats and mice) are dead-end hosts (Berke & Grosse Beilage, 2003). The main route of transmission of ADV among swine populations is the physical contact between infected and susceptible animals. Less important for the transmission are fomites such as feed, vehicles or artificial insemination. Other routes of infection of susceptible populations are long-distance airborne transmission (Gloster et al., 1984; Christensen et al., 1990, 1993; Casal et al., 1997) and vector-mediated transmission (Wright & Thawley, 1980; Kirkpatrick et al., 1980; Medveczky et al., 1988). The distances among swine populations and the surrounding topographical features might be important for the last two routes. The health status of the pig population and the success in eradication of Aujeszky's disease (AD) are influenced by the geographic population density (Weigel et al., 1991; Austin & Weigel, 1992; Austin et al., 1993; Siegel et al., 1993; Leontides et al., 1994; Stegeman et al., 1995; Norman et al., 1996; Boelaert et al., 1999; Siegel & Weigel, 1999; Rodríguez-Buenfil et al., 2002; Tamba et al., 2002). Nevertheless, the effect of the geographical features on the spread of infection has been a rarely investigated topic. Marsh et al. (1991) investigated the distance to the nearest county road, highway, quarantined herd, river or lake and the density of swine herds within

a 5 km radius as possible risk factors; high density of pig herds within 5 km was found as a possible risk-increasing factor whilst lakes or rivers within 1 km seemed protective. The distance between pig units and geographical characteristics of the region can influence the effectiveness of AD eradication programs (Scheidt et al., 1991).

Our aim was to investigate the association between sero-status of pig units and presence of certain geographical features as well as large-scale swine units and villages (including small towns) in their neighborhood.

TETYN: tool for extracting climatic parameters from Tyndall data sets

In recent years the importance of climate change has been rapidly growing in most disciplines, including agriculture, health, ecology, economy and sociology. In human and veterinary epidemiology it is an emerging topic to study the changes in spatial distribution of vector-borne diseases due to global changes (Wittmann & Baylis, 2000; Purse et al., 2005; Casimiro et al., 2006; Gloster et al., 2007; Jaykus et al., 2008). Many publications deal with the necessity of adaptation to climate change, therefore researchers and decision makers need adequate data-sources to analyse, model or forecast consequences of these changes. Realising this, some institutions and organisations produced data-sets of a projected future climate.

One of these is the **P**rediction of **R**egional scenarios and **U**ncertainties for **D**efining **E**uropeaN **C**limate change risks and **E**ffects (PRUDENCE), which contains datasets from different models for the interval 2071-2100 (Christensen et al., 2007). Although this data-source is very useful for long-term modelling, it can not be used for near-term climate change based analyses.

For those who would work on near-term modelling, the Tyndall Centre for Climate Change Research produced two future climate data-sets the TYN SC 1.0 and the TYN SC 2.0 (Mitchell et al., 2004; Mitchell & Jones, 2005). They offer a dataset (CRU TS) from the past (1990-2003) based on observations.

Unfortunately the reading of the Tyndall datasets is not too easy for people without programming knowledge. In applied research most people needs data tables readable by convenient tools (e.g. R, MS Excel). Our **t**ool for **e**xtracting climatic parameters from **Tyn**dall datasets (TETYN) was developed with an aim to make it easier to extract certain parts of the Tyndall datasets.

The Tyndall Centre projected climate datasets contain precursors for the different parameters which are actually calculated from these precursors by an algorithm based on the equation published by Mitchell et al. (2004). The results of TETYN queries can be saved as comma separated values (CSV) or ESRI shapefiles. In both formats the records represent the gridboxes and columns contain the monthly data.

2 Material and methods

2.1 Data

Rabies database

Data have been collected from the rabies case registry of the Animal Health and Food Control Department of the Ministry of Agriculture. Data were provided in different formats so we needed different processing techniques. A database management software (Rabidb) was developed for the digitalisation of the hand written records of the 1990-1993 period. The 1994, 2000 and 2001 records were provided in a Microsoft Word document file and they could be directly imported into the database. The 1995-1997 records were provided as printouts so we digitalized them using the Microsoft Office XP Professional Edition Microsoft Office Document Scanning software and imported the recognised text into the database. The database had to be unified after the data entry process because the settlement registry code system differed before and after 1993. The next task was to connect the database of rabies cases with the GIS database. The resulting database contains the data of all documented rabies cases for the 1990-2001 period, including the location and date of the occurrence as well as the affected species. Cases are linked to the location of the nearest settlement on a digital map.

For time series analysis this detailed database was completed by count of monthly rabies cases in period 1958-1989. These data was avaiable at county level, and with this spatial reference were recorded into the database. We constructed monthly summerized time series for the period 1967-2001 for western (Transdanubia) and for eastern part (Eastern Hungary) of the country (Figure 2.1).

Both the Rabidib database management software and the RabiGIS integrated GIS database management software were developed with Visual Basic 6.0 Professional Edition. The database was developed using the Microsoft Office XP Professional Edition Microsoft Access software. We used MSDE 1.0 through ADO as a database motor. RabiGIS uses Mapinfo MapX 4.51 for processing operations on a 1:100.000 vector-graphic map of Hungary.

For the Bayesian relative risk estimations the database was reconstructed in PostgreSQL-PostGIS system. In this study the reported red fox rabies case numbers were projected to 150

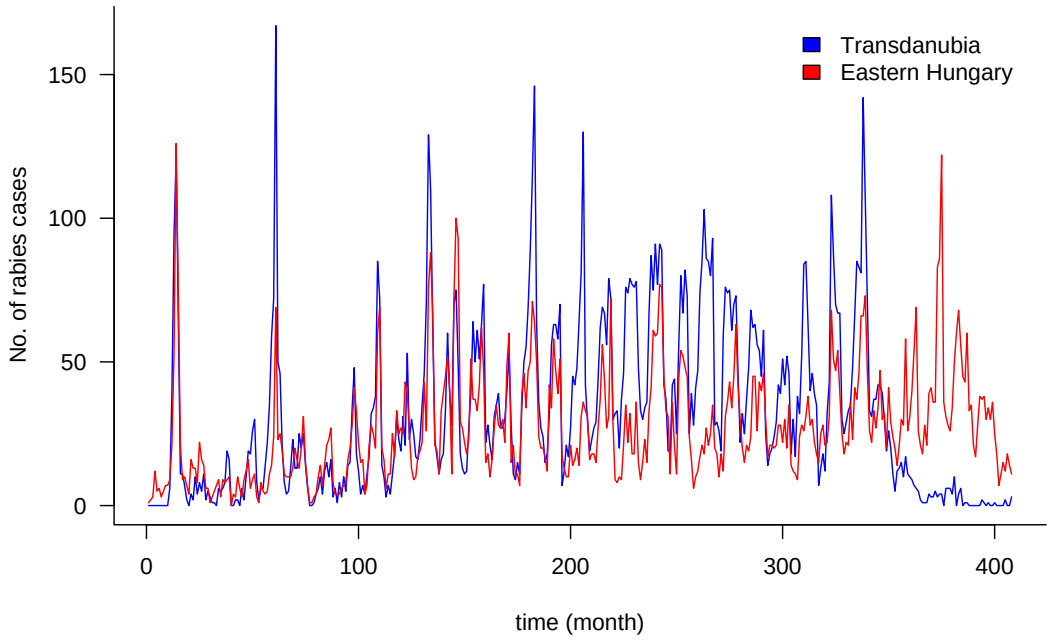


Figure 2.1: Monthly number of rabies cases in Transdanubia and in Eastern Hungary (1967-2001).

local administrative units in Hungary. For this districts the case numbers were summarized for each year. Between 1990 and 2001 for each year relative risk estimation was done. To analyze the time trend of the red fox rabies relative risk spatial–temporal analysis was done for the 1990–1996, and the 1997–2001 time periods.

Aujeszky's disease database

Data of all known pig units (except pure fattening units) registered in Csongrád county in 1998-2000 were used in the study. Csongrád county is located on the Great Plain, in the southeastern part of Hungary. The neighbouring counties are Bács-Kiskun, Békés, Jász-Nagykun-Szolnok, whereas it is bounded in the south by state border with Romania and Serbia-and-Montenegro. This is a rural area; the Tisza river divides it into two parts. Pig production is important in the economy of the county. Swine units are classified as large and small-scale units. Small-scale units represent traditional pig-keepers having a few pigs in the backyard; according to animal-health regulations, the upper limit is 100 pigs. Largescale units are located outside villages. The

number of units and villages is displayed in Table 2.1. The ADV sero-status is known for all units from the results of the yearly ADV serological screening of sows organized by the Animal Health and Food Inspection Service. The geographic coordinates of the large-scale pig units were determined by a Garmin eMap GPS-receiver. The determination of the coordinates was made from May to August 2000 with an accuracy of 10-15 m. The geographic coordinates were appointed in the WGS84 coordinate system (the available GPS-receiver worked with this one) and these data were transformed into the Uniform National Projection (EOV) system.[1] The precise locations of small-scale units within villages were not determined because we regarded each village as one epidemiological unit. The reason for this is that the small-scale units are close together and the hygiene barriers between them typically are limited. A village was regarded as infected if there was at least one infected unit within the village.

Table 2.1: Number of large- and small-scale pig units and villages in the study of Aujeszky's disease herd sero-status in a county in Hungary.

	1998	1999	2000
Number of villages in which there were some small-scale pig units (total number of small-scale units)	77 ([a])	60 ([a])	71 (5345)
Large-scale pig units	33	39	44
Percentage of large-scale units infected	30	26	18

[a] Total number of small-scale units not available.

Tyndall datasets

Future climate scenario datasets

The future climate scenarios are based on global climate (or general circulation) models (GCM) and future greenhouse gas (GHG) emission scenarios. GCMs describe the interactions of the components of the climate system: the atmosphere, the oceans, the terrestrial and marine biospheres, the cryosphere and the land surface. The different GCMs are varying in parameterization of physical processes. The Intergovermental Panel on Climate Change (IPCC) Special Report on Emissions Scenarios (SRES) provides a comprehensive set of 40 different scenarios for GHG emission sorted into four storylines (Nakicenovic et al., 2000). The four storylines represent different demographic, social, economic, technological and environmental developments.

[1] http://lazarus.elte.hu/gb/geodez/geodind.htm

Since storing the parameters for all scenarios and all time and space points in a pre-calculated form requires too much hard disc space, the developers organized the datasets into base files. From these files the user can calculate any of the five parameters. In this form the datasets need much less disc space. In the data-set there are six types of data files: key values of model-scenario combinations (scenario selector file), GCM patterns of change between 2070-2099 (response pattern files), time-series of global temperature change between 2001-2100 (global warming files), de-trended inter-annual variability from 1901-2000 (residual files), the 1961-1990 average climate data (climatology files), and the minimum and maximum permissible value (minimum and maximum files).

For the different GCMs (PCM, CGCM2, CSIRO2, HadCM3, ECHam4), emission scenarios (A1FI, A2, B2, B1) and climate parameters different files also contain the necessary data sources. Files are identified by the combinations of these three elements in their names. The five parameters (and their mark) are the temperature (TMP), precipitation (PRE), diurnal temperature range (DTR), vapour pressure (VAP) and cloud cover (CLD).

Structure of data files. The scenario selector file is an ASCII ordered list. The global warming files are ASCII columns with four line headers. The other file format is a standard grid with a five line header. The response pattern, the climatology, the minimum and maximum files contain two lines for every grid boxes. The grid identifier is stored in the first line (e.g. Grid-ref= 4,109). The second line contains twelve five-character width fields respectively for the months. In the residual files for every grid-box the first line is the same as in the other grids. The grid identifier line is followed by one hundred lines for the years between 2001-2100 respectively. Every year-line contains twelve five-character width fields for the months.

TYN SC 1.0

The TYN SC 1.0 contains data files for calculation of the five parameters for all possible combinations (16) of the four emission scenarios and four GCMs (PCM, CGCM2, CSIRO2, HadCM3). The dataset is on a 10 minute spatial resolution grid, which has 31,143 grid boxes. The spatial

range of the grid is between 11.00°-32.00° in longitude and 34.00°-72.00° in latitude. The dataset is free on request.[2]

The file name constructions are the following: `ateam.iavar.1901-2000.VAR` (residual files), `MODEL.SCEN.2080s.ateam.VAR` (response pattern files), `MODEL-SCEN.ann` (global warming files), `min.VAR.ann`, `max.VAR.ann` (minimum and maximum files), `obs.clim6190.ateam.VAR` (climatology files).

`VAR` identifies the parameter, the `MODEL` is the `GCM` and `SCEN` is the emission scenario. The scenario selector file is general (scenarios.txt).

TYN SC 2.0

The TYN SC 2.0 contains data files for calculation of the five parameters for all possible combinations (20) of the four emission scenarios and four GCMs (PCM, CGCM2, CSIRO2, HadCM3, ECHam4). The dataset is on a 30 minute spatial resolution grid over the globe (excluding Antarctica), which has 67,420 grid boxes. The dataset is free on request.[3] The file name constructions are the following: `iavar.1901-2000.VAR` (residual files), `hd.MODEL.SCEN.2080s.VAR` (response pattern files), `MODEL-SCEN.ann` (global warming files), `min.VAR.ann, max.VAR.ann` (minimum and maximum files), `obs.clim6190.globe.VAR` (climatology files).

Observed climate dataset

The CRU TS 2.0 contains data of the observed monthly meteorological data in the twentieth century containing nine parameters (with extension): cloud cover (`.cld`), diurnal temperature range (`.dtr`), ground frost frequency (`.frs`), precipitation (`.pre`), temperature (`.tmp`), minimum temperature (`.tmn`), maximum temperature (`.tmx`), vapour pressure (`.vap`), wet day frequency (`.wet`).

The latest dataset version (2.10) includes climatic data for the period 1901-2002. The spatial property of the observed data is the same as in TYN SC 2.0, 30 minute resolution grid over the globe (excluding Antarctica), with 67,420 grid boxes.

The format of files is a standard grid with a five line header. The grid identifier is stored in the first line. The grid identifier line is followed by 102 lines for the years between 1901-2002

[2] http://www.cru.uea.ac.uk/~timm/grid/TYN_SC_1_0.html
[3] http://www.cru.uea.ac.uk/~timm/grid/TYN_SC_2_0.html

respectively. Every year-line contains twelve five-character width fields for the months. The dataset is freely downloadable.[4]

2.2 Methods, software

Rabies

Power law tails and highly optimized tolerance in epidemiology

Distributions with power law tails arise in many natural and engineered systems. During the last three decades much effort has been made to explain their widespread occurrence. Systems influenced by many independent random factors in their environment would more likely behave randomly dictated by the law of large numbers and the central limit theorem. It is then always puzzling why such systems so often violate these simple concepts. To understand the mechanism behind power laws the key ideas came first from statistical physics. In conventional physics power laws arise at special points of the parameter space of systems. These are critical points. Systems at their critical points go over phase transitions and show very strong fluctuations obeying power laws. However, this mechanism does not provide a sufficient explanation since critical points are isolated in the parameter space. It is very unlikely to find a large number of systems in nature that are accidentally tuned to their critical points. The next step was to explain why systems naturally evolve into their critical points and stay there. Simple systems are not able to do this spontaneously. In complex systems this can happen more easily as they are able to adapt to external conditions via self-organization. At the beginning of the eighties the concept of self-organized criticality (SOC) emerged (Bak, 1996), where it was shown that many complex systems naturally seek their critical states and go into a self-organized critical state without external tuning of their parameters. This explanation can account for a very large class of natural systems showing power law distributed fluctuations.

During the last couple of years it became clear that in certain systems engineered or designed by humans or by evolutionary forces distributions with power law tails can arise from another reason. The theory of HOT (Doyle & Carlson, 2000) claims that in systems strongly optimized to avoid failures the size distribution of failures naturally develops power tails. The basic idea

[4] http://www.cru.uea.ac.uk/cru/data/hrg/cru_ts_2.10/data_all/

behind HOT theory can be best demonstrated by an idealized forest fire model. In this model one should find the optimal distribution of the trees on a grid so as to maximize tree harvest in the face of occasional fires that burn complete connected clusters of trees and are started by sparks that arrive with a given spatial distribution. It turned out that optimizing the harvest gives rise to a segmented forest consisting of contiguous patches of trees separated by firebreaks, and that the resulting distribution of fire sizes follows a power law. Newman et al. (2002) showed that the tail of the complementary cumulative distribution behaves as an IPL $F_c(l) \propto l^{-1-1/\beta}$ parameter β is intimately related to the geometry of the problem. For a d dimensional forest fire model with optimized barriers $\beta = d$ holds, resulting in an IPL exponent $b = -1 - 1/d$.

Spatial, spatio-temporal clustering

Spatial clustering methods. In spatial clustering analysis three area based global technics were used: Black-black (BB), Moran's I and Geary's c. All of them use a neighbouring matrice (W) for the calculation of statistic. This matrice descripts the spatial relation (w_{ij}) of i and j observational units, in our case the settlements or the local administrative units:

$$w_{ij} = \begin{cases} 1 & \text{if district } i \text{ and } j \text{ have a common boundary} \\ 0 & \text{if district } i \text{ and } j \text{ have no common boundary.} \end{cases}$$

BB statistic:

$$BB = \frac{1}{2} S_0 x_i x_j$$

$$S_0 = \sum_{i,j} w_{ij},$$

where x_i is the number of cases or a ratio of the i-th region. The observational units (polygons) may be black (infected) or white (uninfected). Fundamentaly the statistic compares the number of black-black connections to the number of all connections (Gebhardt, 1998). Test statistic is not standardized. If there is no black-black pair, then the value of it is 0. When the number of black regions is growing, then the value of statistic is also increasing. To test it, Monte-Carlo method was used.

The global form of Moran's I statistic:

$$I = \frac{n}{S_0 \sum_i z_i^2} S_0 z_i z_j$$

$$z_i = x_i - \bar{x},$$

wehere n is the number of districts (observational units). Moran's I typically ranges from -1 to 1. An uncorrelated process has an expected $I = 0$. Negative values of I indicate negative autocorrelation. Positive values indicate positive autocorrelation.

If the observations (x_i) are spatially independent normal random variables, the expected value of I is:

$$E(I) = -\frac{1}{n-1}$$

hence for large n the expected value is approximately 0. The variance of I under these conditions:

$$Var(I) = \frac{n^2 S_1 - nS_2 + 3S_0^2}{(n^2-1)S_0^2}.$$

The variance of I under the assumption of complete randomisation:

$$\begin{aligned} Var(I) &= \frac{1}{(n-1)(n-2)(n-3)S_0^2}\Big\{n[(n^2-3n+3)S_1 - nS_2 + 3S_0^2] \\ &\quad -\frac{n\sum z_i^4}{(\sum z_i^2)}[(n^2-n)S_1 - 2nS_2 + 6S_0^2]\Big\} \end{aligned}$$

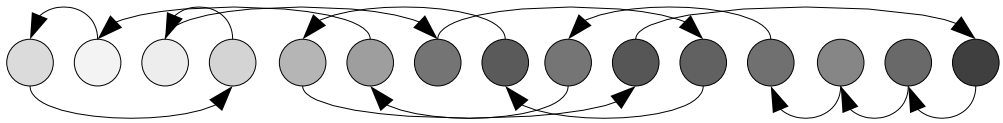
$$S_1 = \frac{1}{2}\sum_{i,j}(w_{ij} + w_{ji})^2$$

$$S_2 = \sum_i (w_{i.} + w_{.i})^2.$$

In other cases to test Monte-Carlo permutations are used. We proposed a stratified MC method to obtain the null distribution of statistic. Assume that we want to assess the null distribution by the permutation method, but exchangeability is suspect to be violated because our data contains a covariate (population size, region size or something similar) which is in a presumably monotone but unknown relationship to the variable of interest. If the relationship is known to be linear, then standardization by calculating rates is an appropriate solution to provide for exchangeability (Figure 2.2(a)).

Furthermore, if the relationship is nonlinear but known, a similar standardization by the known function offers a solution. Our proposal for the situation when the type of the relationship is unknown is to stratify according to the value of the covariate and to permute within strata (Figure 2.2(b)).

(a) Conventional Monte Carlo



(b) Stratified Monte Carlo

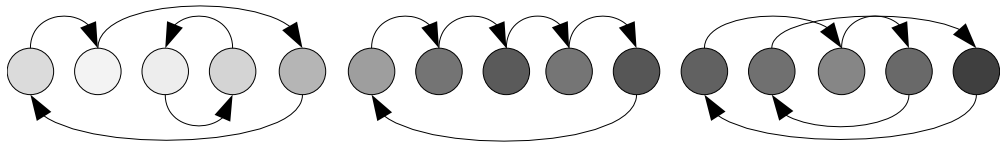


Figure 2.2: Permuting the whole dataset (a) or within strata (b).

The rationale behind this is that if the covariate is nearly constant, then its influence on the case counts diminishes. In our example this means that we build strata according to the size of region. Permuting within strata means that we permute the values belonging to big regions among each other, medium sized regions among each other, etc. The question arises how many strata should be built, and the pragmatic answer is, as many as results in the number of observations needed to permute within strata.

In the example, we built strata so that each stratum contains 30 observations. Of course it is generally not true that the covariate is constant in a stratum, so in this sense the method is not exact. But it may be approximately valid, and it is more robust than standardization by calculating rates (in our example: case count divided by the size of region) as it allows for a nonlinear relationship and it works without knowledge of the form of the relationship.

In a simulation experiment we generated random data for all regions of Hungary with mean values depending on the size of the region, but otherwise free from spatial autocorrelation (except for those due to dependence on size). The dependence was modelled as nonlinear, illustrating the situation related to a contagious disease, namely we used $Y_i \sim Poisson(\lambda_i)$ with $\lambda_i = c\sqrt{a_i}$, where Y_i is the case count in region i, a denotes the area of region i, and c a constant (same for all regions) to provide realistic case counts. We wanted to see what happens if (1) we analyse the raw case counts, (2) we standardize by the size of region (i.e. by a natural but not the right covariate), or (3) we apply the proposed stratified randomization. The table of the frequencies and the size of the districts was sorted by the area size. This new table was stratified to not

Table 2.2: Meaning of Geary's c and Moran's I statistic.

Spatial distribution	Geary's c	Moran's I
Clustered	$0 \leq c < 1$	$I > 0$
Random	$c = 1$	$I = 0$
Uniform	$1 < c < 3$	$I < 0$

bigger than 80 records parts. In this subsets we made 999 permutations using the `pps` package (Gambino, 2005). In every permutations we calculated the test statistic and recorded the result. In the end as the 1000^{th} row we inserted the Moran I result of the raw data. The final 1000 records were ordered by the Moran I, and the rank number of the raw test statistic was divided by the total length (1000) resulting in the significance level.

The global form of Geary c statistic:

$$c = \frac{n-1}{2S_0 \sum z_i^2} S_0 (x_i - x_j)^2$$

The test statistic of Geary c is between 0 and 3. Around 1 it means random spatial distribution. Under 1 indicate positive over 1 negative autocorrelation. Table 2.2 shows the contexts of Geary's c and Moran's I.

Time-space clustering. In veterinary epidemiology, methods of spatio-temporal analysis are used to acquire information about the characteristics of the observed disease patterns. In our study Knox's test was applied to detect clusters in space and time of rabies cases. The test statistic is

$$X = \sum_{i=1}^{n} \sum_{j=1}^{i-1} s_{ij} t_{ij},$$

where $s_{ij} = 1$ if the spatial distance between events i and j is less than the critical distance and 0 otherwise, and $t_{ij} = 1$ if the length of time between events i and j is less than the critical time and 0 otherwise. The number of event pairs in a data set containing n objects is $n * (n-1)/2$ (Ward & Carpenter, 2000). For testing χ^2 test or any randomisation procedure can be performed. In both cases the null hypothesis is that time distances between event pairs are independent of the spatial distances between them.

We used 4 km and 30 days as critical space and time distances. These critical distances were choosen arbitrary in a certain sense, but based on the home range size of red foxes (Smith &

Wilkinson, 2003; White, 2003; Moorcroft & Lewis, 2006) and the incubation time of disease (Blancou, 1988; Zienius et al., 2007) respectively.

Relative risk estimation

Besag, York and Mollié model. For modelling the relative risk of rabies the Besag, York and Mollié (BYM) model was used (Besag et al., 1991). In this model the area-specific random effects of relative risk comprise two components, the correlated (clustering) heterogeneity and the uncorrelated heterogeneity component (Lawson et al., 2003). The model is formulated as follows:

$$y_i \sim Poisson(e_i\theta_i),$$

where the $(y_i, ..., y_n)$ and $(e_i, ..., e_n)$ is the observed and expected number of the rabies cases in the i^{th} region, respectively. The expected rabies case number was calculated as proportional to the area of the local administrative districts.

$$log\ \theta_i = \alpha + u_i + v_i,$$

where α is the overall level of the relative risk, u_i is the clustering and v_i is the uncorrelated heterogeneity. As Besag et al. (1991) proposed, the clustering component is modelled as a spatial correlation structure with a conditional autoregressive (CAR) model

$$[u_i|u_j, i \neq j, \tau_u^2] \sim N(\bar{u}_i, \tau_i^2),$$

where

$$\bar{u}_i = \frac{1}{\sum_j w_{ij}} \sum_j u_j w_{ij}$$

$$\tau_i^2 = \frac{\tau_u^2}{\sum_j w_{ij}}.$$

The prior distribution model for the uncorrelated heterogeneity is

$$v_i \sim N(0, \tau_v^2).$$

To control the variability of v and u parameter τ_u^2 and τ_v^2 is used with gamma distribution (Bernardinelli et al., 1995a).

Space × time model. For modelling the temporal trend of the relative risk a spatio-temporal model was used (Bernardinelli et al., 1995b). In this model the area-specific intercept and the temporal trend are modelled as random effects. The temporal trends assumed to be linear. If y_{ik} is the count of disease and e_{ik} is the expected count in the i^{th} region and k^{th} time period. The model for the relative risk formulated as

$$log\ \theta_{ik} = \alpha + u_i + v_i + \beta \cdot t_k + \delta_i \cdot t_k,$$

where α is the overall rate, u_i and v_i are clustering and unstructured heterogeneity components (like in the BYM model), $\beta \cdot t_k$ is the linear trend term in time t_k, and δ_i random effect is a space–time interaction (Lawson et al., 2003).

Model fitting. The model fitting was carried out using two separate chains with different initial values in the spatial relative risk estimation and the spatio-temporal modeling as well. The convergence was checked by the multivariate version of Gelman and Rubin's diagnostic (Brooks & Gelman, 1998). Achieving the convergence the samples were discarded as a burn-in and further 15 000 iterations were run on each chain.

Software. To store the spatial and rabies data the PostgreSQL-PostGIS database system was used[5]. The Bayesian modelling was performed by using WinBUGS[6]. For preliminary mapping the maps were imported into WinBUGS from the PostgreSQL-PostGIS database with the help of maps2WinBUGS tool.[7] The final mapping was done using R-environment (R Development Core Team, 2009), the connection between R and PostgreSQL-PostGIS database was built by RpostGIS package (Solymosi et al., 2006). The WinBUGS Marcov Chain Monte Carlo (MCMC) simulations were called from R using the R2WinBUGS package (Sturtz et al., 2005). The MCMC output was analysed using the coda package (Plummer et al., 2007).

[5]http://postgis.refractions.net/documentation/
[6]http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml
[7]http://maps2winbugs.sourceforge.net/

Environmental association studies

Spatial risk assessment of herd sero-status of Aujeszky's disease

Software. A civil vector graphical digital map of Hungary (OTAB1) was used with a resolution of 1:100,000. The map contains the following layers: villages and towns, large-scale pig units, lakes, rivers, small natural waterways and artificial canals, forests, and four types of roads from unpaved road to controlled-access high-speed motorway. The projection system is EOV. This map was integrated into a database-management software (VetEpiGIS) developed by me. The software was built by using the Visual Basic 6.0 Professional Edition environment (Microsoft Corp.) with a MSDE 1.0 (Microsoft Corp.) as database engine through ADO and MapX 4.51 (MapInfo Corp.) for mapping.

Analysis. Buffer generation (Sanson et al., 1991; Norstrøm, 2001) was applied to define circular zones around large-scale swine units and to determine the presence of topographical features in these zones. The radius of the zone was increased from 1 to 10 km in 1-km steps as Casal et al. (1997) suggested. The feature (like, e.g. forest) was counted as being present in the zone if an instance of the feature had any presence with the buffer zone. Data of all known swine units from years 1998 to 2000 were analyzed. Because largescale units were the same in all 3 years (with a few exceptions due to reorganization, etc.), and sometimes neither their status nor the features in their neighbourhood changed, our data contained duplicate records. These duplicates were deleted from analysis (i.e. just one record of the identical cases was kept). However, if the status of the unit changed due to infection or eradication, and/or there were changes in the neighbourhood, we used data from both or all years. Data were analyzed by stepwise logistic regression using the backward conditional method by SPSS 8.0. The dependent variable was status (1 if seropositive, 0 if not). Explanatory variables were indicators of the topographical features, seropositive/seronegative large-scale units, and seropositive/seronegative villages in the zone around the unit (1 if feature/unit/village is present, 0 if not). No interaction term was included in the model. This analysis was repeated for each zone radius. The reported Odds Ratios (OR) were those obtained after removing the non-significant ($P > 0.05$, by Wald's test) factors from the model. Multicollinearity was checked calculating the variance inflation factors, which were < 1.5 (except for "seropositive village" and "seronegative village", for which it was

2.5 and 2.1 at zone radius of 4 and 5 km, respectively) indicating no serious collinearity. We also checked the correlations between the estimated regression coefficients. Highly correlated estimates were found only in case of the 4 km buffer zone for "lake" and "forest" ($r > 0.999$) therefore, we removed "lake" from the equation. For all other zone radii, all |r| were ≤ 0.53.

Extracting climatic parameters from Tyndall data sets

Parameter calculation. The following equation was used for the calculation of all parameters:

$$x_{vgsiym} = o_{vim} + o'_{viym} + (p_{vgsim} * t_{gsy}),$$

where x is the datum in any scenario (s), defined by the parameter (v), GMC (g), grid box (i) year (y) and month (m). The o is the observed climatology, o' the observed climatology, p is the normalised response pattern and t is the global warming. The algorithm for calculation of mean temperature in the grid box 4,109, in March 2010 based on the HadCM3 GCM and A1FI emission scenario using the TYN SC 1.0 dataset is presented in Table 2.3.

Table 2.3: The steps of calculation the monthly mean temperature in the grid box 4,109 based on the HadCM3 GCM and A1FI emission scenario using the TYN SC 1.0 dataset resulted 7.9 ℃ for March 2010.

```
GRIDREF = 4109
MODEL = HadCM3
SCENARIO = A1FI
YEAR = 2010
MONTH = 3
VAR = TMP

ovim = 71*0.1000                  <- defined by VAR, GRIDREF, MONTH
o'viym = 5*0.1000                 <- defined by VAR, GRIDREF, YEAR, MONTH
response_pattern = 20*0.1000      <- defined by MODEL, SCENARIO, VAR, GRIDREF, MONTH
amount_of_global_warming = 4.863  <- defined by MODEL, SCENARIO
tgsy = 0.785                      <- defined by MODEL, SCENARIO, YEAR

pvgsim = response_pattern / amount_of_global_warming = 0.411
xvgsiym = 7.1+0.5+(0.411*0.785) = 7.923
```

3 Results and discussions

3.1 Rabies in Hungary

Time series analysis

Descriptive analysis

In the monthly rabies cases time series we could identify three main components. There is a deterministic trend on long time scales (several years), then a 12-month seasonal periodicity on medium scale and finally a superimposed monthly random like fluctuation on short time scales. These fluctuations are related to the stochastic factors governing the processes. Our first goal was to identify the deterministic trends and the seasonal periodicity in the time series. Then we removed the deterministic trends and the seasonal periodicity and analysed the fluctuations with statistical methods, especially the extremely large fluctuations.

Annual trends

During the years, the average number of cases changed significantly (see Figure 2.1). In the Transdanubian data there was a growing trend until 1989 (month 256 in our data set starting in January 1967), followed by a decreasing trend until the start of the immunization in 1996 (month 348). In Figure 3.1, we show for each year the monthly average $\bar{f}(j) = \sum_{i=1}^{12} f(i,j)/12$, where $f(i,j)$ is the number of cases in month i of year j.
It is clear that the number of cases gradually increases from the mid-sixties to the mid-eighties in both parts of the country and then saturates. In Transdanubia it starts declining sharply from 1996 due to a mass immunization campaign.

Seasonal trends

Next, we concentrate on the seasonal variation of the number of cases. To be able to compare years with different average number of monthly cases, we divided the monthly values in each year by the yearly average shown in Figure 3.1. Thus, after dividing the monthly frequency by the yearly average, we get the detrended time series $e(i,j) = f(i,j)/\bar{f}(j)$. This time series shows

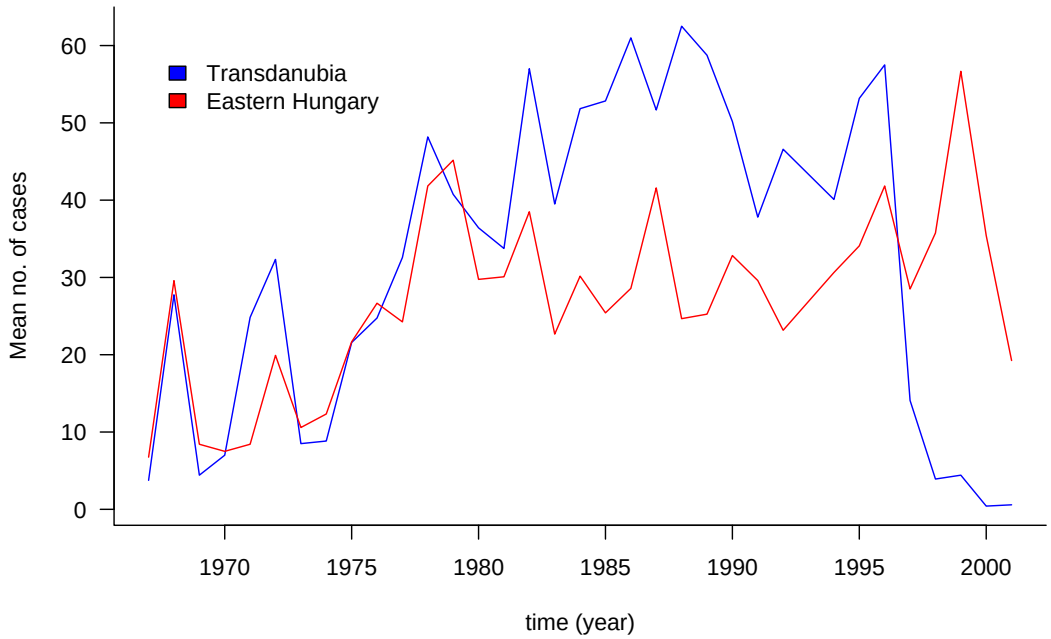


Figure 3.1: Monthly average of rabies cases in each year for Transdanubia and Eastern Hungary.

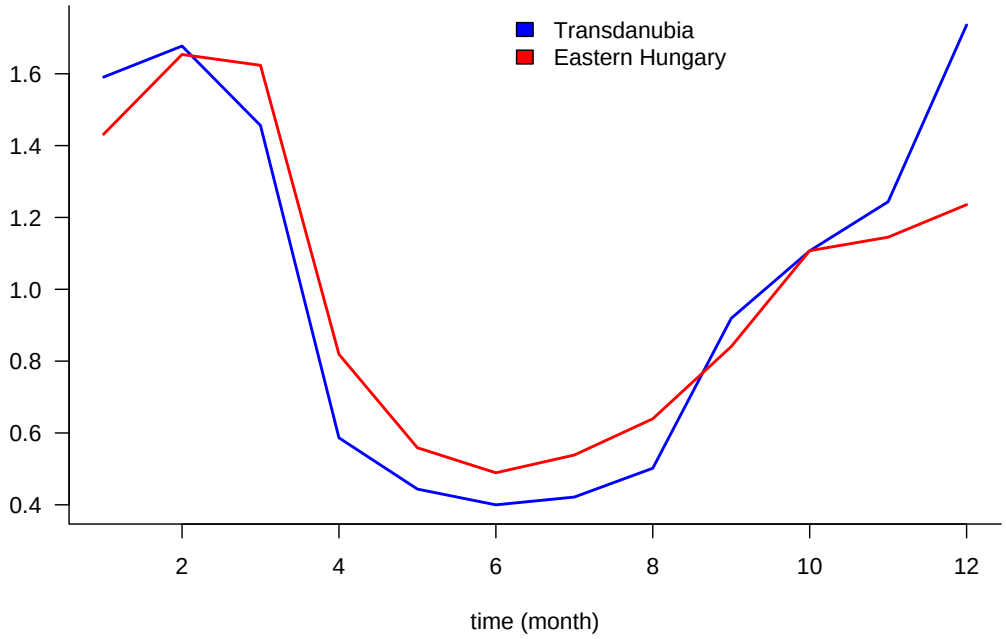


Figure 3.2: Seasonal trends in Transdanubia and Eastern Hungary (the value 1 represents the average of all years).

a 12-month seasonality due to yearly variation of the external conditions. In Figure 3.2, we show for each month the average number of cases calculated from all 34 years $\bar{e}(i) = \frac{1}{34}\sum_{j=1}^{34} e(i,j)$, $i = 1, \ldots, 12$. We can see that in January–March the number of cases is about 150% of the yearly average and in November and December it is also somewhat above the average. This is in accordance with other studies (Thulke et al., 2000) where this variation is attributed to the elevated number of contacts between foxes during periods of mating and dispersal.

Monthly random fluctuations

The monthly random fluctuations can be uncovered by removing the seasonal trend from the detrended time series $e_f(i,j) = e(i,j) - \bar{e}(i)$. In Figure 3.3 we show these fluctuations for Transdanubia and for Eastern Hungary. These time series are approximately stationary (Harnos et al., 2002), except for the beginning of both time series and the end of the Transdanubia time series. At the beginning, the relatively large fluctuations are related to the increasing trend in that period. The Transdanubia time series shows some large relative fluctuations at the end due to the decreasing trend caused by the immunization. Our next goal is to understand the statistics of extremely large fluctuations related to the immunization.

Analysis of extreme fluctuations

To analyse the large fluctuations identified in the previous section we refine our methods. Previously, we considered fluctuations relative to the long term and the seasonal trends. This method cannot be used if the average of the time series changes rapidly. In such cases it is more useful to compare the number of actual cases to a moving average over a preceding period. In our case, data show 12-month seasonal variations, so it is natural to consider a 12-month moving average and values relative to the moving average $l(i) = f(i)/(\sum_{j=1}^{12} f(i-j)/12)$, where $f(i)$ is the number of cases in the ith month in the series. The best way to study the statistics of the extremely large outbreaks is to consider the complementary probability distribution $F_c(l) = Prob\{l(i) > l\}$, which gives the probability that the relative fluctuation is larger than l. In Figure 3.3, we show the complementary cumulative distribution for the Transdanubia time series including data only before the start of the immunization campaign in 1992.

Next we show that this distribution develops an IPL tail. The method of fitting is called scaling analysis and it consists of fitting an IPL function $F_c(l) = al^b$ via fitting a linear function on the logarithmic scale log $F_c(l) = log\ a + b\ logl$, where l is the epidemic size (in our case normalized with the moving average), a and b (a negative number) are estimable parameters of the epidemic size distribution. On Figure 3.4, we show the IPL fitted to the Transdanubia data before immunization.

To study the effect of immunization we show in Figure 3.4, the complementary distribution for the Transdanubia time series including the immunized cases. The main effect of immunization on the statistics of large outbreaks is the change of the scaling exponent b of the IPL from about -1.5 before immunization to 2.0 after immunization.

In the next section, we will argue that this change in the exponent is related to the change of geometry of the spread of disease during the immunization campaigns. This can be understood within the framework of HOT introduced recently by Doyle & Carlson (2000) , and also applied in a theoretical biology context by Zhoua et al. (2005).

Highly Optimized Tolerance

The HOT forest fire model can also be applied in epidemiology: the animal population plays the role of forest and the "fire" is the epidemic itself. Animal populations are often separated by natural barriers such as mountains and rivers, and epidemics break out in the separated domains, when infected animals migrate into the area ("sparks"). The difference between the forest fire model and its epidemiological counterpart is that unlike trees, animals are able to move and also the barriers to the spreading of the disease can be more complex. Beyond natural barriers or artificial quarantines the spread of disease can be blocked by the presence of naturally or artificially immunized animal populations. In such cases the effective dimensionality of the problem may be different from its spatial dimensionality.

Since our rabies data set involves an immunization experiment, it gives us a unique opportunity to test the epidemiological applicability of the HOT model. In case of unimmunized data we expect that the fluctuations are determined by the spatial dimension ($d = 2$) of the problem and we expect that the complementary cumulative distribution of the fluctuations is going to have a power law tail with exponent predicted by the HOT theory: $b = -(1 + 1/d) = -1.5$. One

can see that the data collected at Transdanubia before the immunization supports our concept as the tail of the distribution is clearly consistent with an exponent 1.5. On the other hand, immunization changes dramatically the scaling behaviour of the tail of the Transdanubia data set involving the immunization, the full Transdanubia data set scales with an exponent near $b = -2$. This means that the effective dimensionality of our problem changes to $d = 1$. Figure 3.4 displays the two distributions in a single plot. The bodies of the distributions are almost the same and the change in the slope of the tail is clearly visible.

It is easy to understand the change in the effective dimensionality of the problem. Before immunization the disease can spread in two spatial dimensions as animals move around. After immunization large patches of immunized populations exist and the disease can spread only at the quasi one-dimensional borders separating large immunized populations where unimmunized populations can still exist.

Discussion

As a zoonosis and an incurable disease, Rabies has always been given due respect both in human and veterinary medicine. Most rabies cases in Hungary are still diagnosed in red foxes (*Vulpes vulpes*). This is in accordance with the European situation and confirms the role of the red fox as the reservoir and primary perpetrator of the disease. In this paper, we concentrated on extreme fluctuations in rabies cases. Extreme fluctuations are low-probability, high-consequence events (Englehardt, 2002) representing the majority of total losses. Understanding the size distribution of extreme events makes it possible to assess the risks of outbreaks.

We determined the distributions of extreme fluctuations which is a scale invariant power law distribution for both time series investigated. Scale invariance in the distributions of the sizes of fluctuations in complex dynamical systems has been explained on the basis of mechanical models of natural and engineered systems, such as models of SOC or HOT. In our case we could find an HOT theory explanation.

Owing to the fact that the exponent of the power law is related to the dimensionality of the process, we could show how immunization changed the structure of the problem. This kind of analysis can help to estimate the effects of eradication programs and to asses the risk of epidemic outbreaks. We think that HOT theory gives a reasonable explanation for the value of the IPL

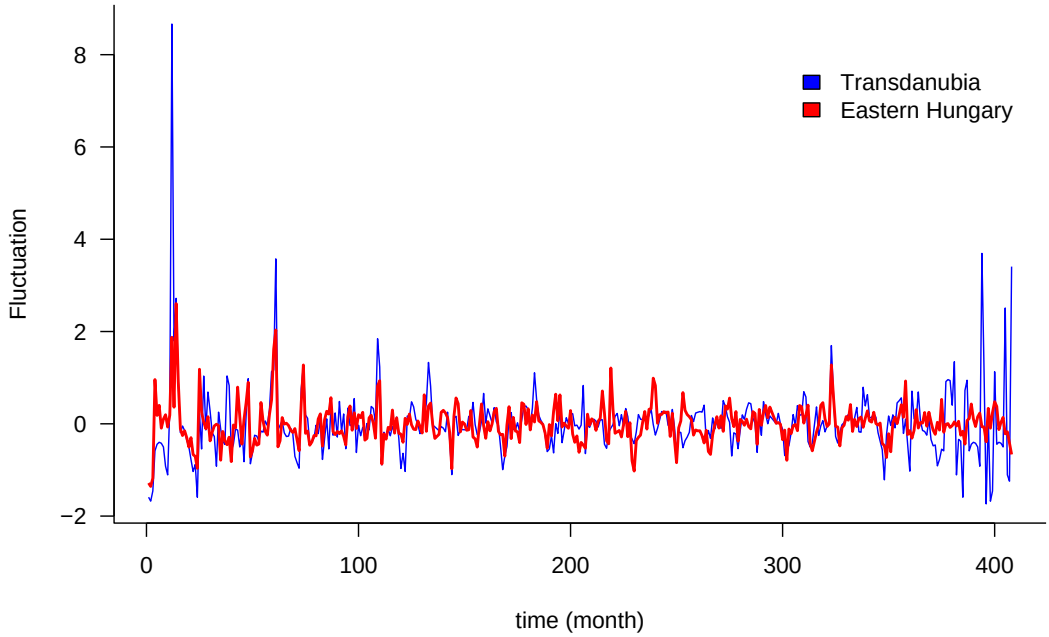


Figure 3.3: Monthly fluctuations in Transdanubia and Eastern Hungary.

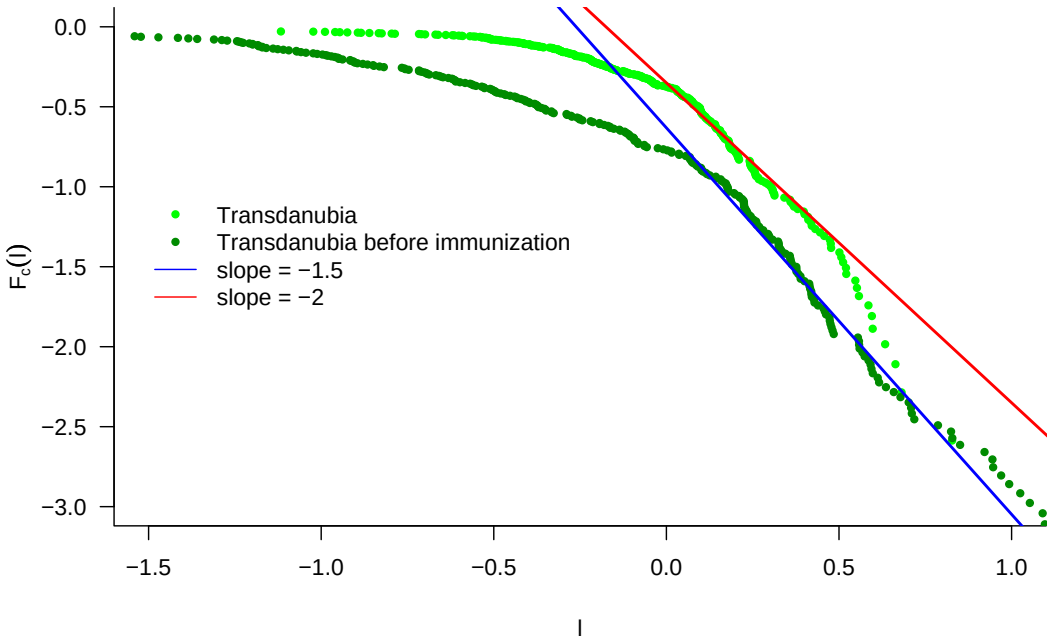


Figure 3.4: The complementary cumulative distributions of the fluctuations relative to 12 months moving average for the full and after immunization Transdanubian dataset on doubly logarithmic plot. The straight lines represent the fitted power law tails.

Table 3.1: The results of spatial clustering of rabies cases in Transtisza region.

Year	No. infected area	Case number	BB	Moran I	Geary c
1990	79	121	*87	*0.122	1.287
1991	93	152	*87	*0.161	1.103
1992	64	95	*42	*0.127	1.512
1993	91	164	*94	*0.147	1.541
1994	99	146	*115	*0.145	1.238
1995	86	118	*67	0.032	1.104
1996	114	181	*120	*0.075	1.734
1997	74	130	*66	*0.119	1.832
1998	94	170	*119	*0.179	1.359
1999	69	242	*70	*0.212	1.689
2000	87	150	*91	*0.191	1.183
2001	67	114	*59	*0.113	1.305

$*$: $p \leq 0.05$

exponent and for the change in it. In the future, we plan to investigate the spatiotemporal distribution of outbreaks in order to verify the HOT-based explanation in the details of the process.

Spatial clustering in the Transtisza region

The case number, number of affected areas and the test statistic of applied clustering methods are presented in Table 3.1. BB-test shows the affected areas were culstered for every years. Moran's I results are similar except for the year 1995, when the small positive autocorrelation is not significant. Although Geary's c indicates negative autocorrelation for every year, these were not significant in any year. These means that in Transtisza region between 1990-2001 rabies cases were aggregated and not sporadic. Since the rabies is a highly contagues disease it is not surprising the affected areas aggregated in the area where the eradication was not going on.
Figure 3.5 shows the correlations among test statistic of BB, Geary's c, Moran's I and case number (CN) and number of affected areas (AN). While BB test has a very strong positive correlation with number of affected areas, Moran's I and Geary's c have just a small negative. With case number the strongest correlation showed by Moran's I.

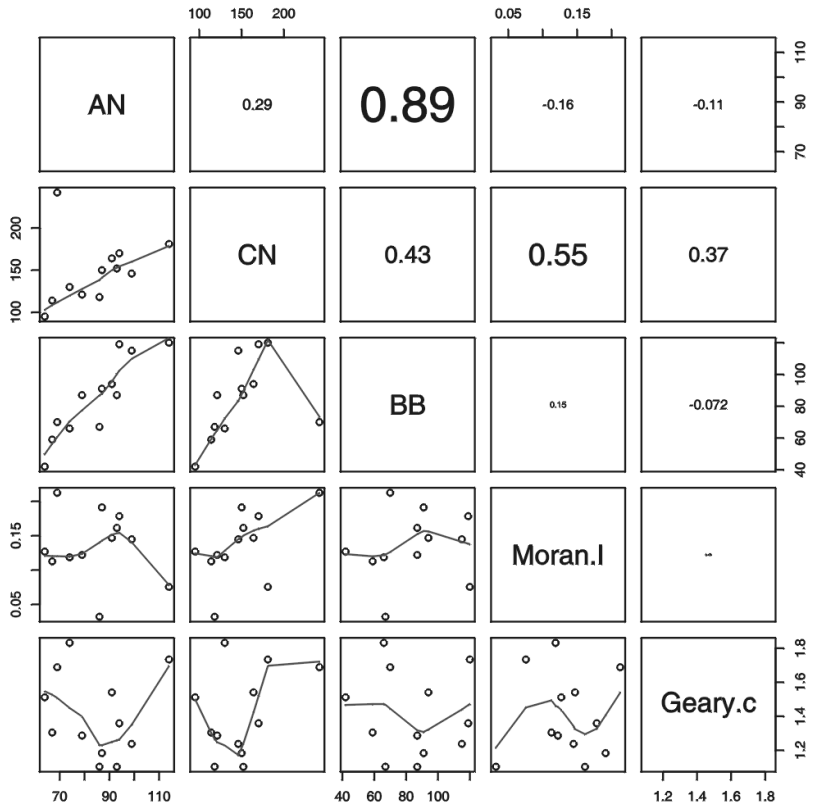


Figure 3.5: Correlations among number of cases (CN), of affected areas (AN) and statistic of BB, Geary's c and Moran's I.

Spatio-temporal clustering

The results of Knox's tests are presented in Table 3.2. The results showed ($*:\ p < 0.05$) that the clustering of case occurence is altering by years and regions. In the Transdanube region until 1996 (except for 1992) there was significant clustering in space-time. It seems that in years with high case number the test found a clustered occurence of the disease. After 1996 in three years (1998, 2000, 2001) there were no significant clustering demonstrable. One can interpret this as the result of oral immunization. This can mean also that with lower case number the occurence of disease is more probably sporadic than aggregated. But as rabies is a highly contagious disease and spreads onto an animal from an animal, its aggreagtion is necessary. So the absence of clusters may be caused by the relatively lower recorded cases in years with small incidence. In the region between Danube and Tisza we can find clusters in years with higher and not with

Table 3.2: The results of the Knox's test, with 4 km and 30 days critical distances.

	Transdanube		Between Danube and Tisza		Transtisza	
Year	χ^2	P	χ^2	P	χ^2	P
1990	12.757	0.000	45.851	0.000	9.528	0.002
1991	34.719	0.000	0.585	0.445	0.332	0.565
1992	1.958	0.162	0.619	0.431	0.031	0.861
1993	10.290	0.001	6.278	0.012	0.064	0.800
1994	4.228	0.040	8.975	0.003	1.839	0.175
1995	7.322	0.007	4.816	0.028	0.764	0.382
1996	5.655	0.017	4.618	0.032	20.204	0.000
1997	27.356	0.000	0.779	0.377	4.162	0.041
1998	0.088	0.767	6.519	0.011	0.116	0.733
1999	49.979	0.000	412.864	0.000	97.726	0.000
2000	0.000	1.000	62.927	0.000	1.312	0.252
2001	0.085	0.770	0.084	0.771	12.502	0.000

a lower case number. The oral vaccination program started in this region in 2000, in that year the cases were clustered. In the following year it had no clusters, just like in the Transdanube region. In the Transtisza region we found clusters in fewer years, than in the two other parts of the country. In years with higher case number in this region clusters were detectable as well. In three years, in 1990, 1996 and 1999 in every parts of Hungary the rabies cases among red foxes were clustered.

Relative risk estimations

The main results of the estimation of relative risks in small areas are summarized on Figure 3.6 and in Table B.1. On the Figure B.1–B.6 posterior expected relative risks and the correlation heterogeneities are mapped. Between 1990 and 2001 there were four district (*Csornai, Csurgói, Hevesi, Kapuvári*) where the posterior relative risk never exceeded 1. In four district (*Budapest, Dunakeszi, Mórahalomi, Váci*) during the studied period the posterior relative risk was above 1 ten times. In the *Miskolci* district the RR was above 1 eleven times.

Figure 3.7 shows the temporal trend estimates of relative risk in districts of Hungary for the period 1990-1996 and 1997-2001. This computed as $\exp(\beta + \delta_i)$ from the temporal model term. Districts where the posterior relative risk was decreasing in both periods (43): *Ajkai, Barcsi, Bicskei, Celldömölki, Csepregi, Csornai, Esztergomi, Győri, Hevesi, Kapuvári, Keszhely-Hévízi,*

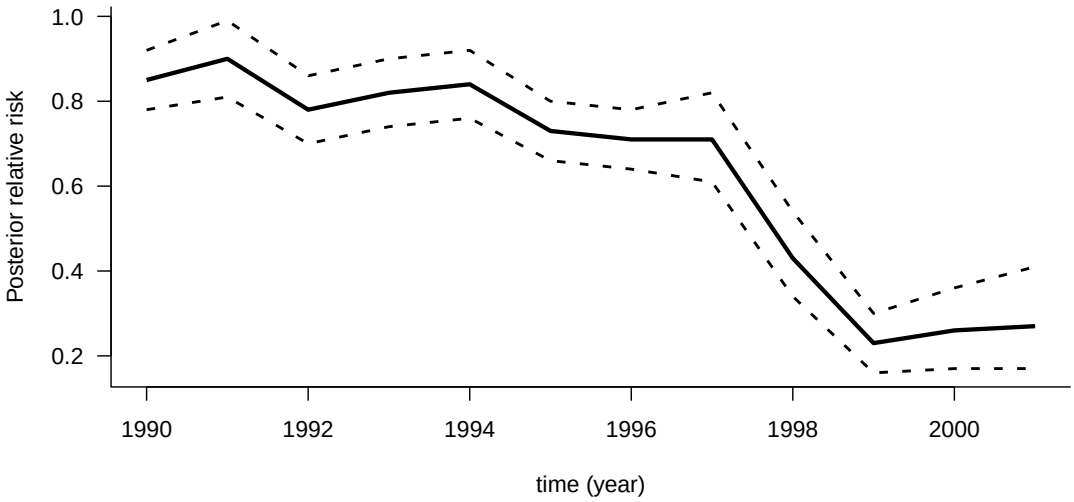


Figure 3.6: Yearly overall posterior relative risk (with 95% credible interval) for red fox rabies, Hungary.

Kisbéri, Komlói, Körmendi, Kőszegi, Lengyeltóti, Lenti, Letenyei, Marcali, Móri, Mosonmagyaróvári, Nagyatádi, Oroszlányi, Őriszentpéteri, Pápai, Pilisvörösvári, Sárvári, Sellyei, Siklósi, Sopronfertődi, Sümegi, Szentendrei, Szentgothárdi, Szigetvári, Szombathelyi, Tapolcai, Téti, Várpalotai, Vasvári, Veszprémi, Zalaegerszegi, Zalaszentgróti, Zirci.

Districts where the posterior relative risk was increasing in both periods (39): *Bajai, Baktalórántházai, Berettyóújfalui, Budapest, Ceglédi, Csengeri, Debreceni, Dunakeszi, Edelényi, Egri, Encsi, Fehérgyarmati, Hajdúböszörményi, Kazincbarcikai, Kecskeméti, Kiskunhalasi, Kisvárdai, Mátészalkai, Mezőkovácsházi, Mezőkövesdi, Miskolci, Mórahalomi, Nagykállói, Nagykátai, Nyírbátori, Nyíregyházi, Ózdi, Pétervásárai, Ráckevei, Salgótarjáni, Sárospataki, Sátoraljaújhelyi, Szécsényi, Szegedi, Szerencsi, Szikszói, Tiszaújvárosi, Tiszavasvári, Vásárosnaményi.*

Districts where the posterior relative risk was increasing in the period of 1990–1996 and decreasing in the period of 1997–2001 (27): *Balatonalmádi, Balatonfüredi, Bonyhádi, Budaörsi, Csurgói, Dombóvári, Dorogi, Dunaújvárosi, Enyingi, Ercsi, Fonyódi, Kaposvári, Komáromi, Mohácsi, Nagykanizsai, Paksi, Pécsi, Pécsváradi, Sárbogárdi, Sásdi, Siófoki, Székesfehérvári, Szekszárdi, Tabi, Tamási, Tatabányai, Tatai.*

Districts where the posterior relative risk was decreasing in period 1990–1996 and increasing in period 1997–2001 (41): *Aszódi, Bácsalmási, Balassagyarmati, Balmazújvárosi, Bátonyterenyei,*

Csongrádi, Dabasi, Füzesabonyi, Gödöllői, Gyáli, Gyöngyösi, Gyulai, Gyulai, Hajdúszoboszlói, Hatvani, Hódmezővásárhelyi, Jánoshalmai, Jászberényi, Kalocsai, Karcagi, Kiskőrösi, Kiskunfélegyházi, Kiskunmajsai, Kisteleki, Kunszentmártoni, Kunszentmiklósi, Makói, Monori, Orosházai, Pásztói, Polgári, Püspökladányi, Rétsági, Szarvasi, Szeghalomi, Szentesi, Szobi, Szolnoki, Tiszafüredi, Törökszentmiklósi, Váci.

In the Transdanube region the red fox rabies relative risk was clustered before the immunization. Since 1997, this elevated risk was decreasing due to immunization. The relative risk decreasing was clustered, which can be interpreted that the immunization was able to create blocks on the region. But it could be an interesting phenomenon at the western border with the districts *Csepregi, Körmendi, Kőszegi, Sárvári* and *Szombathelyi*. These districts show a cluster in 1990 (Figure B.1). At the end of study period, 2001 these districts also show a cluster (Figure B.6). Although in this year the relative risk is below 1 for these districts, but higher than in the neighboring districts. The maps show that while in the western part of the country the RR was decreasing, in the eastern part there was an increase. This increase seems to be clustered in the Great Plain.

3.2 Null distribution of Moran's I based on stratified MC

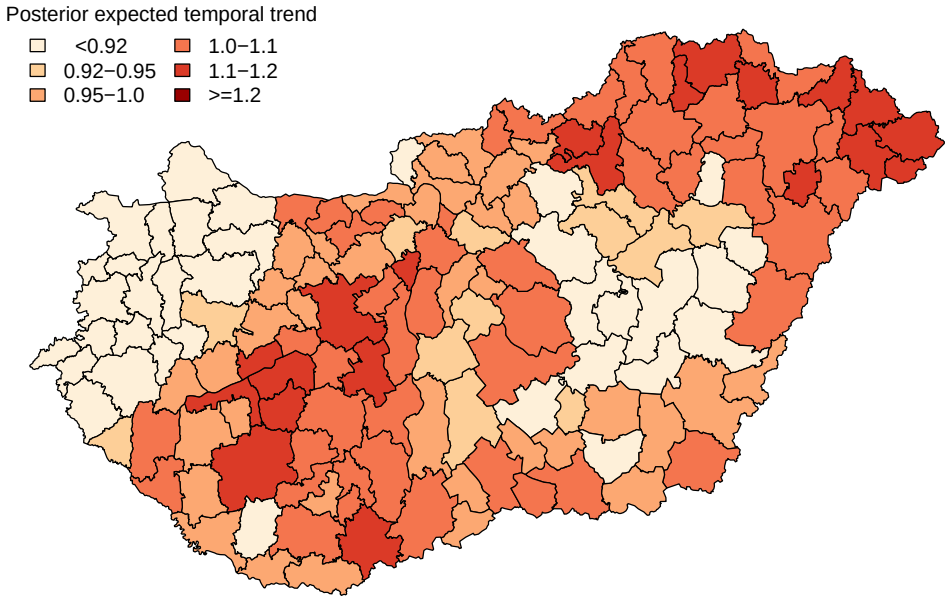
Conclusion. Due to the results of simulations (Table 3.3) we can conclude that if the variable of interest has high correlation with the size of observational units or population (Table 3.4), then there is a good choice to stratify the recordset to homogeneous subsets as the base of the Monte Carlo permutation. Here the subset size was arbitrary. Further research may focus on developing a method to define homogeneous strata. For example to reduce the correlation within the subsets by spalling.

3.3 Environmental association studies

Spatial risk assessment of herd sero-status of Aujeszky's disease

Figure 3.8 shows the results of the logistic regression analysis. The first two factors ("lake" and "highway") are positively associated with positive sero-status, whereas the other two factors ("forest" and "seronegative large-scale unit") are negatively associated with it.

(a)



(b)

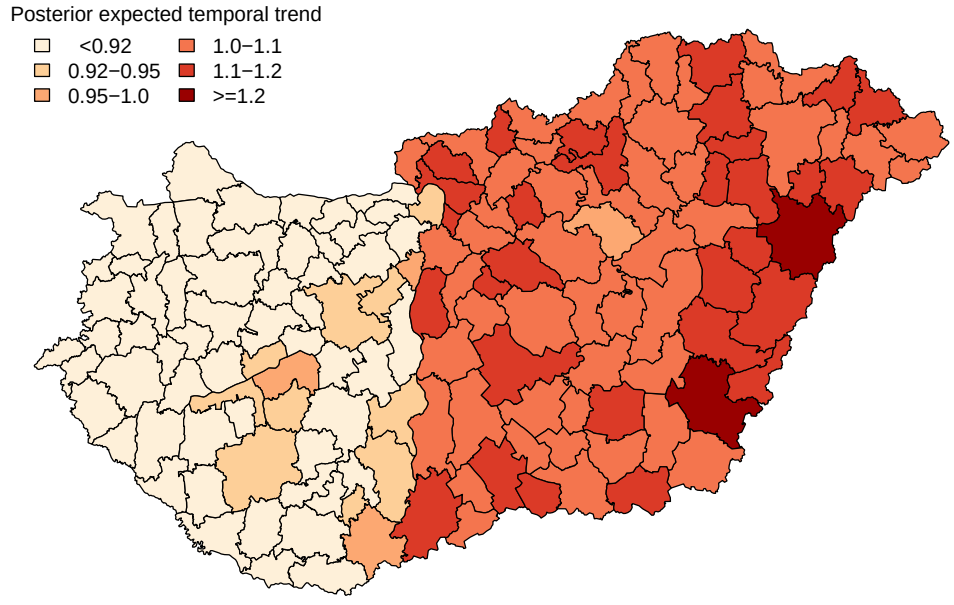


Figure 3.7: Posterior expected temporal trend (space × time model (Bernardinelli et al., 1995b)) for red fox rabies, Hungary, 1990-1996 and 1997-2001.

Table 3.3: Type I error rates based on different methods obtaining null distribution of Moran's I. The standardization n_i/a_i illustrates (n_i is case number, a_i is area of i^{th} subregion) the effect of incorrectly chosen standardization on the results, while $n_i/\sqrt{a}_i$ represents the correct standardization, which corresponds to the true model.

Method	x_i	Region Transdanube	Between the Danube and Tisza	Transtisza
Randomization				
	n_i	0.069	0.192	0.363
	n_i/a_i	0.068	0.109	0.132
	$n_i/\sqrt{a}_i$	0.067	0.052	0.060
Monte Carlo				
	n_i	0.061	0.176	0.347
	n_i/a_i	0.057	0.093	0.106
	$n_i/\sqrt{a}_i$	0.061	0.045	0.050
Stratified MC				
	n_i	0.028	0.046	0.056

Table 3.4: Correlation between the size and case number of districts in the three parts of Hungary.

	Transdanube	Between the Danube and Tisza	Transtisza
R	0.629	0.585	0.761

In their study, Marsh et al. (1991) concluded that a distance < 5 km between pig units might increase the risk of disease spread. However, their data covered just 41% of the known pig units in the study area. Norman et al. (1996) showed that within 3.22 km, a high density of non-quarantined herds or low density of quarantined herds is risk decreasing – whereas low density of vaccinated herds is a risk-increasing factor. These results were based on data of about 75% of the farms in the study area. Similar results hold for Csongrád county with radii of 4–7 km. In another publication, the farms within 2.5 km seemed to be a risk factor (Rodríguez-Buenfil et al., 2002). Tamba et al. (2002) found that if there are at least 10,000 pigs within 6 km, the risk increases. In our study, the proximity of infected villages seemed associated with seropositivity. We suppose that workers living in neighbouring villages acted as vectors between their own herds and the swine unit they were working at but the possibility of air-born spread also exists. While the results from Csongrád county suggest that lakes close to pig units increase the risk of seropositivity, Marsh et al. (1991) indicated that the nearby presence of lakes reduces the risk of infection spreading. We speculate that more abundant vegetation around the lake

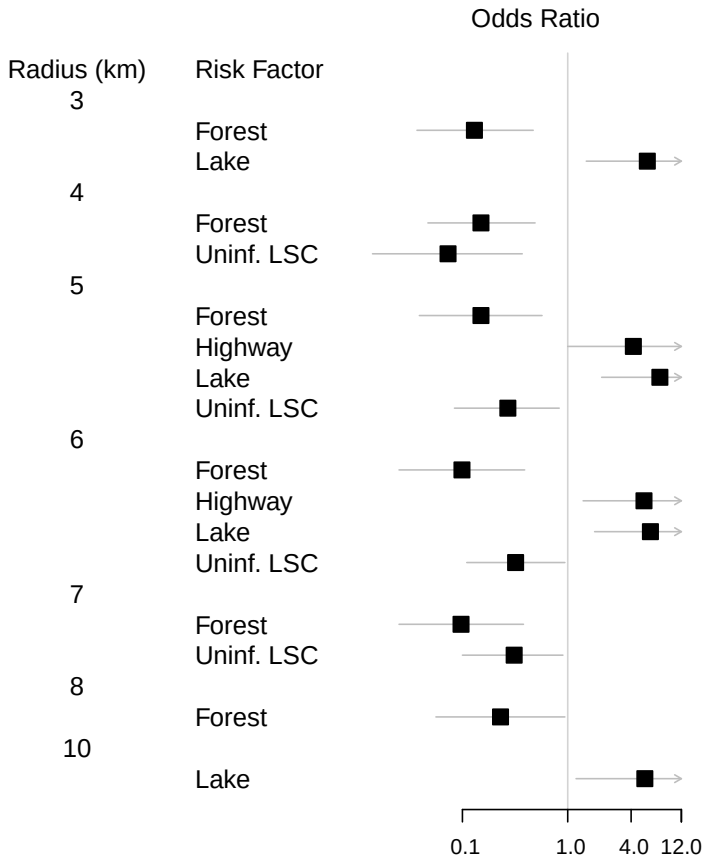


Figure 3.8: Final logistic multiple-regression models of risk factors (within indicated radii) for seropositivity of swine units for Aujeszky's disease in a county in Hungary, 1998–2000. (*Uninf. LSC: Uninfected large-scale swine unit*).

increases the risk of infection through feral swine or wildlife vectors (e.g. rodents, foxes, birds). Another interpretation is that fog and higher humidity increases the possibility of airborne virus transmission. On the other hand, lakes and rivers might act as barriers (allowing no animal or vector movements). It is possible that in one study area, this effect dominates the other one while in another study area it is the other way round. Another result of our study indicates that the presence of a highway might increase the risk, but Marsh et al. (1991) could not detect association between ADV status and the nearest county road or highway. Csongrád county data

lead to the hypothesis that forests might decrease the risk of infection transmission (perhaps by decreasing air-borne transmission).

Our results must be interpreted with caution, because the investigation area is rather small and we did not correct for edge effects (arising from spatial censoring at the borders; Lawson, 2001). On the other hand, a larger study area would have a higher chance to be inhomogeneous making it difficult to draw conclusions valid for the whole area. Presence of a lake or highway showed positive association with seropositive status of the unit, while in case of forests and uninfected farms the association was negative.

TETYN: tool for extracting climatic parameters from Tyndall data sets

The result of development is a tool that helps the user calculate climate variables based on the TYN SC 1.0, TYN SC 2.0 and the CRU TS climate datasets (Timbal et al., 2009). Without programming abilities and knowledge of the various data structures one can set up a complex query using an easy to use graphical interface. The tool helps the user to query spatially not just the whole grid of the datasets, but only a certain region. Spatial queries can be performed by one geolocation, bounding box, Tyndall grid coordinates and by choosing one or more countries depending on interest. Results of the queries are easy to import into spreadsheet management software (R Development Core Team, 2009) and GIS tools.[1]

Based on Tyndall datasets we are studying the associations of spatial pattern of climate parameters and vectors of *Leishmania infantum*. Preliminary results showed that the spatial distribution of some *Phlebotomus* species has strong relationship with the monthly mean temperature patterns in the period 1961-1990. As a next step we will generalize the relations to predict the expectable pattern of spatial distribution of this kind of vectors in the near future due to different climate scenarios.

[1] http://www.qgis.org/

4

Aknowledgements

I would like to say thanks to †*István Medveczky* who initiated my work on veterinary epidemiology. Many thanks for the help of my supervisor *Lajos Rózsa*. During the years I worked in epidemiology and computational biology helped me with their knowledge and friendship *Andrea Harnos*, *Jenő Reiczigel*, *Olaf Berke*. My studies were technically strongly supported by †*Zsolt Harnos*, *János Fodor*. The financial support (project numbers: OTKA T035150, T032437 and ACTION Austrian-Hungarian Foundation 2003-2004 (53ÖU2)) is gratefully acknowledged. I could always count on my family, many thanks for their perseverance.

5 Summary

5.1 Introduction

For years, my main topic has been studying spatial aspects of various phenomena. Mostly I worked on veterinary epidemiology, but sometimes I was involved in human health related or social science studies. In this dissertation I collected some of my works in connection with veterinary spatial, spatio-temporal epidemiology.

As a zoonosis and an incurable disease, rabies has always been given due respect both in human and veterinary medicine. In Hungary up to 1954 there were just the urban rabies, spread by dogs. The sylvatic form rabies appeared in 1954, from the North-East direction. In Europe the reservoir and primary perpetrator of this form of disease is the red fox. Up to 1970 sylvatic rabies became endemic in Hungary. To decrease the risk of infection an oral vaccination campaign was started in Hungary to immunize the foxes. In the spreading of the infection by foxes there are two natural borders in Hungary: the Danube and the Tisza River. There is low possibility that the foxes go trough these rivers. By this cogitation these different regions are handled in the immunization as independent areas. First the oral vaccination was started in the region beyond the Danube. From October 1996 up to 2000, the whole western part of the country was involved to the immunization. Since 2000, the oral immunization process is being done in the central part of the country. According to our goal to understand the epidemiological properties of rabies in Hungary, to analyse the effectiveness of eradication program we performed analyses on the temporal, spatial and spatio-temporal properties of reported rabies cases in Hungary.

In the estimation relative risk of rabies occurence I used Bayesian models as well. In spatial epidemiology the usage of the Bayesian methods is increasing. To make the spatial Bayesian modelling easier I have developed a tool for preparation of data and maps to help the user of GeoBUGS.

We had some interesting experiences using Moran's I method in analyses of spatial clustering of rabies cases. Moran's I is a frequently used statistic for quantifying spatial clustering. I measures the spatial autocorrelation of the variable of interest with respect to the neighborhood structure of the spatial regions to be analyzed. It serves as a basis of local as well as global tests.

The variable of interest may vary from count data, i.e. incidence or prevalence, to rates and proportions. The null hypothesis of Moran's test is the lack of autocorrelation. Traditionally the null distribution of I is determined on the basis of distributional assumptions and asymptotic normal theory. Another method to determine the null distribution, which is regarded more reliable in practice, is based on a permutation argument relying on the assumption of exchangeability under the null hypothesis. It simply means that under the null hypothesis permuting the values of the variable of interest does not affect the distribution of I. Under this assumption, the null distribution can be determined either theoretically or by Monte Carlo simulation, generating a large number of random permutations of data. If exchangeability does not hold, the p-value obtained from the test may be invalid, i.e., the test may falsely detect significant autocorrelation or clustering (the actual Type I error rate may be higher than the nominal one). It is often quite clear how exchangeability depends on the choice of the variable to analyse. If incidence is known to be proportional to population size and population size varies from region to region, then exchangeability requires the use of rate (incidence per 10000 inhabitants) rather than raw incidence counts. This is the case for many chronic or non-infectious diseases. For a contagious disease, however, the assumption of proportionality may not hold. Dependence of incidence on population density may not be linear, that is, a two times bigger population density may result in an incidence more than twice as much. In this case, the use of rate does not ensure exchangeability.

Researchers typically do not search for spatial clustering due to known factors like population size or age distribution. They would rather eliminate the effects of those explanatory variables to avoid results which trivially mirror the pattern of a known factor, like e.g. the clustering of urban-rural areas in the country. So the natural null hypothesis sounds like lack of autocorrelation after controlling for the uninteresting or known explanatory variables. Sometimes it is rather puzzling how to eliminate the effects of some factors or how to adjust or standardize observed values to ensure exchangeability. These prompted us to deal with the null distribution of the test statistic. We developed a stratified Monte Carlo method to obtain the null distribution of I.

The Aujeszky's disease virus (ADV) belongs to the family of Herpesviridae which establishes a lifelong persistent infection with intermittent shedding of the virus. The infection can cause clinical disease with neurological symptoms in suckling and weaned pigs and respiratory or

reproductive disorders in grower-finisher and adult animals. Pigs are the only natural host of ADV. Other animals found on farms (such as rats and mice) are dead-end hosts. The main route of transmission of ADV among swine populations is the physical contact between infected and susceptible animals. Less important for the transmission are fomites such as feed, vehicles or artificial insemination. Other routes of infection of susceptible populations are long-distance airborne transmission and vector-mediated transmission. The distances among swine populations and the surrounding topographical features might be important for the last two routes. The health status of the pig population and the success in eradication of Aujeszky's disease (AD) are influenced by the geographic population density. Nevertheless, the effect of the geographical features on the spread of infection has been a rarely investigated topic. At the end of the nineties I was invited by †*István Medveczky* into the analysis of the results of countrywide Aujeszky's disease eradication program. At county level we published an evironmental association paper from this disease.

In recent years the importance of climate change has been rapidly growing in most disciplines, including agriculture, health, ecology, economy and social science. Since 2007 I have been working on studying the possible associations of animal health and climate change. Based on climate and other environmental (e.g. landcover) datasets, researchers can model environmental associations and/or predict spatial distribution of diseases. Due to climate change expectedly new vector-borne diseases may appear in the temperate zone. For modelling the spatial pattern changes in disease distributions due to global changes we are studying the environmental similarity of differenet sites of Earth. In this work we try to find sites on the Globe where the climate in the past is close to the future climate in Hungary. Identifying similar areas we could predict for which pathogens the climate of Hungary will be suitable. For this modelling we need adequate data sources. For researchers working on near-term and far-term modelling the Tyndall Centre for Climate Change Research produced two data sets based on different climate scenarios: the TYN SC 1.0 and the TYN SC 2.0. The first one is projected onto a 10-min spatial resolution grid over Europe, while the second is projected onto the whole globe with a 0.5°grid. These data sets contain future climate data comprising five parameters on a monthly basis. Besides these data sets the Tyndall Centre also provides the CRU TS data sets of the observed monthly meteorological data in the 20^{th} century containing nine parameters. The CRU TS data set is on

the same grid as TYN SC 2.0. The TYN SC data sets are free on request. The newest version of CRU TS data set is also freely downloadable. Although some Fortran programs are available for managing the TYN SC data sets, their use may prove to be difficult for a non-expert user. As a first step of my work I developed a tool to make easier the usage of climate datasets produced by Tyndall Centre.

5.2 Results

Rabies

In one of our studies the goal was to analyse the fox rabies cases time series from 1967 to 2001 and to identify its main characteristics. In this study, we concentrated on extreme fluctuations in rabies cases. Extreme fluctuations are low-probability, high-consequence events representing the majority of total losses. Understanding the size distribution of extreme events makes it possible to assess the risks of outbreaks. We determined the distributions of extreme fluctuations which is a scale invariant power law distribution for both time series investigated. Scale invariance in the distributions of the sizes of fluctuations in complex dynamical systems has been explained on the basis of mechanical models of natural and engineered systems. In our case we could find an Highly Optimized Tolerance (HOT) theory explanation. Due to the fact that the exponent of the power law is related to the dimensionality of the process, we could show how immunization changed the structure of the problem. This kind of analysis can help estimate the effects of eradication programs and to asses the risk of epidemic outbreaks. We think that HOT theory gives a reasonable explanation for the value of the IPL exponent and for the change in it. Our results showed that the effective dimensionality of our problem changes from two to one.
Spatial clustering analyses showed that in the Transtisza region between 1990-2001 rabies cases were aggregated and not sporadic. The results of Knox's spatio-temporal clustering tests resulted in the fact that the clustering of case occurence is altering according to year and region. In the Transdanube region until 1996 (except for 1992) there was significant clustering in space-time. It seems that in years with high case number the test found clustered occurence of disease. Since 1996 in three years (1998, 2000, 2001) there was no significant clustering demonstrable. In the region between the Danube and Tisza we have found clusters in years with a higher and not with

lower case number. The oral vaccination program started in this region in 2000, in that year the cases were clustered. In the following year it had no clusters, just like in the Transdanube region. In the Transtisza region we found clusters in fewer years, than in the two other parts of the country. In years with a higher case number in this region clusters were detectable as well. In three years in 1990, 1996 and 1999 in every parts of Hungary the rabies cases among red foxes were clustered.

Null distribution of Moran's I

We proposed a method for the case in which the variable of interest is in a presumably monotone but unknown relationship to an observed covariate. (If we know the type of relationship, e.g. quadratic, logarithmic, etc., then the best to do is to standardize the values by this function.) Our results showed that, if the variable of interest has high correlation with the size of observational units or population then it is a good choice to stratify the recordset to homogeneous subsets as the base of the Monte Carlo permutation. Here the subset size was arbitrary. Further research may focus on developing a method to define homogeneous strata. For example, to reduce the correlation within the subsets by spalling.

Aujeszky's disease

A Geographic Information System (VetEpiGIS) was developed to analyze the ADV (Aujeszky's disease virus) sero-status in large-scale pig units regarding certain geographical features in a county of southern Hungary. The ADV sero-statuses were collected from all swine units in Csongrád county in 1998-2000. The units' coordinates were combined with a vector graphical digital map of the county. Logistic regression tested the associations between sero-status of large-scale units and presence of topographical features, other units and villages in the neighborhood. The term "neighborhood" was defined by circular zones with a radius of 1-10 km around the unit. Presence of a lake or highway showed positive association with seropositive status of the unit, while in case of forests and uninfected farms the association was negative.

Software development

maps2WinBUGS

For Bayesian spatial modelling, GeoBUGS is a widely used tool incorporated in WinBUGS and OpenBUGS. The map format that GeoBUGS own map format uses differs from the standard formats used in geographical information systems (GIS). I have developed a tool (maps2WinBUGS) which helps the user prepare maps and tabular data to be used in GeoBUGS. With this tool one can obtain adjacency lists, manipulate maps and visualize the results of runs. By the script wizard one can generate BUGS or R scripts using three different models based on own map and data. The resulting maps can be exported into different image and GIS file formats (ESRI shape and Google Earth files).

TETYN

I developed TETYN with the aim of making it easier to extract certain parts from the above described data sets. The climatic data sets contain precursors for the different parameters that are actually calculated from these precursors. The results of TETYN queries can be saved as comma separated values (CSV) or ESRI shape files. In both formats the records represent the gridboxes and columns containing the monthly data.

6 Összefoglalás

6.1 Bevezetés

Jelen dolgozatban a térbeli-, tér-időbeli állatorvosi epidemiológia témakörében végzett olyan kutatások eredményeit gyűjtöttem össze, amelyekben meghatározó szerepem volt.

Mint zoonózis és gyógyíthatatlan betegség, a veszettség mind humán, mind állategészségügyi szempontból fontos. Hazánkban a veszettség 1954-ig úgy nevezett „urbanus" formában fordult elő, aminek terjesztésében központi szerepe a kutyáknak volt. A betegség ún. „sylvaticus" formája a II. világháború után a Baltikumból Nyugat, Dél-Nyugat felé terjedt és érte el hazánkat is. Ebben a formában a fertőzés fenntartója a vörös róka egész Európában. Magyarországon a betegség előfordulásának visszaszorítása céljából 1992 óta különböző méretű területeken folyik a rókák orális immunizációja. Vizsgáltuk a veszettség hazai előfordulásának időbeli- és térbeli mintázatát, mind az immunizáció előtti, mind pedig az azt követő időszakra vonatkozóan. A betegség térbeli rizikó-elemzése során térbeli Bayes-i modelleket alkalmaztunk. A térbeli Bayes-i modellezésben általánosan elterjedt szoftver a GeoBUGS, ami része a WinBUGS és OpenBUGS alkalmazásoknak. A GeoBUGS által használt térképi formátum eltér a térinformatikai rendszerekben (GIS) általánosan használt formátumoktól, valamint egyes, az elemzésekhez fontos funkciója nem ismert módon működik. Az egészséggel kapcsolatos adatok térbeli mintázatelemzése során több nehézségbe ütköztem a szoftver használata közben. A modellezést megkönnyítendő fejlesztettem a maps2WinBUGS alkalmazást.

A veszettség térbeli halmozódásával kapcsolatos elemzéseink során a témakörben használtos Moran-féle I statisztika null-eloszlásának meghatározására egy rétegzett Monte Carlo eljárást dolgoztunk ki. A Moran-teszt nullhipotézise a korrelálatlanság, azaz a térbeli autokorreláció hiánya. A statisztika null-eloszlását vagy aszimptotikusan (normális vagy permutációs aszimptotikát használva), vagy pedig Monte Carlo-módszerrel szokták meghatározni (permutációs vagy randomizációs teszt). Gyakran nyilvánvaló, hogy a felcserélhetőség teljesülése miképpen függ a vizsgált változó megválasztásától. Például ha a területeken élő népesség létszáma eltérő, akkor jobb, ha a változó nem az esetszám (a megbetegedések megfigyelt száma), hanem a populációra vetített ráta (az 1000 lakosra jutó megbetegedések száma). A vizsgált változó ilyen módon

történő megválasztása tulajdonképpen a mért értékeknek a felcserélhetőséget biztosító „közös nevezőre hozása”, amelyet a továbbiakban standardizálásnak nevezünk. Ha a felcserélhetőség nem teljesül, akkor az elsőfajú hiba valószínűsége a névleges 5%-nál magasabb lehet. A nyers, megfigyelt adat gyakran az esetek abszolút száma. Ha azt gyanítjuk, hogy ezt használva a felcserélhetőség nem teljesül, akkor kereshetünk olyan változót, amellyel standardizálva vélhetően már teljesülni fog. Elképzelhető azonban, hogy nem a megfelelő változót választjuk, sőt az is, hogy a megfelelő változó nem is áll rendelkezésünkre.

Az Aujeszky-betegség sertésállományokban jelentős gazdasági károkat előidézni képes vírusos fertőző betegség. A fertőzés leggyakrabban állatról-állatra, közvetlenül terjed. Emellett más terjedési útvonalak is ismeretesek. A sertésállományok közötti egyéb terjedésben szerepet játszhatnak az azt elősegítő, illetve gátló környezeti tényezők. Sertéstelepek Aujeszky-betegségre vonatkozó szerostátuszának környezeti tényezőkkel való összefüggéseire vonatkozóan Csongrád megyében készítettünk tanulmányt.

Az utóbbi években számos területen (pl. mezőgazdaság, egyészségügy, ökológia, közgazdaság, társadalomtudományok) növekszik a figyelem a klímaváltozással kapcsolatban. Számos tanulmány foglalkozik a változásokhoz való alkalmazkodás fontosságával Az emberek és állatok egészségét számos oldalról érinthetik a globális változások. A fertőző betegségek földrajzi elterjedtségének mintázatváltozása is ilyen. Munkánk során mi is foglalkozunk olyan modellezésekkel, amelyek segítségével előrejelezhetjük, hogy milyen új fertőző betegségek megjelenésére számíthatunk, illetve az endémiások térbeli mintázatában milyen átalakulások várhatók a globális változások függvényében. A kutatóknak, döntéshozóknak szükségük van olyan, a jövőre vonatkozó klíma-modellekből származó adatforrásokra, amelyek alapján kvantitatív elemzéseket, modellezéseket végezhetnek az egyes témakörökben. Ezt felismerve számos intézmény, szervezet hoz létre a kutatók számára elérhető adatállományokat klíma-szcenáriók alapján. Ilyen adatállomány például azok számára, akik rövidtávú vagy hosszútávú modellezéssel foglalkoznak, a Tyndall Centre for Climate Change Research által létrehozott TYN SC 1.0 és a TYN SC 2.0. Előbbi Európára 10 perces térbeli felbontású rácson, míg az utóbbi a teljes Földre vonatkozóan tartalmaz adatokat 0.5°-os felbontásban. Az adatállományok öt klimatikus paraméter havi értékeinek kiszámítására tartalmaznak adatokat. A jövőre vonatkozó adatok mellett fontosak a múltra vonatkozó klimatikus adatok is, egységes térbeli felbontásban. A Tyndall Centre ilyen adatál-

lományt is rendelkezésre bocsát, mint pl. a CRU TS 2.0, amely a XX. századra vonatkozóan tartalmaz havi klimatikus adatokat ugyanazon a térbeli felbontáson, mint a TYN SC 2.0. Mindegyik adatállomány ingyenes és szabadon felhasználható. Habár néhány Fortan program elérhető az adatállományok kezelésére, ezek alkalmazásához speciális szoftver-környezetre és programozói ismeretekre van szükség, ami az alkalmazott tudományok művelői esetén sokszor hiányzik. Az adatok könnyebb kezelését segítendő fejlesztettem a TETYN elnevezésű alkalmazást.

6.2 Eredmények

Veszettség

Létrehoztunk egy adatbázist, amely térbeli referenciával tartalmazza a hazánkban 1958 és 2001 között bejelentett veszettség-eseteket. A bejelentett veszettség-esetek időbeli mintázatának elemzése során különös figyelemmel fordultunk a betegség előfordulásának extrém ingadozása felé. Az extrém ingadozások alacsony valószínűségű, de komoly következményeket magukban rejtő események. Az extrém események eloszlásának vizsgálata lehetőséget adhat a kitörések veszélyének becslésére. Az 1967 és 2001 közötti összes dunántúli esetet és az immunizálás utáni eseteket vizsgáltuk. Az extrém ingadozások eloszlásának vizsgálatakor azt találtuk, hogy az jól közelíthető egy skálafüggetlen hatványfüggvény-eloszlással, mindkét idősorra vonatkozóan. Komplex dinamikus rendszerekben az ilyen jellegű eloszlások interpretációjára számos elméleti modellt alkalmaznak. Egyik ilyen a Highly Optimized Tolerance elmélet, melyet a hatványfüggvény kitevőjének értelmezését segítendő mi is használtunk. Figyelembevéve a hatványfüggvény kitevőjének sajátosságát arra vonatkozóan, hogy a vizsgált folyamat dimenzionalitásával mutat összefüggést, arra a következtetésre jutottunk, hogy az immunizálás a terjedés térbeli szerkezetét változtatta meg. Az eredmények azt jelzik, hogy a terjedés korábbi kétdimenziós volta az immunizálás következtében egydimenzióssá változott, mivel a fertőző forrás környezetében az immunizált egyedek jelenléte a minden irányba történő terjedés ellen hat.

A térbeli mintázat elemzése során a Tiszántúlon 1990 és 2001 között rókákban előfordult és bejelentett eseteket vizsgáltuk. Az eredmények azt mutatták, hogy egy év (1995) kivételével szignifikáns, gyenge pozitív autokorrelációt mutatnak az esetek. A Knox-féle tér-idő klaszterelemzés a 4 km-es térbeli, illetve 30 napos időbeli kritikus távolságokon belül országrészenként

és évenként különböző klaszterez̋ődést jelzett. A Dunántúlon 1996-ig (1992 kivételével) szignifikáns tér-idő beli halmozódást láttunk. A magas esetszámú években halmozódás tapasztalható. Az immunizáció megkezdése után három évben (1998, 2000 és 2001) nem volt kimutatható tér-idő aggregáció. A Duna-Tisza közén a magas esetszámú években szintén halmozódást figyeltünk meg, míg alacsony esetszámú években az esetek előfordulása sporadikus volt. Ezen a területen az immunizálás megkezdésének évében (2000) halmozódás volt megfigyelhető, azonban az azt követő évben már nem, ugyanúgy, ahogy a Dunántúlon. A Tiszántúlon kevesebb évben láttunk halmozódást, mint a másik két országrészben. A magasabb esetszámú években itt is halmozódást mutathattunk ki. Ugyanakkor figyelembevéve ennek a területnek a térbeli halmozódásra vonatkozó elemzések eredményeit, azt láthatjuk, hogy az időbeli határérték jelentős limitáló tényező az esetek klaszterez̋ődésében. Hiszen míg a térbeli autokorreláció majd minden évben megállapítható volt, addig a tér-idő beli aggregáció ritkábban. Három magas esetszámú évben (1990, 1996 és 1999) az ország mindhárom részén szignifikáns tér-idő beli klaszterez̋ődést mutattunk ki.

Moran I null-eloszlása

A standardizálás helyett rétegzett permutációs módszert javaslunk. A módszer lényege, hogy ha van egy „gyanús változónk”, vagy egy azzal vélhetően korrelált változónk, akkor képezzünk rétegeket e változó szerint, és csak a rétegeken belül permutáljunk. A terület szerinti rétegzést annyiban tartjuk jobbnak a területtel való standardizálásnál – amely itt „esetszám per a terület nagysága” rátát jelentene – hogy a rétegzés akkor is működik, amikor az összefüggés nem lineáris, míg a ráta lineáris összefüggést feltételez.

Aujeszky-betegség

Csongrád megye összes nagylétszámú sertésállományának Aujeszky-betegség vírus (ADV) fertőzöttségre vonatkozó szerológiai státuszának környezeti tényezőkkel való összefüggését vizsgáltuk. A feladat megvalósítása érdekében egy térinformatikai rendszert (VetEpiGIS) fejlesztettünk. Összegyűjtöttük Csongrád megye összes sertésállományának 1998 és 2000 között rögzített ADV szerológiai státuszára vonatkozó vizsgálati eredményét. A nagylétszámú sertésállományok térbeli pozícióját GPS-el határoztuk meg. Ezeket a megye vektorgrafikus digitális térképével

szerveztük egységes rendszerbe. Logisztikus regressziót használtunk a nagylétszámú állományok szerológiai státusza és a „szomszédosságukban” előforduló topográfiai objektumok összefüggésének elemzésében. A „szomszédosság”-ot a sertésállományok körül különböző sugarú (1-10km, kilométerenkénti lépésekkel) körkörös zónák létrehozásával definiáltuk.
A következő földfelszíni objektumokra vonatkozóan találtunk szignifikáns összefüggést a szerostátusszal kapcsolatban: a tavak és műutak pozitív, míg az erdők és nem fertőzött állományok negatív kapcsolatot mutattak. Az eredményekből arra következtethetünk, hogy az erdő csökkenti a fertőzés átvitelét, a tavak, illetve a főutak közelsége pedig növeli az állományok fertőzöttségét.

Szoftver-fejlesztések

maps2WinBUGS

Az alkalmazás segítséget nyújt a GeoBUGS-szal való Bayes-i térbeli elemzések során a felhasználónak a térképek, tabuláris adatok előkészítésében, a szomszédossági mátrixok létrehozásában, térképek manipulációjában, valamint az eredmények vizualizációjában. Emellett három gyakran használt modell futtatásához szükséges BUGS- vagy R-szkript generálását segítendő „varázslót” is beépítettem. Az eszköz lehetőséget nyújt az elemzések eredményeként létrejött térképek exportálására, pixelgrafikus- vagy GIS fájl-formátumokban (ESRI shape vagy Google Earth fájl).

TETYN

Az alkalmazás segítségével egyszerűen lehet a szükséges klimatikus paramétereket kigyűjteni (CRU TS), vagy kiszámítani (TYN SC) térbeli és időbeli lekérdezéseken keresztül. Az eredmény vesszővel határolt állományba (CSV) vagy ESRI shape fájlba menthető, e formátumokat már könnyen lehet felhasználni a további elemzésekhez.

Bibliography

Allepuz, A., Saez, V., Solymosi, N., & Casal, J. (2009). The role of geographical factors on the success of Aujeszky's disease eradication programme in a high pig density area (northeast of Spain, 2003 – 2007). *Prev. Vet. Med. in press.*

Anselin, L. (1992). Local indicators of spatial association – LISA. *Geogr. Anal.*, 27(2), 93–115.

Austin, C. C. & Weigel, R. M. (1992). Factors affecting the geographic distribution of pseudorabies (Aujeszky's disease) virus infection among swine herds in Illinois. *Prev. Vet. Med.*, 13(4), 239–250.

Austin, C. C., Weigel, R. M., Hungerford, L. L., & Biehl, L. G. (1993). Factors affecting the risk of infection with pseudorabies virus in Illinois swine herds. *Prev. Vet. Med.*, 17(3-4), 161–173.

Bak, P. (1996). *How Nature Works: The Science of Self-Organized Criticality.* New York: Copernicus.

Barton, D. E., David, F. N., & Merrington, M. (1965). A criterion for testing contagion in time and space. *Ann. Human Genetics*, 29(1), 97–102.

Berke, O. & Grosse Beilage, E. (2003). Spatial relative risk mapping of pseudorabies-seropositive pig herds in an animal-dense region. *J. Vet. Med. B*, 50(7), 322–325.

Bernardinelli, L., Clayton, D. G., & Montomoli, C. (1995a). Bayesian estimates of disease maps: how important are priors? *Stat. Med.*, 14(21-22), 2411–2431.

Bernardinelli, L., Clayton, D. G., Pascutto, C., Montomoli, C., Ghislandi, M., & Songini, M. (1995b). Bayesian analysis of space-time variation in disease risk. *Stat. Med.*, 14(21-22), 2433–2443.

Besag, J. & Newell, J. (1991). The detection of clusters in rare diseases. *J. R. Stat. Soc. Ser. A Stat. Soc.*, 154(1), 143–155.

Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Ann. Inst. Stat. Mat.*, 43(1), 1–20.

Bjerkdal, T. & Bakketeig, L. S. (1975). Surveillance of congenital malformations and other conditions of the newborn. *Int. J. Epidemiol.*, 4(1), 31–36.

Blancou, J. (1988). Ecology and epidemiology of fox rabies. *Rev. Infect. Dis.*, 10(Suppl. 4), S606–609.

Boelaert, F., Deluyker, H., Maes, D., Godfroid, J., Raskin, A., Varewijck, H., Pensaert, M., Nauwynck, H., Castryck, F., Miry, C., Robijns, J. M., Hoet, B., Segers, E., van Vlaenderen, I., Robert, A., & Koenen, F. (1999). Prevalence of herds with young sows seropositive to pseudorabies (Aujeszky's disease) in northern Belgium. *Prev. Vet. Med.*, 41(4), 239–255.

Brooks, S. & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.*, 7(4), 434–455.

Bánhidy, J., Solymosi, N., Reiczigel, J., & Harnos, A. (2002). A magyarországi veszettség esetek téridő mintázatának analízise. In *VI. Magyar Biometriai és Biomatematikai Konferencia* (pp. 3–4). Budapest, Hungary.

Carpenter, T. E. (2001). Methods to investigate spatial and temporal clustering in veterinary epidemiology. *Prev. Vet. Med.*, 48(4), 303–320.

Casal, J., Moreso, J. M., Planas-Cuchí, E., & Casal, J. (1997). Simulated airborne spread of Aujeszky's disease and foot and mouth disease. *Vet. Rec.*, 140(26), 672–676.

Casimiro, E., Calheiros, J., Santos, F. D., & Kovats, S. (2006). National assessment of human health effects of climate change in Portugal: Approach and key findings. *Environ. Health Perspect.*, 114(12), 1950–1956.

Centers for Disease Control and Prevention (1990). Guidelines for investigating clusters of health events. *MMWR Morb. Mortal. Wkly. Rep.*, 39(RR-11), 17–23.

Chen, R. (1979). Statistical techniques in birth defects surveillance systems. *Contr. Epidemiol. Biostatist.*, 1, 184–189.

Chen, R. (1986). Revised values for the parameters of the sets techniques for monitoring the incidence rate of a rare disease. *Meth. Inform. Med.*, 25(1), 47–49.

Christensen, J. H., Carter, T. R., Rummukainen, M., & Amanatidis, G. (2007). Evaluating the performance and utility of regional climate models: the PRUDENCE project. *Climatic Change*, 81(Suppl. 1), 1–6.

Christensen, L. S., Mortensen, S., Bøtner, A., Strandbygaard, B. S., Rønsholt, L., Henriksen, C. A., & Andersen, J. B. (1993). Further evidence of long distance airborne transmission of Aujeszky's disease (pseudorabies) virus. *Vet. Rec.*, 132(13), 317–321.

Christensen, L. S., Mousing, J., Mortensen, S., Soerensen, K. J., Strandbygaard, S. B., Henriksen, C. A., & Andersen, J. B. (1990). Evidence of long distance airborne transmission of Aujeszky's disease (pseudorabies) virus. *Vet. Rec.*, 127(19), 471–474.

Cuzick, J. & Edwards, R. (1990). Spatial clustering for inhomogeneous populations. *J. R. Statist. Soc. Ser. B*, 52(1), 73–104.

Dale, M. R. T., Dixon, P., Fortin, M.-J., Legendre, P., Myers, D. E., & Rosenberg, M. S. (2002). Conceptual and mathematical relationships among methods for spatial analysis. *Ecography*, 25(5), 558–577.

Diggle, P., Chetwynd, A., Haggkvist, R., & Morris, S. (1995). Second-order analysis of space-time clustering. *Stat. Methods. Med. Res.*, 4(2), 124–136.

Doyle, J. & Carlson, J. M. (2000). Power laws, highly optimized tolerance, and generalized source coding. *Phys. Rev. Lett.*, 84(24), 5656–5659.

Ederer, F., Myers, M. H., & Mantel, N. (1964). A statistical problem in space and time: do leukaemia cases come in clusters? *Biometrics*, 20(3), 626–638.

Englehardt, J. (2002). Scale invariance of incident size distributons in response to sizes of their causes. *Risk Anal*, 22(2), 369–381.

Erdélyi, K., Dencső, L., Szeredi, L., Lehoczki, R., Heltai, M., Sonkoly, K., Csányi, S., & Solymosi, N. (2009). Roe deer fibropapillomatosis, identification of the distribution area and clinico-pathological features of an endemic papillomavirus infection. *Vet. Microbiol.*, 138(1-2), 20–26.

Ewan, W. D. & Kemp, K. W. (1960). Sampling inspection of continuous processes with no autocorrelation between successive results. *Biometrika*, 47(3-4), 363–380.

Farkas, R., Gyurkovszky, M., Solymosi, N., & Beugnet, F. (2009). Studies on flea infestion of dogs and cats combined with a questionnaire survey in Hungary. *Med. Vet. Entomol. in press.*

Flynt, J. W. (1974). Trends in surveillance of congenital malformations. In Janerich, et al. (Ed.), *Congenital Defects: New Directions in Research* (pp. 119–128). New York: Academic Press.

Földvári, G., Márialigeti, M., Solymosi, N., Lukács, Z., Majoros, G., Kósa, J. P., & Farkas, R. (2007). Hard ticks infesting dogs in Hungary and their infection with *Babesia* and *Borrelia* species. *Parasitol. Res.*, 101(Suppl. 1), 25–34.

Galamb, O., Győrffy, B., Sipos, F., Dinya, E., Krenács, T., Berczi, L., Szőke, D., Spisák, S., Solymosi, N., Németh, A., Juhász, M., Molnár, B., & Tulassay, Z. (2008a). *Helicobacter pylori* and antrum erosion-specific gene expression patterns: the discriminative role of CXCL13 and VCAM1 transcripts. *Helicobacter*, 13(2), 112–126.

Galamb, O., Sipos, F., Solymosi, N., Spisák, S., Krenács, T., Tóth, K., Tulassay, Z., & Molnár, B. (2008b). Diagnostic mRNA expression patterns of inflamed, benign, and malignant colorectal biopsy specimen and their correlation with peripheral blood results. *Cancer Epidemiol. Biomarkers Prev.*, 17(10), 2835–2845.

Gambino, J. G. (2005). *pps: Functions for PPS sampling.* R package version 0.94.

Geary, R. (1954). The contiguity ratio and statistical mapping. *Incorp. Statist.*, 5, 115–145.

Gebhardt, F. (1998). *Survey on cluster test for spatial area data.* Technical report, GMD, Stenzelbergweg 8, D-53229 Bonn.

Gloster, J., Donaldson, A. I., & Hough, M. N. (1984). Analysis of series of outbreaks of Aujeszky's disease in Yorkshire in 1981-82: the possibility of airborne disease spread. *Vet. Rec.*, 114(10), 234–239.

Gloster, J., Mellor, P. S., Burgin, L., Sanders, C., & Carpenter, S. (2007). Will bluetongue come on the wind to the United Kingdom in 2007? *Vet. Rec.*, 160(13), 422–426.

Hardy, R. J., Schroder, G. D., Cooper, S. P., Buffler, P. A., Prichard, H. M., & Crane, M. (1990). A surveillance system for assessing health effects from hazardous exposures. *Am. J. Epidemiol.*, 132(Suppl. 1), 32–42.

Harnos, A., Reiczigel, J., Rubel, F., & Solymosi, N. (2006a). Scaling properties of epidemiological time series. *AEER*, 4(2), 151–158.

Harnos, A., Reiczigel, J., Solymosi, N., & Vattay, G. (2006b). Analysis of the effect of immunization in rabies time series. *J. Theor. Biol.*, 240(1), 72–77.

Harnos, A., Solymosi, N., & Vattay, G. (2002). Az immunizació hatásának vizsgálata veszettség idősorokban. In *VI. Magyar Biometriai és Biomatematikai Konferencia* (pp. 28–29). Budapest, Hungary.

Harnos, Z., Gaál, M., & Hufnagel, L. (2008). *Klímaváltozásról mindenkinek.* Budapest: Budapesti Corvinus Egyetem, Kertészettudományi Kar, Matematika és Informatika Tanszék.

Healy, M. J. R. (1968). The disciplining of medical data. *Br. Med. J.*, 24(3), 210–214.

Herszényi, L., Sipos, F., Galamb, O., Solymosi, N., Hritz, I., Miheller, P., Berczi, L., Molnár, B., & Tulassay, Z. (2008). Matrix metalloproteinase-9 expression in the normal mucosa-adenoma-dysplasia-adenocarcinoma sequence of the colon. *Pathol. Oncol. Res.*, 14(1), 31–37.

Hill, G. B., Spicer, C. C., & Weatherall, J. A. C. (1968). The computer surveillance of congenital malformations. *Br. Med. Bull.*, 24(3), 215–218.

Hill, G. B., Spicer, C. C., & Weatherall, J. A. C. (1976). Surveillance of malformations. *Br. Med. Bull.*, 32(1), 39–44.

Jacquez, G. M. (1996). A k nearest neighbour test for space-time interaction. *Stat. Med.*, 15(17-18), 1935–1949.

Jaykus, L.-A., Woolridge, M., Frank, J. M., Miraglia, M., McQuatters-Gollop, A., Tirado, C., Clarke, R., & Friel, M. (2008). Climate change: implications for food safety. Food and Agriculture Organization of the United Nations (FAO).

Kennett, R. & Pollack, M. (1983). On sequential detection of a shift in the probability of a rare event. *J. Am. Statist. Assoc.*, 78(382), 389–395.

Kerekes, B. (1996). Will the silvatic rabies be 'urbanized'? Recent Hungarian data. *Magy. Állatorv. Lapja*, 51(6), 327.

Kerekes, B. (2000). Rabies – Current situation. *Magy. Állatorv. Lapja*, 122(4), 195–200.

Kirkpatrick, C. M., Kanitz, C. L., & McCrocklin, S. M. (1980). Possible role of wild mammals in transmission of pseudorabies to swine. *J. Wildl. Dis.*, 16(4), 601–614.

Knox, G. (1964). The detection of space-time interactions. *Appl. Stat.*, 13, 25–29.

Kulldorff, M. & Nagarwall, N. (1995). Spatial disease clusters: detection and inference. *Stat. Med.*, 14(8), 799–810.

Lawson, A. B. (2001). *Statistical Methods in Spatial Epidemiology.* New York: Wiley.

Lawson, A. B. (2008). *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology.* Interdisciplinary Statistics. Boca Raton, Florida, USA: Chapman & Hall/CRC.

Lawson, A. B., Browne, W. J., & Vidal Rodeiro, C. L. (2003). *Disease Mapping with WinBUGS and MLwiN.* Statistics in Practice. Chichester, England: Wiley.

Leontides, L., Ewald, C., Mortensen, S., & Willeberg, P. (1994). Factors associated with circulation of Aujeszky's disease virus in fattening herds of an intensively vaccinated area of northern Germany. *Prev. Vet. Med.*, 20(1-2), 63–78.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.*, 27(2), 209–220.

Marsh, W. E., Damrongwatanapokin, T., Larntz, K., & Morrison, R. B. (1991). The use of a Geographic Information System in an epidemiological study of pseudorabies (Aujeszky's disease) in minnesota swine herds. *Prev. Vet. Med.*, 11(3-4), 249–254.

McGrath, D. & Zharig, C. (2003). Spatial distribution of soil organic carbon concentrations in grassland of ireland. *Appl. Geochem.*, 18(10), 1629–1639.

Medveczky, I., Kovács, L., Kovács, F., & Papp, L. (1988). The role of the housefly, *Musca domestica*, in the spread of Aujeszky's disease (pseudorabies). *Med. Vet. Entomol.*, 2(1), 81–86.

Mitchell, T. D., Carter, T. R., Jones, P. D., Hulme, M., & New, M. (2004). *A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901-2000) and 16 scenarios (2001-2100).* Working Paper 55, Tyndall Centre.

Mitchell, T. D. & Jones, P. D. (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.*, 25(6), 693–712.

Mocsári, E., Kerekes, B., Heltay, I., Szalay, D., & Csabay, L. (1994). Experiences on the oral vaccination of foxes against rabies in Hungary. *Magy. Állatorv. Lapja*, 49(1), 10–15.

Moorcroft, P. R. & Lewis, M. A. (2006). *Mechanistic Home Range Analysis.* Princeton, New Jersey, USA: Princeton University Press.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1-2), 17–23.

Mészáros, J., Solymosi, N., & Speiser, F. (2007). Spatial distribution of political parties in Hungary 1990-2006. *Polit. Geogr.*, 26(7), 804–823.

Mészáros, J., Solymosi, N., & Speiser, F. P. (2004). Pártok területi eloszálsa. In *Térinformatika 2004, NyME GEO* Székesfehérvár, Hungary.

Mészáros, J., Szakadát, I., Speiser, F. P., & Solymosi, N. (2006). Political and social statistics geodatabase. In *Proceedings of Agile 2006* (pp. 332–338). Visegrád, Hungary.

Nagy, A., Kerekes, B., Heltay, I., Szalay, D., & Csabay, L. (1995). Oral immunization of foxes (wildlife carnivores) - results in Hungary. *Magy. Állatorv. Lapja*, 50(2), 95–100.

Nakicenovic, N., Alcamo, J., Davis, G., de Vries, B., Fenhann, J., Gaffin, S., Gregory, K., Grübler, A., Jung, T. Y., Kram, T., La Rovere, E. L., Michaelis, E., Mori, S., Morita, T., Pepper, W., Pitcher, H., Price, L., Riahi, K., Roehrl, A., Rogner, H., Sankovski, A., Schlesinger, M., Shukla, P., Smith, S., Swart, R., van Rooijen, S., Victor, N., & Dadi, Z. (2000). *Intergovermental Panel on Climate Change Special Report on Emissions Scenarios*. Cambridge: Cambridge Univ. Press.

Naus, J. I. (1965). The distribution of the size of the maximum cluster of points on a line. *J. Am. Statist. Assoc.*, 60, 532–538.

Naus, J. I. (1966). A power comparison of two tests of non-random clustering. *Technometrics*, 8(3), 493–517.

Newman, M. E. J., Girvan, M., & Farmer, J. D. (2002). Optimal design, robustness, and risk aversion. *Phys. Rev. Lett.*, 89, 028301.

Norman, H. S., Sischo, W. M., Pitcher, P., Nesselrodt, A., & Day, R. L. (1996). Spatial and temporal epidemiology of pseudorabies virus infection. *Am. J. Vet. Res.*, 57(11), 1563–1568.

Norstrøm, M. (2001). Geographical Information System (GIS) as a tool in surveillance and monitoring of animal diseases. *Acta Vet. Scand.*, 94(Suppl. 1), 79–85.

Oden, N. (1995). Adjusting Moran's I for population density. *Stat. Med.*, 14(1), 17–26.

Oden, N., Jacquez, G., & Grimson, R. (1996). Realistic power simulations compare point- and area-based disease cluster tests. *Stat. Med.*, 15(7-9), 783–806.

Ohno, Y. & Aoki, K. (1981). Cancer deaths by city and county in Japan (1969-1971): a test of significance for geographic clusters of disease. *Soc. Sci. Med. Med. Geogr.*, 15D(1), 251–258.

Ohno, Y., Aoki, K., & Aoki, N. (1979). A test of significance for geographic clusters of disease. *Int. J. Epidemiol.*, 8(3), 273–280.

Openshaw, S., Charlton, M., Wymer, C., & Craft, A. W. (1987). A Mark 1 geographical analysis machine for the automated analysis of point data sets. *Int. J. GIS*, 1, 335–358.

Overmars, K., de Koning, G., & Veldkamp, A. (2003). Spatial autocorrelation in multi-scale land use models. *Ecol. Model.*, 164(2), 257–270.

Pfeiffer, D. U. (2000). Spatial analysis - a new challenge for veterinary epidemiologist. In M. V. Thrusfield & E. A. Goodall (Eds.), *Proceedings of Annual Meeting of Society for Veterinary Epidemiology and Preventive Medicine* (pp. 86–106).: Society for Veterinary Epidemiology and Preventive Medicine.

Philippe, P. (2000). Epidemiology and self-organized critical systems: An analysis in waiting times and diseases heterogeneity. *Nonlinear Dyn. Psychol. Life Sci.*, 4(4), 275–295.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2007). *coda: Output analysis and diagnostics for MCMC*. R package version 0.12-1.

Potthoff, R. F. & Whittinghill, M. (1966a). Testing for homogeneity. I. The binomial and multinomial distributions. *Biometrika*, 53(1-2).

Potthoff, R. F. & Whittinghill, M. (1966b). Testing for homogeneity. II. The Poisson distribution. *Biometrika*, 53(1-2).

Purse, B. V., Mellor, P. S., Rogers, D. J., Samuel, A. R., Mertens, P. P. C., & Baylis, M. (2005). Climate change and the recent emergence of bluetongue in Europe. *Nat. Rev. Microbiol.*, 3(2), 171–181.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Reiczigel, J., Solymosi, N., Könyves, L., Maróti-Agóts, A., & Kern, A. (2008). Estimation of the effect of heat stress on milk production. In *Deutsche Veterinärmedizinische Gesellschaft* Graz, Austria.

Reiczigel, J., Solymosi, N., Könyves, L., Maróti-Agóts, A., Kern, A., & Bartyik, J. (2009). A hőstressz okozta tejtermelés-kiesés vizsgálata hőmérséklet-páratartalom indexek alkalmazásával. *Magy. Állatorv. Lapja*, 131(3), 137–144.

Rhodes, C. J. & Anderson, R. M. (1996a). Power laws governing epidemics in isolated populations. *Nature*, 381, 600–602.

Rhodes, C. J. & Anderson, R. M. (1996b). A scaling analysis of measles epidemics in a small population. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 351(1348), 1679–1688.

Rhodes, C. J., Butler, A. R., & Anderson, R. M. (1998). Epidemiology of communicable diseases in small population. *J. Mol. Med.*, 76(2), 111–116.

Rhodes, C. J., Jensen, H. J., & Anderson, R. M. (1997). On the critical behaviour of simple epidemics. *Proc. Biol. Sci.*, 264(1388), 1639–1646.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *J. Appl. Probab.*, 13, 255–266.

Ripley, B. D. (1977). Modelling spatial patterns. *J. Roy. Statistical Society, Series B*, 39, 172–212.

Ripley, B. D. (1981). *Spatial Statistics*. New York: John Wiley & Sons.

Rodríguez-Buenfil, J. C., Alvarez-Fleites, M., Alzina-Lopez, A., Arjona-Torres, M. G., Segura-Correa, J. C., & Villegas-Pérez, S. (2002). Risk factors for Aujeszky's disease in pig herds and detection of field virus antibodies in fattening pigs in the state of Yucatan, Mexico. *Prev. Vet. Med.*, 53(3), 205–213.

Rogerson, P. A. (1997). Surveillance systems for monitoring the development of spatial patterns. *Stat. Med.*, 16(18), 2081–2093.

Rosenberg, M. S., Sokal, R. R., Oden, N. L., & DiGiovanni, D. (1999). Spatial autocorrelation of cancer in Western Europe. *Eur. J. Epidemiol.*, 15(1), 15–22.

Sanson, R. L., Pfeiffer, D. U., & Morris, R. S. (1991). Geographic Information Systems: their application in animal disease control. *Rev. Sci. Technol. Off. Int. Epizoot.*, 10(1), 179–195.

Scheidt, B., Rueff, L. R., Grant, R. H., Teclaw, R. F., Hill, M. A., Meyer, K. B., & Clark, L. K. (1991). Epizootic of pseudorabies among ten swine herds. *J. Am. Vet. Med. Assoc.*, 199(6), 725–730.

Shaoxiang, N., Jeffrey, A. L., Yuchun, W., Jiangjun, J., Yong, Z., & Hongliang, Z. (2003). Spatial clustering of rangeland grasshoppers (Orthoptera: Acrididae) in the Qinghai Lake region of northwestern China. *Agricult. Ecosys. Environ.*, 95(1), 61–68.

Siegel, A. M. & Weigel, R. M. (1999). Herd factors affecting the selection and success of intervention strategies in the program for eradication of pseudorabies (Aujeszky's disease) virus from Illinois swine farms. *Prev. Vet. Med.*, 40(3-4), 243–259.

Siegel, A. M., Weigel, R. M., & Taft, A. C. (1993). Factors associated with time under quarantine for swine herds in the voluntary phase of the Illinois pseudorabies eradication program. *Prev. Vet. Med.*, 16(4), 295–310.

Sipos, F., Galamb, O., Herszényi, L., Molnár, B., Solymosi, N., Zágoni, T., Berczi, L., & Tulassay, Z. (2008). Elevated insulin-like growth factor 1 receptor, hepatocyte growth factor receptor and telomerase protein expression in mild ulcerative colitis. *Scand. J. Gastroenterol.*, 43(3), 289–298.

Smith, G. C. & Wilkinson, D. (2003). Modeling control of rabies outbreaks in red fox populations to evaluate culling, vaccination, and vaccination combined with fertility control. *J. Wildl. Dis.*, 39(2), 278–286.

Solymosi, N. (2008). Vektorokkal terjedő kórokozók elterjedtségének predikciója környezeti tényezők alapján. In *Magyar Parazitológusok Társasága, tudományos ülés* Magyar Természettudományi Múzeum, Budapest, Hungary.

Solymosi, N., Harnos, A., & Reiczigel, J. (2003). A Tiszántúlon, állatok között előfordult veszettség esetek térbeli halmozódásának vizsgálata, 1990-2001. In *VIII. Geomatemaikai Ankét* Szeged, Hungary.

Solymosi, N., Harnos, A., Reiczigel, J., & Speiser, F. P. (2006). RpostGIS an R-library for using PostGIS spatial structures and functions. In *Book of Abstracts of The second international R user conference useR! 2006* (pp. 154). Vienna, Austria.

Solymosi, N., Kern, A., Maróti-Agóts, A., Horváth, L., & Erdélyi, K. (2008a). TETYN: An easy to use tool for extracting climatic parameters from Tyndall data sets. *Environ. Modell. Softw.*, 23(7), 948–949.

Solymosi, N., Maróti-Agóts, A., Könyves, L., Kern, A., & Berke, O. (2008b). Climate change scenarios and the projection of geographic risk for heat stress in Hungarian dairy cattle. In *XXV. World Buiatrics Congress* Budapest, Hungary.

Solymosi, N., Maróti-Agóts, A., Ózsvári, L., Könyves, L., Horváth, L., & Kern, A. (2007). Region specific heat stress forecast for cattle production based on climate change scenarios. In *GISVET'07* Coppenhagen, Denmark.

Solymosi, N., Molnár, D. L., & Mészáros, J. (2005). Választási eredmények térbeli-statisztikai elemzése. In J. Mészáros & I. Szakadát (Eds.), *Magyarország Politikai Atlasza 2004*. Gondolat, Budapest. ISBN: 963-9510-01-1.

Solymosi, N., Reiczigel, J., Berke, O., Harnos, A., Szigeti, S., Fodor, L., Szigeti, G., & Bódis, K. (2004a). Spatial risk assessment of herd sero-status of Aujeszky's disease in a county in Hungary. *Prev. Vet. Med.*, 65(1-2), 9–16.

Solymosi, N., Reiczigel, J., Edélyi, K., Bánhidy, J., Földi, Z., & Bódis, K. (2002). *Spatial analysis of rabies cases in foxes in Hungary between 1990 and 2001: a preliminary report*, volume 90 of *Studies in Health Technology and Informatics*, (pp. 770–773). Budapest, Hungary. ISBN: 978-1-58603-279-1.

Solymosi, N., Reiczigel, J., Harnos, A., Rubel, F., & Berke, O. (2004b). Null distribution of Moran's I by stratification. In *GisVet Conference 2004* Guelph, Ontario, Canada.

Stegeman, A., Elbers, A. R. W., van Oirschot, J. T., Hunneman, W. A., Kimman, T. G., & Tielen, M. J. M. (1995). A retrospective study into characteristics associated with the seroprevalence of pseudorabies virus infected breeding pigs in vaccinated herds in the southern Netherlands. *Prev. Vet. Med.*, 22(4), 273–283.

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *J. Stat. Softw.*, 12(3), 1–16.

Szőke, D., Molnár, B., Solymosi, N., Klausz, G., Gyulai, Z., Tóth, B., Mándi, Y., & Tulassay, Z. (2008). T-251A polymorphism of IL-8 relating to the development of histological gastritis and G-308A polymorphism of TNF-α relating to the development of macroscopic erosion. *Eur. J. Gastroen. Hepat.*, 20(3), 191–196.

Szőke, D., Molnár, B., Solymosi, N., Rácz, K., Gergics, P., Blasko, B., Vásarhelyi, B., Vannay, Á., Mándy, Y., Klausz, G., Gyulai, Z., Galamb, O., Spisák, S., Hutkai, B., Somogyi, A., Berta, K., Szabó, A., Tulassay, T., & Tulassay, Z. (2009a). Polymorphisms of the ApoE, HSD3B1, IL-1β and p53 genes are associated with the development of early uremic complications in diabetic patients: Results of a DNA resequencing array study. *Int. J. Mol. Med.*, 23(2), 217–227.

Szőke, D., Molnár, B., Solymosi, N., Sipos, F., Galamb, O., Győrffy, A., & Tulassay, Z. (2009b). The RR genotype of codon 72 of p53 gene reduces the development of intestinal metaplasia. *Dig. Liver Dis.*, 41(3), 179–184.

Tamba, M., Calabrese, R., Finelli, E., & Cordioli, P. (2002). Risk factors for Aujeszky's disease seropositivity of swine herds of a region of northern Italy. *Prev. Vet. Med.*, 54(3), 203–212.

Tango, T. (1995). A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Stat. Med.*, 14(21-22), 2323–2334.

Thulke, H. H., Tischendorf, L., Staubach, C., Selhorst, T., Jeltsch, F., Müller, T., Schlüter, H., & Wissel, C. (2000). The spatio-temporal dynamics of a post-vaccination resurgence of rabies in foxes and emergency vaccination planning. *Prev. Vet. Med.*, 47(1-2), 1–21.

Timbal, B., Fernandez, E., & Li, Z. (2009). Generalization of a statistical downscaling model to provide local climate change projections for Australia. *Environ. Modell. Softw.*, 24(3), 341–358.

Trottier, H. & Philippe, P. (2005). Scaling properties of childhood infectious diseases epidemics before and after mass vaccination in Canada. *J. Theor. Biol.*, 235(3), 326–337.

Turnbull, B. W., Iwano, E. J., Burnett, W. S., Howe, H. L., & Clark, L. C. (1990). Monitoring for clusters of disease: application to leukemia incidence in upstate New York. *Am. J. Epidemiol.*, 132(Suppl. 1), 136–143.

Wagner, S. E., Burch, J. B., Hussey, J., Temples, T., Bolick-Aldrich, S., Mosley-Broughton, C., Liu, Y., & Hebert, J. R. (2009). Soil zinc content, groundwater usage, and prostate cancer incidence in South Carolina. *Cancer Causes Control*, 20(3), 345–353.

Wakefield, J. C., Kelsall, J. E., & Morris, S. E. (2000). Clustering, cluster detection, and spatial variation in risk. In P. Elliott, J. C. Wakefield, N. G. Best, & D. J. Briggs (Eds.), *Spatial epidemiology: methods and applications* (pp. 128–152). New York: Oxford University Press Inc.

Wallenstein, S. (1980). A test for detection of clustering over time. *Am. J. Epidemiol.*, 111(3), 367–371.

Wallenstein, S. & Neff, N. (1987). An approximation for the distribution of the scan statistic. *Stat. Med.*, 6(2), 197–207.

Ward, M. P. & Carpenter, T. E. (2000). Analysis of time-space clustering in veterinary epidemiology. *Prev. Vet. Med.*, 43(4), 225–237.

Weigel, R. M., Austin, C. C., Siegel, A. M., Biehl, L. G., & Taft, A. C. (1991). Risk factors associated with the seroprevalence of pseudorabies virus in Illinois swine herds. *Prev. Vet. Med.*, 12(1-2), 1–13.

Weinstock, M. A. (1981). A generalized scan statistic test for the detection of clusters. *Am. J. Epidemiol.*, 10(3), 289–293.

Weishampel, J. E., Bagley, D. A., Ehrhart, L. M., & Rodenbeck, B. L. (2003). Spatiotemporal patterns of annual sea turtle nesting behaviors along an East Central Florida beach. *Biol. Conserv.*, 110(2), 295–303.

White, J. (2003). An assessment of habitat manipulation as a fox control strategy. Final Report to the National Feral Animal Control Program, Bureau of Rural Science, Australian Government.

Wittmann, E. J. & Baylis, M. (2000). Climate change: effects on culicoides–transmitted viruses and implications for the UK. *Vet. J.*, 160(2), 107–117.

Wright, J. C. & Thawley, D. G. (1980). Role of the raccoon in the transmission of pseudorabies: a field and laboratory investigation. *Am. J. Vet. Res.*, 4(41), 581–583.

Zhang, C. & Selinus, O. (1997). Spatial analyses for copper, lead and zinc contents in sediments of the Yangtze river basin. *Sci. Total Environ.*, 204(3), 251–262.

Zhoua, T., Carlson, J. M., & Doyle, J. (2005). Evolutionary dynamics and highly optimized tolerance. *J. Theor. Biol.*, 236(4), 438–447.

Zienius, D., Sereika, V., & Lelesius, R. (2007). Rabies occurence in red fox and racoon dog population in lithuania. *EKOLOGIJA*, 53(1), 59–64.

Appendix

I.

Spatial risk assessment of herd sero-status of Aujeszky's disease in a county in Hungary

Norbert Solymosi, Jenő Reiczigel, Olaf Berke, Andrea Harnos, Sándor Szigeti, László Fodor, Gábor Szigeti, Katalin Bódis

Preventive Veterinary Medicine 65: 9 – 16

Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Preventive Veterinary Medicine 65 (2004) 9–16
PREVENTIVE VETERINARY MEDICINE
www.elsevier.com/locate/prevetmed

Spatial risk assessment of herd sero-status of Aujeszky's disease in a county in Hungary

Norbert Solymosi[a,*], Jenő Reiczigel[a], Olaf Berke[b,c], Andrea Harnos[a], Sándor Szigeti[d], László Fodor[d], Gábor Szigeti[d], Katalin Bódis[e]

[a]*Department of Biomathematics and Informatics, Faculty of Veterinary Science, Szent István University, István u. 2, H-1078 Budapest, Hungary*
[b]*Department of Population Medicine, Ontario Veterinary College, University of Guelph, Guelph, Ont., Canada N1G 2W1*
[c]*Department of Biometry, Epidemiology and Information Processing, School of Veterinary Medicine Hannover, D-30559 Hannover, Germany*
[d]*Animal Health and Food Inspection Service Office of County Csongrád, Vasas Szent Péter u. 9, H-6724 Szeged, Hungary*
[e]*Department of Physical Geography and Geoinformatics, University of Szeged, Egyetem u. 2-6, H-6722 Szeged, Hungary*

Received 27 February 2003; received in revised form 13 July 2004; accepted 13 July 2004

Abstract

A Geographic Information System (VetEpiGIS) was used to analyze the ADV (Aujeszky's disease virus) sero-status in large-scale pig units regarding certain geographical features in a county of southern Hungary. The ADV sero-statuses were collected from all swine units in Csongrád county in 1998–2000. The units' coordinates were combined with a vector graphical digital map of the county, with a resolution of 1:100,000. Logistic regression tested the associations between sero-status of large-scale units and presence of topographical features, other units and villages in the neighborhood. ''Neighborhood'' was defined by circular zones with radius 1–10 km around the unit (in 1 km increments; one logistic regression for each radius). The following topographical features showed significant positive association with the ADV seropositivity: lake (3 km OR: 5.7; 5 km OR: 7.5; 6 km OR: 6.1; 10 km OR: 5.4) and highway (5 km OR: 4.2; 6 km OR: 5.3). Other features had negative association with ADV seropositivity: forest (3 km OR: 0.13; 4 km OR: 0.15; 5 km OR: 0.15; 6 km

* Corresponding author. Tel.: +36 1 478 42 13; fax: +36 1 478 42 17.
E-mail address: solymosinorbert@freemail.hu (N. Solymosi).

0167-5877/$ – see front matter
doi:10.1016/j.prevetmed.2004.07.006

OR: 0.10; 7 km OR: 0.10; 8 km OR: 0.23) and uninfected large-scale unit (4 km OR: 0.07; 5 km OR: 0.27; 6 km OR: 0.32; 7 km OR: 0.31).

Keywords: Aujeszky's disease; Spatial epidemiology; Odds ratio; Logistic regression

1. Introduction

The Aujeszky's disease virus (ADV) belongs to the family of Herpesviridae which establish a lifelong persistent infection with intermittent shedding of the virus. The infection can cause clinical disease with neurological symptoms in suckling and weaned pigs and respiratory or reproductive disorders in grower-finisher and adult animals. Pigs are the only natural host of ADV. Other animals found on farms (such as rats and mice) are dead-end hosts (Berke and Beilage, 2003). The main route of transmission of ADV among swine populations is the physical contact between infected and susceptible animals. Less important for the transmission are fomites such as feed, vehicles or artificial insemination. Other routes of infection of susceptible populations are long-distance airborne transmission (Gloster et al., 1984; Christensen et al., 1990, 1993; Casal et al., 1997) and vector-mediated transmission (Wright and Thawley, 1980; Kirkpatrick et al., 1980). The distances among swine populations and the surrounding topographical features might be important for the last two routes. The health status of the pig population and the success in eradication of Aujeszky's disease (AD) are influenced by the geographic population density (Weigel et al., 1991; Austin and Weigel, 1992; Austin et al., 1993; Siegel et al., 1993; Leontides et al., 1994; Stegeman et al., 1995; Norman et al., 1996; Boelaert et al., 1999; Siegel and Weigel, 1999; Rodríguez-Buenfil et al., 2002; Tamba et al., 2002). Nevertheless, the effect of the geographical features on the spread of infection has been a rarely investigated topic. Marsh et al. (1991) investigated the distance to the nearest county road, highway, quarantined herd, river or lake and the density of swine herds within a 5 km radius as possible risk factors; high density of pig herds within 5 km was found as a possible risk-increasing factor whilst lakes or rivers within 1 km seemed protective. The distance between pig units and topographical characteristics of the region can influence the effectiveness of AD eradication programs (Scheidt et al., 1991). Our aim was to investigate the association between sero-status of pig units and presence of certain geographical features as well as large-scale swine units and villages (including small towns) in their neighborhood.

2. Material and methods

2.1. Data

Data of all known pig units (except pure fattening units) registered in Csongrád county in 1998–2000 were used in the study. Csongrád county is located on the Great Plain, in the southeastern part of Hungary. The neighbouring counties are Bács-Kiskun, Békés, Jász-

Table 1
Number of large- and small-scale pig units and villages in the study of Aujeszky's disease herd sero-status in a county in Hungary

	1998	1999	2000
Number of villages in which there were some small-scale pig units (total number of small-scale units)	77 ([a])	60 ([a])	71 (5345)
Large-scale pig units	33	39	44
Percentage of large-scale units infected	30	26	18

[a] Total number of small-scale units not available.

Nagykun-Szolnok, whereas it is bounded in the south by state border with Romania and Serbia-and-Montenegro. This is a rural area; the Tisza river divides it into two parts. Pig production is important in the economy of the county. Swine units are classified as large- and small-scale units. Small-scale units represent traditional pig-keepers having a few pigs in the backyard; according to animal-health regulations, the upper limit is 100 pigs. Large-scale units are located outside villages. The number of units and villages is displayed in Table 1. The ADV sero-status is known for all units from the results of the yearly ADV serological screening of sows organized by the Animal Health and Food Inspection Service. The geographic coordinates of the large-scale pig units were determined by a Garmin eMap GPS-receiver. The determination of the coordinates was made from May to August 2000 with an accuracy of 10–15 m. The geographic coordinates were appointed in the WGS84 coordinate system (the available GPS-receiver worked with this one) and these data were transformed into the Uniform National Projection (EOV) system (http://lazarus.elte.hu/gb/geodez/geodind.htm). The precise locations of small-scale units within villages were not determined because we regarded each village as one epidemiological unit. The reason for this is that the small-scale units are close together and the hygiene barriers between them typically are limited. A village was regarded as infected if there was at least one infected unit within the village.

2.2. Software

A civil vector graphical digital map of Hungary (OTAB1) was used with a resolution of 1:100,000. The map contains the following layers: villages and towns, large-scale pig units, lakes, rivers, small natural waterways and artificial canals, forests, and four types of roads from unpaved road to controlled-access high-speed motorway. The projection system is EOV. This map was integrated into a database-management software (VetEpiGIS) developed by the first author. The software was built by using the Visual Basic 6.0 Professional Edition environment (Microsoft Corp.) with a MSDE 1.0 (Microsoft Corp.) as database engine through ADO and MapX 4.51 (MapInfo Corp.) for mapping.

2.3. Analysis

Buffer generation (Sanson et al., 1991; Norstrøm, 2001) was applied to define circular zones around large-scale swine units and to determine the presence of topographical features in these zones. The radius of the zone was increased from 1 to 10 km in 1−km steps

as Casal et al. (1997) suggested. The feature (like, e.g. forest) was counted as being present in the zone if an instance of the feature had any presence with the buffer zone.

Data of all known swine units from years 1998 to 2000 were analyzed. Because large-scale units were the same in all 3 years (with a few exceptions due to reorganization, etc.), and sometimes neither their status nor the features in their neighborhood changed, our data contained duplicate records. These duplicates were deleted from analysis (i.e. just one record of the identical cases was kept). However, if the status of the unit changed due to infection or eradication, and/or there were changes in the neighborhood, we used data from both or all years.

Data were analyzed by stepwise logistic regression using the backward conditional method by SPSS 8.0. The dependent variable was status (1 if seropositive, 0 if not). Explanatory variables were indicators of the topographical features, seropositive/seronegative large-scale units, and seropositive/seronegative villages in the zone around the unit (1 if feature/unit/village is present, 0 if not). No interaction term was included in the model. This analysis was repeated for each zone radius. The reported ORs are those obtained after removing the non-significant ($P > 0.05$, by Wald's test) factors from the model. Multicollinearity was checked calculating the variance inflation factors, which were <1.5 (except for ''seropositive village'' and ''seronegative village'', for which it was 2.5 and 2.1 at zone radius of 4 and 5 km, respectively) indicating no serious collinearity. We also checked the correlations between the estimated regression coefficients. Highly correlated estimates were found only in case of the 4 km buffer zone for ''lake'' and ''forest'' ($r > 0.999$) therefore, we removed ''lake'' from the equation. For all other zone radii, all $|r|$ were ≤ 0.53.

3. Results

Table 2 shows the frequencies of presence of the topographical features, seropositive and seronegative large-scale units and villages in the zones around the units. Features for which the frequency was 0 or 100% for all zones (e.g. motorway, one kind of road, and artificial canal) were dropped.

Table 3 shows the results of the logistic regression analysis. The first two factors in the table (''lake'' and ''highway'') are positively associated with positive sero-status, whereas the other two factors (''forest'' and ''seronegative large-scale unit'') are negatively associated with it.

4. Discussion

In their study, Marsh et al. (1991) concluded that a distance <5 km between pig units might increase the risk of disease spread. However, their data covered just 41% of the known pig units in the study area. Norman et al. (1996) showed that within 3.22 km, a high density of non-quarantined herds or low density of quarantined herds is risk decreasing — whereas low density of vaccinated herds is a risk-increasing factor. These results were based on data of about 75% of the farms in the study area. Similar results hold for Csongrád

Table 2
Percentage of seropositive (P) and seronegative (N) large-scale swine units, for which the listed features occurred within the given distance (Aujeszy's disease in a county in Hungary)

Zone radius (km)	River		Lake		Forest		Highway		Unpaved road		Channel		Infected large-scale unit		Uninfected large-scale unit		Infected village		Uninfected village	
	P	N	P	N	P	N	P	N	P	N	P	N	P	N	P	N	P	N	P	N
1	0	6	8	2	15	33	15	25	69	58	31	50	0	0	0	13	8	23	0	27
2	0	20	19	11	19	55	25	30	81	68	75	89	0	0	0	34	38	36	38	39
3	0	20	40	15	25	66	45	53	85	84	85	90	0	0	0	44	50	41	45	44
4	9	37	55	19	36	72	68	59	86	97	100	99	9	3	9	49	50	42	50	49
5	24	37	60	24	60	78	80	65	100	100	100	100	12	13	28	62	52	47	60	58
6	39	40	58	31	62	81	81	66	100	100	100	100	31	23	50	71	65	59	58	69
7	35	39	57	42	61	94	83	69	100	100	100	100	39	31	48	76	74	66	65	72
8	35	39	57	49	78	94	91	73	100	100	100	100	39	33	65	79	74	72	74	73
9	36	43	73	60	86	100	91	76	100	100	100	100	41	40	64	79	73	72	77	75
10	48	44	91	64	91	100	91	83	100	100	100	100	48	49	76	85	76	73	76	77

Table 3
Final logistic multiple-regression models of risk factors (within indicated radii) for seropositivity of swine units for Aujeszky's disease in a county in Hungary, 1998–2000

Model and risk factors	*b*	S.E.(*b*)	*P*	OR	95% CI (OR)
3 km radius (deviance = 73.37, d.f. = 2, model $P = 0.0002$)					
Forest	−2.0	0.65	0.002	0.13	0.037, 0.47
Lake	1.7	0.69	0.011	5.7	1.5, 22
Constant	−0.68	0.38			
4 km radius (deviance = 74.33, d.f. = 2, model $P < 0.0001$)					
Forest	−1.89	0.60	0.002	0.15	0.047, 0.49
Uninf. LSC[a]	−2.6	0.83	0.002	0.073	0.014, 0.37
Constant	0.64	0.47			
5 km radius (deviance = 82.96, d.f. = 4, model $P < 0.0001$)					
Forest	−1.9	0.68	0.006	0.15	0.039, 0.57
Highway	1.4	0.71	0.043	4.2	1.0, 17
Lake	2.0	0.65	0.002	7.5	2.1, 27
Uninf. LSC[a]	−1.3	0.58	0.023	0.27	0.084, 0.83
Constant	−1.1	0.69			
6 km radius (deviance = 88.15, d.f. = 4, model $P = 0.0001$)					
Forest	−2.3	0.70	0.001	0.099	0.025, 0.39
Highway	1.7	0.67	0.014	5.3	1.4, 20
Lake	1.8	0.61	0.003	6.1	1.8, 20
Uninf. LSC[a]	−1.1	0.55	0.038	0.32	0.11, 0.94
Constant	−0.70	0.69			
7 km radius (deviance = 85.97, d.f. = 2, model $P = 0.0001$)					
Forest	−2.3	0.69	0.001	0.097	0.025, 0.38
Uninf. LSC[a]	−1.2	0.55	0.031	0.31	0.10, 0.90
Constant	1.5	0.72			
8 km radius (deviance = 98.18, d.f. = 1, model $P = 0.0422$)					
Forest	−1.5	0.72	0.041	0.23	0.056, 0.94
Constant	0.22	0.67			
10 km radius (deviance = 89.78, d.f. = 1, model $P = 0.0115$)					
Lake	1.7	0.79	0.031	5.4	1.2, 25
Constant	−2.5	0.74			

[a] Uninfected large-scale swine unit.

county with radii of 4–7 km. In another publication, the farms within 2.5 km seemed to be a risk factor (Rodríguez-Buenfil et al., 2002). Tamba et al. (2002) found that if there are at least 10,000 pigs within 6 km, the risk increases. In our study, the proximity of infected villages seemed associated with seropositivity (Tables 2 and 3). We suppose that workers living in neighboring villages acted as vectors between their own herds and the swine unit they were working at but the possibility of air-born spread also exists.

While the results from Csongrád county suggest that lakes close to pig units increase the risk of seropositivity, Marsh et al. (1991) indicated that the nearby presence of lakes reduces the risk of infection spreading. We speculate that more abundant vegetation around the lake increases the risk of infection through feral swine or wildlife vectors (e.g. rodents,

foxes, birds). Another interpretation is that the fog and higher humidity increases the possibility of airborne virus transmission. On the other hand, lakes and rivers might act as barriers (allowing no animal or vector movements). It is possible that in one study area, this effect dominates the other one while in another study area it is the other way round.

Another result of our study indicates that the presence of a highway, might increase the risk but Marsh et al. (1991) could not detect association between ADV status and the nearest county road or highway. Csongrád county data lead to the hypothesis that forests might decrease the risk of infection transmission (perhaps by decreasing air-borne transmission).

Our results must be interpreted with caution, because the investigation area is rather small and we did not correct for edge effects (arising from spatial censoring at the borders; Lawson, 2001). On the other hand, a larger study area would have a higher chance to be inhomogeneous making it difficult to draw conclusions valid for the whole area.

5. Conclusion

Presence of a lake or highway showed positive association with seropositive status of the unit, while in case of forests and uninfected farms the association was negative.

Acknowledgements

First of all we would say thanks to István Medveczky†, who encouraged us dealing with this research topic. The authors appreciate the support of György Elefánti†, Katalin Faragó, Attila Farkas, Sándor Hajdu, István Kovács, Kornél Körffy, Mária Müller, Zoltán Pap, Szilárd Pinnyey, Zoltán Rácz, Éva Rankl, István Szabó, Tibor Szántó, László Szűts, Ms. Anikó Kiss and the administrators of district national animal health offices in Csongrád county, colleagues of Department of Surgery and Ophthalmology, Faculty of Veterinary Science, Szent István University, György Velák (LandInfo Ltd.). This work was partly supported by grants OTKA T035150 and T032437.

References

Austin, C.C., Weigel, R.M., 1992. Factors affecting the geographic distribution of pseudorabies (Aujeszky's disease) virus infection among swine herds in Illinois. Prev. Vet. Med. 13, 239–250.

Austin, C.C., Weigel, R.M., Hungerford, L.L., Biehl, L.G., 1993. Factors affecting the risk of infection with pseudorabies virus in Illinois swine herds. Prev. Vet. Med. 17, 161–173.

Berke, O., Grosse, B.E., 2003. Spatial relative risk mapping of pseudorabies-seropositive pig herds in an animal-dense region. J. Vet. Med. B 50, 322–325.

Boelaert, F., Deluyker, H., Maes, D., Godfroid, J., Raskin, A., Varewijck, H., Pensaert, M., Nauwynck, H., Castryck, F., Miry, C., Robijns, J.M., Hoet, B., Segers, E., van Vlaenderen, I., Robert, A., Koenen, F., 1999. Prevalence of herds with young sows seropositive to pseudorabies (Aujeszky's disease) in northern Belgium. Prev. Vet. Med. 41, 239–255.

Casal, J., Moreso, J.M., Planas-Cuchí, E., Casal, J., 1997. Simulated airborne spread of Aujeszky's disease and foot and mouth disease. Vet. Rec. 140, 672–676.

Christensen, L.S., Mousing, J., Mortensen, S., Soerensen, K.J., Strandbygaard, S.B., Henriksen, C.A., Andersen, J.B., 1990. Evidence of long distance airborne transmission of Aujeszky's disease (pseudorabies) virus. Vet. Rec. 127, 471–474.

Christensen, L.S., Mortensen, S., Bøtner, A., Strandbygaard, B.S., Rønsholt, L., Henriksen, C.A., Andersen, J.B., 1993. Further evidence of long distance airborne transmission of Aujeszky's disease (pseudorabies) virus. Vet. Rec. 132, 317–321.

Gloster, J., Donaldson, A.I., Hough, M.N., 1984. Analysis of series of outbreaks of Aujeszky's disease in Yorkshire in 1981–82: the possibility of airborne disease spread. Vet. Rec. 114, 234–239.

Kirkpatrick, C.M., Kanitz, C.L., McCrocklin, S.M., 1980. Possible role of wild mammals in transmission of pseudorabies to swine. J. Wildl. Dis. 16, 601–614.

Lawson, A.B., 2001. Statistical Methods in Spatial Epidemiology, Wiley, New York.

Leontides, L., Ewald, C., Mortensen, S., Willeberg, P., 1994. Factors associated with circulation of Aujeszky's disease virus in fattening herds of an intensively vaccinated area of northern Germany. Prev. Vet. Med. 20, 63–78.

Marsh, W.E., Damrongwatanapokin, T., Larntz, K., Morrison, R.B., 1991. The use of a Geographic Information System in an epidemiological study of pseudorabies (Aujeszky's disease) in Minnesota swine herds. Prev. Vet. Med. 11, 249–254.

Norman, H.S., Sischo, W.M., Pitcher, P., Nesselrodt, A., Day, R.L., 1996. Spatial and temporal epidemiology of pseudorabies virus infection. Am. J. Vet. Res. 57, 1563–1568.

Norstrøm, M., 2001. Geographical Information System (GIS) as a tool in surveillance and monitoring of animal diseases. Acta Vet. Scand. Suppl. 94, 79–85.

Rodríguez-Buenfil, J.C., Alvarez-Fleites, M., Alzina-Lopez, A., Arjona-Torres, M.G., Segura-Correa, J.C., Villegas-Pérez, S., 2002. Risk factors for Aujeszky's disease in pig herds and detection of field virus antibodies in fattening pigs in the state of Yucatan, Mexico. Prev. Vet. Med. 53, 205–213.

Sanson, R.L., Pfeiffer, D.U., Morris, R.S., 1991. Geographic Information Systems: their application in animal disease control. Rev. Sci. Technol. Off. Int. Epizoot. 10, 179–195.

Scheidt, B., Rueff, L.R., Grant, R.H., Teclaw, R.F., Hill, M.A., Meyer, K.B., Clark, L.K., 1991. Epizootic of pseudorabies among ten swine herds. J. Am. Vet. Med. Assoc. 199 (6), 725–730.

Siegel, A.M., Weigel, R.M., Taft, A.C., 1993. Factors associated with time under quarantine for swine herds in the voluntary phase of the Illinois pseudorabies eradication program. Prev. Vet. Med. 16, 295–310.

Siegel, A.M., Weigel, R.M., 1999. Herd factors affecting the selection and success of intervention strategies in the program for eradication of pseudorabies (Aujeszky's disease) virus from Illinois swine farms. Prev. Vet. Med. 40, 243–259.

Stegeman, A., Elbers, A.R.W., van Oirschot, J.T., Hunneman, W.A., Kimman, T.G., Tielen, M.J.M., 1995. A retrospective study into characteristics associated with the seroprevalence of pseudorabies virus infected breeding pigs in vaccinated herds in the southern Netherlands. Prev. Vet. Med. 22, 273–283.

Tamba, M., Calabrese, R., Finelli, E., Cordioli, P., 2002. Risk factors for Aujeszky's disease seropositivity of swine herds of a region of northern Italy. Prev. Vet. Med. 54, 203–212.

Weigel, R.M., Austin, C.C., Siegel, A.M., Biehl, L.G., Taft, A.C., 1991. Risk factors associated with the seroprevalence of pseudorabies virus in Illinois swine herds. Prev. Vet. Med. 12, 1–13.

Wright, J.C., Thawley, D.G., 1980. Role of the Raccoon in the transmission of pseudorabies: a field and laboratory investigation. Am. J. Vet. Res. 41, 581–583.

II.

Analysis of the effect of immunization in rabies time series

Andrea Harnos, Jenő Reiczigel, *Norbert Solymosi*,

Gábor Vattay

Journal of Theoretical Biology 240: 72 – 77

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 240 (2006) 72–77
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Analysis of the effect of immunization in rabies time series

Andrea Harnos[a,*], Jenő Reiczigel[a], Norbert Solymosi[a], Gábor Vattay[b]

[a]Department of Biomathematics and Informatics, Faculty of Veterinary Science, Szent István University, H-1078 István u. 2., Budapest, Hungary
[b]Department of the Physics of Complex Systems, Eötvös University, H-1117 Pázmány Péter sétány 1/A, Budapest, Hungary

Received 3 March 2005; received in revised form 10 August 2005; accepted 29 August 2005
Available online 3 October 2005

Abstract

The authors analysed time series of rabies cases diagnosed in Hungary between 1967 and 2001. In Transdanubia (West Hungary), an oral immunization program started in 1992 and in East Hungary in 2001. Both long term and seasonal trends were identified in the time series of rabies cases. In order to characterize the underlying processes governing the behaviour of the epidemic, the fluctuations around the trend were analysed before and after immunization separately. It turned out that the tail of the complementary cumulative distribution functions differ. The tail of the distribution follows an inverse power law (IPL) function and describes the distribution of extreme events. The significant difference in the IPL exponents before and after immunization can be explained by the theory of Highly Optimized Tolerance (HOT).

Keywords: Epidemiology; Immunization; Rabies; Time-series analysis; Highly optimized tolerance

1. Introduction

In Hungary only the urban type of rabies—spread by dogs—existed until 1954. The sylvatic type rabies—spread by foxes—appeared in 1954 coming from the North–East. The epidemic propagated at a speed of 50–60 km/year until rabies became an endemic disease in 1970 (Mocsári et al., 1994).

It is convenient to handle separately the western and eastern parts of Hungary because they are separated by the Danube river, a natural barrier for the spread of infection in Hungary, as it is quite unlikely for the foxes to pass through. Fig. 1a shows the location of rabies cases in Hungary in 1992, when a major immunization campaign started to eliminate the disease.

The oral vaccination campaign of the red foxes had several steps. The first campaign was carried out in October 1992 covering an area of 5000 km^2 in the North-West part of Transdanubia (along the Austrian border). In April 1993, this area was expanded to 6000 km^2 (Nagy et al., 1995). Then the vaccinated area was increased to 7000 km^2 in April 1995, to 10,000 km^2 in October 1995 and to 15,000 km^2 in April 1996 (Kerekes, 1996). From October 1996 until 2000 the whole western part of the country (38,000 km^2), and finally in 2001 also the central part of the country (region between the rivers Danube and Tisza) was covered by the program (Kerekes, 2000). Fig. 1b shows rabies cases in 1998, after the immunization in Transdanubia.

The goal of this paper is to analyse the fox rabies cases time series from 1967 to 2001 and to identify its main characteristics. We are particularly interested in the changes due to the oral immunization campaign and motivated to show that it not only affects the average number of cases but also the way the number of cases fluctuate. We show that the distribution of extremely large fluctuations develops an inverse power law (IPL) tail, a characteristic feature in recent epidemiological models. IPL has been proposed by (Rhodes and Anderson, 1996a, b; Rhodes et al., 1997, 1998) for describing time series from small and large vaccinated human populations. Trottier and Philippe (2005) applied this model for measles, rubella, pertussis and mumps outbreaks in Canada. IPL is a function with no characteristic scale ad self-similar upon

*Corresponding author. Tel.: +36 1 478 42 14; fax: +36 1 478 42 17.
E-mail addresses: Harnos.Andrea@aotk.szie.hu (A. Harnos), Reiczigel.Jeno@aotk.szie.hu (J. Reiczigel), Solymosi.Norbert@aotk.szie.hu (N. Solymosi), vattay@complex.elte.hu (G. Vattay).

0022-5193/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2005.08.026

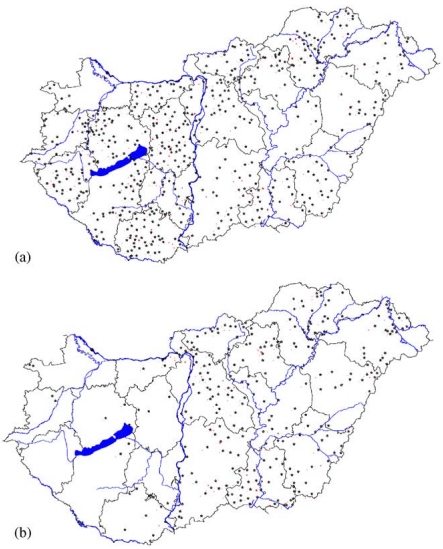


Fig. 1. (a) Rabies cases in Hungary in 1992 (before immunization); (b) rabies cases in Hungary in 1998 (after immunization in Transdanubia).

rescaling (scale invariance), and it may fit to the extreme values of the distribution (power law tail). This can explain the co-existence of small and very large epidemics. Large outbreaks are expected from this type of distributions (Philippe, 2000).

The outline of the paper is as follows. In Section 2, we give a summary of our data and the software used in the analysis. In Section 3, we identify the long time trends and the seasonal variations in the time series. In Section 4, we analyse and compare the extreme fluctuations in the periods before and after immunization. In Section 5, we discuss the results and outline possible explanations for the findings of Section 4, in terms of Highly Optimized Tolerance (HOT) introduced by Doyle and Carlson (2000). In Section 6, we summarize our findings.

2. Materials and methods

2.1. Data

Data have been collected from the rabies case registry of the Animal Health and Food Control Department of the Hungarian Ministry of Agriculture. The resulting database contains data of all documented rabies cases for the period of 1990–2001, including the location and date of occurrences as well as the species affected. In addition to this detailed database, we have monthly count data for the period of 1967–1990 for every county. We constructed monthly summarized time series from the entire data set (1967–2001) for the western (Transdanubia), and for the eastern part (Eastern Hungary) of the country. In Fig. 2, we show these two time series, which we are going to analyse in the rest of the paper.

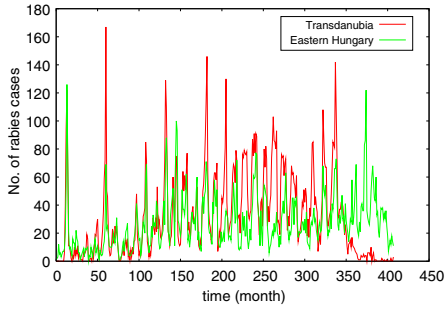


Fig. 2. Monthly number of rabies cases in Transdanubia and in Eastern Hungary (1967–2001).

2.2. Software

The database was developed using the Microsoft Office XP Professional Edition Microsoft Access software. We performed the analysis with Microsoft Excel XP, R and Gnuplot programs.

3. Descriptive analysis

In the monthly rabies cases time series we could identify three main components. There is a deterministic trend on long time scales (several years), then a 12-month seasonal periodicity on medium scale and finally a superimposed monthly random like fluctuation on short time scales. These fluctuations are related to the stochastic factors governing the processes. Our first goal was to identify the deterministic trends and the seasonal periodicity in the time series. Then we removed the deterministic trends and the seasonal periodicity and analysed the fluctuations with statistical methods, especially the extremely large fluctuations.

3.1. Annual trends

During the years, the average number of cases changed significantly (see Fig. 2.) In the Transdanubia data there was a growing trend until 1989 (month 256 in our data set starting in January 1967) followed by a decreasing trend until the start of the immunization in 1996 (month 348). In Fig. 3, we show for each year the monthly average $\tilde{f}(j) = \sum_{i=1}^{12} f(i,j)/12$, where $f(i,j)$ is the number of cases in month i of year j.

It is clear that the number of cases gradually increases from the mid-sixties to the mid-eighties in both parts of the country and then saturates. In Transdanubia it starts declining sharply from 1996 due to a mass immunization campaign.

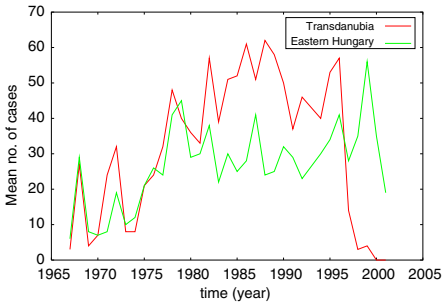


Fig. 3. Monthly average of rabies cases in each year for Transdanubia (red) and Eastern Hungary (green).

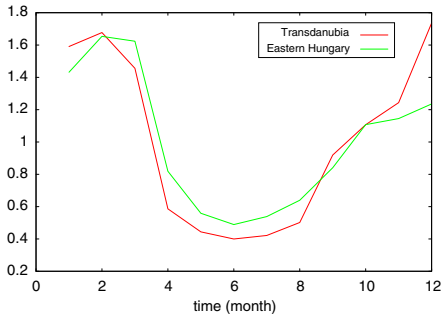


Fig. 4. Seasonal trends in Transdanubia and Eastern Hungary (the value 1 represents the average of all years).

3.2. Seasonal trends

Next, we concentrate on the seasonal variation of the number of cases. To be able to compare years with different average number of monthly cases we divided the monthly values in each year by the yearly average shown in Fig. 3. Thus, after dividing the monthly frequency by the yearly average, we get the detrended time series $e(i,j) = f(i,j)/\bar{f}(j)$. This time series shows a 12-month seasonality due to yearly variation of the external conditions. In Fig. 4, we show for each month the average number of cases calculated from all 34 years $\bar{e}(i) = \frac{1}{34}\sum_{j=1}^{34} e(i,j), i = 1,\ldots,12$. We can see that in January–March the number of cases is about 150% of the yearly average and in November and December it is also somewhat above the average. This is in accordance with other studies (Thulke et al., 2000) where this variation is attributed to the elevated number of contacts between foxes during periods of mating and dispersal periods.

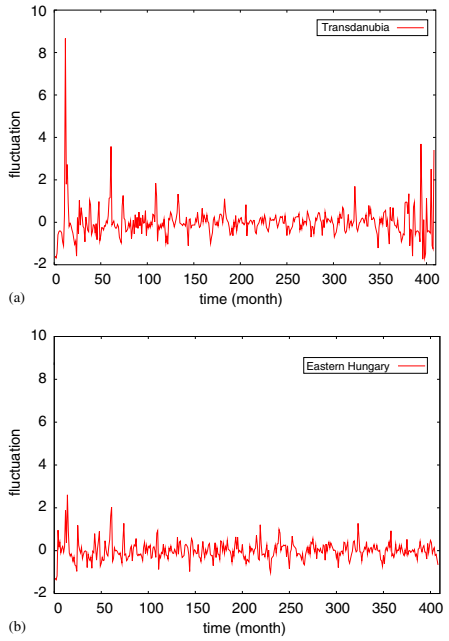


Fig. 5. (a) Monthly fluctuations in Transdanubia, (b) monthly fluctuations in Eastern Hungary.

3.3. Monthly random fluctuations

The monthly random fluctuations can be uncovered by removing the seasonal trend from the detrended time series $e_f(i,j) = e(i,j) - \bar{e}(i)$. In Figs. 5a and b we show these fluctuations for Transdanubia and for Eastern Hungary. These time series are approximately stationary (Harnos et al., 2002), except the beginning of both time series and the end of the Transdanubia time series. At the beginning, the relatively large fluctuations are related to the increasing trend in that period. The Transdanubia time series shows some large relative fluctuations at the end due to the decreasing trend caused by the immunization.

Our next goal is to understand the statistics of extremely large fluctuations related to the immunization.

4. Analysis of extreme fluctuations

To analyse the large fluctuations identified in the previous section we refine our methods. Previously, we considered fluctuations relative to the long term and the seasonal trends. This method cannot be used if the average of the time series changes rapidly. In such cases it is more

useful to compare the number of actual cases to a moving average over a preceding period. In our case, data show 12-month seasonal variations, so it is natural to consider a 12-month moving average and values relative to the moving average $l(i) = f(i)/(\sum_{j=1}^{12} f(i-j)/12)$, where $f(i)$ is the number of cases in the ith month in the series. The best way to study the statistics of the extremely large outbreaks is to consider the complementary probability distribution $F_c(l) = Prob\{l(i) > l\}$, which gives the probability that the relative fluctuation is larger than l. In Fig. 6, we show the complementary cumulative distribution for the Transdanubia time series including data only before the start of the immunization campaign in 1992.

Next we show that this distribution develops an IPL tail. The method of fitting is called scaling analysis and it consists of fitting an IPL function $F_c(l) = al^b$ via fitting a linear function on the logarithmic scale $\log F_c(l) = \log a + b \log l$, where l is the epidemic size (in our case normalized with the moving average), a and b (a negative number) are estimable parameters of the epidemic size distribution. On Fig. 6, we show the IPL fitted to the Transdanubia data before immunization.

To study the effect of immunization we show in Fig. 7, the complementary distribution for the Transdanubia time series including the immunized cases. The main effect of immunization on the statistics of large outbreaks is the change of the scaling exponent b of the IPL from about -1.5 before immunization to 2.0 after immunization.

In the next section, we will argue that this change in the exponent is related to the change of geometry of the spread of disease during the immunization campaigns. This can be understood within the framework of HOT introduced recently by Doyle and Carlson (2000), and also applied in a theoretical biology context by Zhoua et al. (2005).

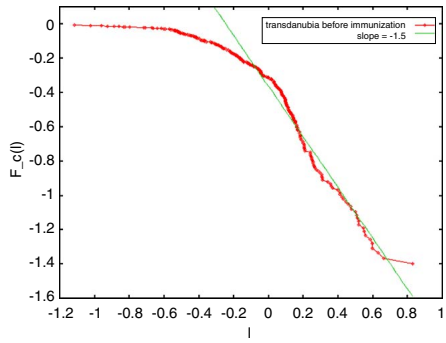


Fig. 6. The complementary cumulative distribution of the fluctuations relative to the 12 months moving average for the Transdanubia data before immunization on a doubly logarithmic plot. The straight line represents the fitted power law tail with exponent $b = -1.5$.

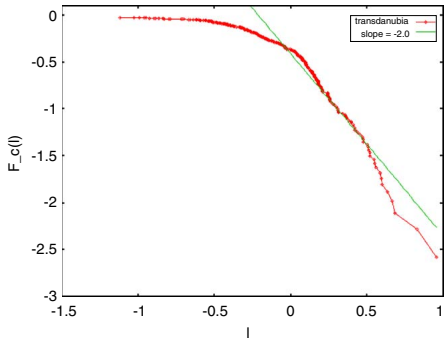


Fig. 7. The complementary cumulative distribution of the full Transdanubia data set on a doubly logarithmic plot and the fitted power law tail with exponent $b = -2$.

5. Power law tails and highly optimized tolerance in epidemiology

5.1. Phase transitions, SOC and HOT: power law tails in complex systems

Distributions with power law tails arise in many natural and engineered systems. During the last three decades much effort has been made to explain their widespread occurrence. Systems influenced by many independent random factors in their environment would more likely behave randomly dictated by the law of large numbers and the central limit theorem. It is then always puzzling why such systems so often violate these simple concepts. To understand the mechanism behind power laws the key ideas came first from statistical physics. In conventional physics power laws arise at special points of the parameter space of systems. These are critical points. Systems at their critical points go over phase transitions and show very strong fluctuations obeying power laws. However, this mechanism does not provide a sufficient explanation since critical points are isolated in the parameter space. It is very unlikely to find a large number of systems in nature that are accidentally tuned to their critical points. The next step was to explain why systems naturally evolve into their critical points and stay there. Simple systems are not able to do this spontaneously. In complex systems this can happen more easily as they are able to adapt to external conditions via self-organization. At the beginning of the eighties the concept of self-organized criticality (SOC) emerged (Bak, 1996), where it was shown that many complex systems naturally seek their critical states and go into a self-organized critical state without external tuning of their parameters. This explanation can account for a very large class of natural systems showing power law distributed fluctuations.

During the last couple of years it became clear that in certain systems—engineered or designed by humans or by evolutionary forces—distributions with power law tails can arise from another reason. The goal of evolution and/or design is usually to improve the efficiency of a system and to protect it from failures at the same time. The theory of HOT (Doyle and Carlson, 2000) claims that in systems strongly optimized to avoid failures the size distribution of failures naturally develops power tails. The basic idea behind HOT theory can be best demonstrated by an idealized forest fire model. In this model one should find the optimal distribution of the trees on a grid so as to maximize tree harvest in the face of occasional fires that burn complete connected clusters of trees and are started by sparks that arrive with a given spatial distribution. It turned out that optimizing the harvest gives rise to a segmented forest consisting of contiguous patches of trees separated by firebreaks, and that the resulting distribution of fire sizes follows a power law. Newman et al. (2002) showed that the tail of the complementary cumulative distribution behaves as an IPL $F_c(l) \propto l^{-1-1/\beta}$, where the parameter β is intimately related to the geometry of the problem. For a d dimensional forest fire model with optimized barriers $\beta = d$ holds, resulting in an IPL exponent $b = -1 - 1/d$.

5.2. Application of the HOT concept in epidemiology and for the rabies data

The HOT forest fire model can also be applied in epidemiology: the animal population plays the role of forest and the "fire" is the epidemic itself. Animal populations are often separated by natural barriers such as mountains and rivers and epidemics break out in the separated domains when infected animals migrate into the area ("sparks"). The difference between the forest fire model and its epidemiological counterpart is that unlike trees, animals are able to move and also the barriers to the spreading of the disease can be more complex. Beyond natural barriers or artificial quarantines the spread of disease can be blocked by the presence of naturally or artificially immunized animal populations. In such cases the effective dimensionality of the problem may be different from its spatial dimensionality.

Since our Rabies data set involves an immunization experiment it gives us a unique opportunity to test the epidemiological applicability of the HOT model. In case of unimmunized data we expect that the fluctuations are determined by the spatial dimension ($d = 2$) of the problem and we expect that the complementary cumulative distribution of the fluctuations is going to have a power law tail with exponent predicted by the HOT theory: $b = -(1 + 1/d) = -1.5$. One can see that the data collected at Transdanubia before the immunization supports our concept as the tail of the distribution is clearly consistent with an exponent –1.5. On the other hand, immunization changes dramatically the scaling behaviour of the tail of the Transdanubia data set involving the immunization, the full Transdanubia data set scales with an exponent near $b = -2$. This means that the effective dimensionality of our problem changes to $d = 1$. Fig. 8 displays the two distributions in a single plot. The bodies of the distributions are almost the same and the change in the slope of the tail is clearly visible.

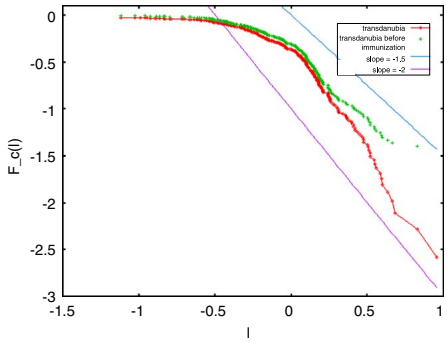


Fig. 8. The two distributions (Figs. 6 and 7) in a single plot.

It is easy to understand the change in the effective dimensionality of the problem. Before immunization the disease can spread in two spatial dimensions as animals move around. After immunization large patches of immunized populations exist and the disease can spread only at the quasi one-dimensional borders separating large immunized populations where unimmunized populations can still exist.

6. Discussion

As a zoonosis and an incurable disease, Rabies has always been given due respect both in human and veterinary medicine. Most rabies cases in Hungary are still diagnosed in red foxes (*Vulpes vulpes*). This is in accordance with the European situation and confirms the role of the red fox as the reservoir and primary perpetrator of the disease. In this paper, we concentrated on extreme fluctuations in rabies cases. Extreme fluctuations are low-probability, high-consequence events (Englehardt, 2002) representing the majority of total losses. Understanding the size distribution of extreme events makes it possible to assess the risks of outbreaks.

We determined the distributions of extreme fluctuations which is a scale invariant power law distribution for both time series investigated. Scale invariance in the distributions of the sizes of fluctuations in complex dynamical systems has been explained on the basis of mechanical models of natural and engineered systems, such as models of SOC or HOT. In our case we could find an HOT theory explanation.

Owing to the fact that the exponent of the power law is related to the dimensionality of the process, we could show how immunization changed the structure of the problem. This kind of analysis can help to estimate the effects of eradication programs and to asses the risk of epidemic outbreaks. We think that HOT theory gives a reasonable explanation for the value of the IPL exponent and for the change in it. In the future, we plan to investigate the spatio-temporal distribution of outbreaks in order to verify the HOT-based explanation in the details of the process.

Acknowledgments

This work was partly supported by the Grants OTKA T049157 and T 037903.

References

Bak, P., 1996. How Nature Works: The Science of Self-Organized Criticality. Copernicus, New York.

Doyle, J., Carlson, J.M., 2000. Power laws, highly optimized tolerance, and generalized source coding. Phys. Rev. Lett. 84, 5656.

Englehardt, J.D., 2002. Scale Invariance of incident size distributons in response to sizes of their causes. Risk Anal. 22, 2.

Harnos, A., Solymosi, N., Vattay, G., 2002. Investigation of the effect of immunization in rabies time series. Abstracts of the 6th Hungarian Coference in Biomathematics and Biometrics, Budapest, August 26–27, p. 28 (in Hungarian).

Kerekes, B., 1996. Will the silvatic rabies be "urbanized"? recent Hungarian data. Hung. Vet. J. 51, 327.

Kerekes, B., 2000. Rabies—current situation. Hung. Vet. J. 122, 195–200.

Mocsári, E., Kerekes, B., Heltay, I., Szalay, D., Csabay, L., 1994. Experiences on the oral vaccination of foxes against rabies in Hungary. Hung. Vet. J. 49, 10–15.

Nagy, A., Kerekes, B., Heltay, I., Szalay, D., Csabay, L., 1995. Oral immunization of foxes (wildlife carnivores)—results in Hungary. Hung. Vet. J. 50, 95–100.

Newman, M.E.J., Girvan, M., Farmer, J.D., 2002. Optimal design, robustness, and risk aversion. Phys. Rev. Lett. 89, 028301.

Philippe, P., 2000. Epidemiology and self-organized critical systems: An analysis in waiting times and diseases heterogeneity. Nonlinear Dyn. Psychol. Life Sci. 4 (4), 275–295.

Rhodes, C.J., Anderson, R.M., 1996a. A scaling analysis of measles epidemics in a small population. Philos. Trans. R. Soc. Lond. B Biol. Sci. 351, 1679–1688.

Rhodes, C.J., Anderson, R.M., 1996b. Power laws governing epidemics in isolated populations. Nature 381, 600–602.

Rhodes, C.J., Jensen, H.J., Anderson, R.M., 1997. On the critical behaviour of simple epidemics. Proc. R. Soc. Lond. B Biol. Sci. 264, 1639–1646.

Rhodes, C.J., Butler, A.R., Anderson, R.M., 1998. Epidemiology of communicable diseases in small population. J. Mol. Med. 76, 111–116.

Thulke, H.-H., Tischendorf, L., Staubach, Ch., Selhorst, T., Jeltsch, F., Müller, T., Schluéter, H., Wissel, Ch., 2000. The spatio-temporal dynamics of a post-vaccination resurgence of rabies in foxes and emergency vaccination planning. Prev. Vet. Med. 47, 1–21.

Trottier, H., Philippe, P., 2005. Scaling properties of childhood infectious diseases epidemics before and after mass vaccination in Canada. J. Theor. Biol. 235, 326–337.

Zhoua, T., Carlson, J.M., Doyle, J., 2005. Evolutionary dynamics and highly optimized tolerance. J. Theor. Biol. 236, 438–447.

III.

TETYN: An easy to use tool for extracting climatic parameters from Tyndall data sets

Norbert Solymosi, Anikó Kern, Ákos Maróti-Agóts, Levente Horváth, Károly Erdélyi

Environmental Modelling & Software 23: 948 – 949

Available online at www.sciencedirect.com
ScienceDirect
Environmental Modelling & Software 23 (2008) 948–949
Environmental Modelling & Software
www.elsevier.com/locate/envsoft

Software, Data and Modelling News

TETYN: An easy to use tool for extracting climatic parameters from Tyndall data sets

Norbert Solymosi [a,*], Anikó Kern [a], Ákos Maróti-Agóts [b], Levente Horváth [a], Károly Erdélyi [c]

[a] Adaptation to Climate Change Research Group, Hungarian Academy of Sciences, Budapest, Hungary
[b] Department of Animal Breeding, Nutrition & Laboratory Animal Science, Faculty of Veterinary Science, Szent István University, Budapest, Hungary
[c] Central Agricultural Office, Veterinary Diagnostic Directorate, Budapest, Hungary

Received 8 October 2007; received in revised form 5 November 2007; accepted 21 November 2007
Available online 3 January 2008

Abstract

The recent and rapid change in climate seems to have strong impact on many aspects of agriculture, health, ecology, economy and the society. To model these impacts researchers need access to future and past climate databases, some of which are publicly available. One possible source of climate data sets is the collection of the Tyndall Centre for Climate Change Research. We developed an easy to use tool to obtain required climatic parameters from the Tyndall future (TYN SC) and past (CRU TS) data sets. Results of the query can be exported as comma separated value or as ESRI shape files.

Keywords: Spatial query; Climate change; Spatial analogy; Data set; Risk assessment

Software availability

Software name: Tetyn (Version 1.00)
Developers: Norbert Solymosi
Contact address: solymosi.norbert@gmail.com
Year first available: 2007
Software required: Windows XP
Program language: Microsoft Visual Studio Express C# 2005
Availability: http://sourceforge.net/projects/tetyn/
Cost: Free of charge

1. Software description

In recent years the importance of climate change has been rapidly growing in most disciplines, including agriculture, health, ecology, economy and social science. Many studies deal with the necessity of adaptation to climate change. Thus, researchers and decision makers need adequate data sources to analyze, model or forecast consequences of these changes. In response to such a need, various institutions and organizations produced data sets of a projected future climate.

For researchers working on near-term and far-term modelling the Tyndall Centre for Climate Change Research produced two data sets based on different climate scenarios: the TYN SC 1.0 and the TYN SC 2.0 (Mitchell et al., 2004; Mitchell and Jones, 2005). The first one is projected onto a 10-min spatial resolution grid over Europe, while the second is projected onto the whole globe with a 0.5° grid. These data sets contain future climate data comprising five parameters (cloud cover, diurnal temperature range, precipitation, temperature, and vapour pressure) on a monthly basis. Beside these data sets the Tyndall Centre also provides the CRU TS data sets of the observed monthly meteorological data in the 20th century containing nine parameters (cloud cover, diurnal temperature range, ground frost frequency, precipitation, temperature, minimum temperature, maximum temperature, vapour

* Corresponding author. Department of Biomathematics and Informatics, 1078 István u. 2., Budapest, Hungary. Tel.: +36 14784213; fax: +36 14784217.
E-mail address: solymosi.norbert@gmail.com (N. Solymosi).

1364-8152/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2007.11.006

pressure, and wet day frequency). The CRU TS is also useful for those researchers, who are trying to find analogies from past climate patterns which are then applicable to the future climate of certain geographical areas. The CRU TS data set is on the same grid as TYN SC 2.0.

The TYN SC data sets are free on request. The newest version of CRU TS data set is also freely downloadable.

Although some Fortran programs are available for managing the TYN SC data sets, their use may prove difficulties for a non-expert user. TETYN was developed with the aim of making it easier to extract certain parts from the above described data sets. The climatic data sets contain precursors for the different parameters that are actually calculated from these precursors by an algorithm based on the equation published by Mitchell et al. (2004). The results of TETYN queries can be saved as comma separated values (CSV) or ESRI shape files. In both formats the records represent the gridboxes and columns containing the monthly data.

The most prominent advantages of using TETYN over other tools are as follows.

- TETYN does not require programming abilities and knowledge of the various data structures. The user can set up a complex query using an easy to use graphical interface.
- Results of the queries are easy to import into spreadsheet management software, and GIS tools.
- TETYN allows the user to build data queries with both temporal and spatial aspects. Spatial queries can be performed by one geolocation, bounding box, Tyndall grid coordinates and by choosing one or more countries depending on interest.

Acknowledgements

Financial support was received from the Hungarian Scientific Research Fund (OTKA T049157) and the NKFP 6-00079/2005 KLIMAKKT project.

References

Mitchell, T.D., Carter, T.R., Jones, P.D., Hulme, M., New, M., 2004. A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901–2000) and 16 scenarios (2001–2100). Tyndall Centre for Climate Change Research, Working Paper 55. http://www.tyndall.ac.uk/publications/working_papers/wp55.pdf.

Mitchell, T.D., Jones, P.D., 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. International Journal of Climatology 25, 693–712.

IV.

maps2WinBUGS: a helper tool for Bayesian spatial analyses

Norbert Solymosi, Sara E. Wagner, Ferenc Speiser,

Alberto Allepuz

Journal of Statistical Software, submitted: 2009.02.11, revision required

Journal of Statistical Software

MMMMMM YYYY, Volume VV, Code Snippet II. http://www.jstatsoft.org/

maps2WinBUGS: a helper tool for Bayesian spatial analyses

Norbert Solymosi **Sara E. Wagner** **Ferenc Speiser** **Alberto Allepuz Palau**

Abstract

For Bayesian spatial modeling, **GeoBUGS** is a widely used tool incorporated in **WinBUGS** and **OpenBUGS**. The map format that **GeoBUGS** own map formatuses differs from the standard formats used in geographical information systems (GIS). The presented tool, **maps2WinBUGS** helps the user to prepare maps and tabular data to be used in **GeoBUGS**. With this tool one can obtain adjacency lists, manipulate maps and visualize the results of runs. By the script wizard one can generate **BUGS** or **R** scripts using three different modell based on own map and data. The resulting maps can be exported into different image and GIS file formats (ESRI shape and Google Earth files).

Keywords: Bayes, spatial, mapping, GIS, BUGS project, R project.

1. Introduction

Bayesian methods for spatial modeling and analysis are being increasingly used in different research fields (e.g., epidemiology, ecology, economy and politology). In Bayesian spatial modeling, perhaps the most widely used tool is **GeoBUGS** (Thomas, Best, Lunn, Arnold, and Spiegelhalter 2007), which is included in **WinBUGS** (Lunn, Thomas, Best, and Spiegelhalter 2000) and **OpenBUGS** (Thomas, O'Hara, Ligges, and Sturtz 2006). These applications are very flexible and user-friendly for modeling on Bayesian bases.

One source of **GeoBUGS** data for modelling is a map, which plays dual role. A map can be used easily and quickly to display the results of runs. Also, for some model, that use spatial relation structure, **GeoBUGS** can generate an adjacency matrix. **GeoBUGS** can import maps in ArcInfo, Epimap or S-Plus format. In its present form, **GeoBUGS** can not read and write certain map formats (eg. ESRI shape file) which are used in Geograhic Information Systems (GIS). Exporting results is also important and **GeoBUGS** can only export the resulting maps in pixel graphical or a map in S-Plus format.

Some tools are avaiable to transform maps to **GeoBUGS** format: the function `sp2WB` of package

maptools can transform `SpatialPolygons` objects to the S-Plus map in **R** (R Development Core Team 2008), or the ArcGIS extension of *W. Thogmartin* and colleagues what helps users to develop adjacency matrices.[1].

For some users a tool, that facilitates interaction with **GeoBUGS**, in the preparation of maps, data and scripts, and in the visualization and exportation of results would be useful. Our suggestion for such a tool is **map2WinBUGS** (Lawson 2008). Our helper tool **maps2WinBUGS**[2] runs only under Microsoft Windows operating system.

2. Program purpose

The aim of the development of **maps2WinBUGS** was to produce an open source, free, relatively easy-to-use tool for preparation maps and tabular data for spatial modeling with **GeoBUGS**. Besides this main goal, some additional functions were implemented for the manipulation of maps, visualization of results, and exportation of maps and tables.

In many cases the Bayesian modeler does not develop a completely new model, but uses a well known model with his own data. Recently, a *Script Wizard* was implemented into the tool to help users generate script for **BUGS** or **R** (using **R2WinBUGS**) based on his own map and data.

3. Development

The tool was developed in the Microsoft Visual Studio Express C# 2008 environment[3] with **SharpMap**[4], **NetTopologySuite**[5] component for mapping and spatial functions. **FWTools**[6] dlls are used for MapInfo tab format import. To manage temporary data **SQLite**[7] database engine was used. For generation random numbers from different distributions as initial values **Repast.NET**[8] library was included.

4. Tools and facilities in maps2WinBUGS

The main window of the graphical user interface(GUI) of **maps2WinBUGS** is the Map window (Figure 1), and all functions start from this form. The main module of the form is a **SharpMap** component for interactive visualization of maps and data. The window has a menu interface (detailed description can be found in Appendix A). It also has a limited GIS function toolstrip, with buttons for *Full Extents*, *Zoom In*, *Zoom Out* and *Pan*.

In the Help menu, Flash demos assist the user to learn the various functions.

[1] http://www.umesc.usgs.gov/management/dss/adjacency_tool.html
[2] http://maps2winbugs.sourceforge.net/
[3] http://www.microsoft.com/express/
[4] http://www.codeplex.com/SharpMap
[5] http://nts.sourceforge.net/
[6] http://fwtools.maptools.org/
[7] http://www.sqlite.org/
[8] http://repast.cvs.sourceforge.net/viewvc/repast/repast.net/

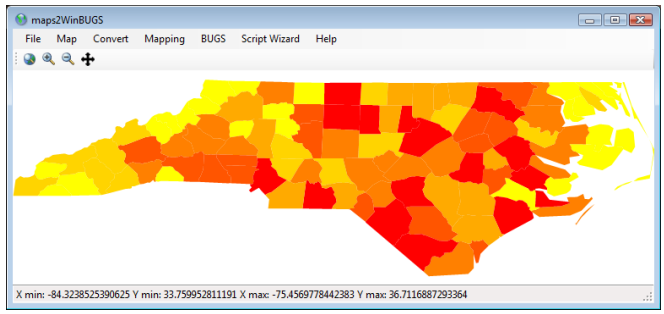


Figure 1: Map window with a thematic map

4.1. Data preparation and script generation for BUGS

Our tool, **maps2WinBUGS** can open different map formats: ESRI shapefiles, MapInfo tab files, PostgreSQL-PostGIS, SQLite spatial tables and from clipboard S-Plus. One can open comma separated value (CSV) as well. From CSV files, a point based map can be produced if the fields contain X and Y coordinates. In this case, the user must designate which field contains the longitude and latitude data. The CSV table can also be transformed into BUGS readable S-Plus format (list or rectangular). Opened maps appear in the Map window.

Functions on map

Useful manipulation can be done on the open map to prepare it for usage in **GeoBUGS**.

Neighbouring List Most spatial models needs a description of neighboring relationships of geographical fetures (e.g. polygons). Although **GeoBUGS** can calculate adjacency matrices for map polygons, **maps2WinBUGS** has three different calculation method for this task including: touches, intersetions, and within distance. In the latter case, a threshold must be setup for calculation on the same scale as the map projection. It is also possible to calculate neighboring list not just for polygons, but for points as well with this method.

Polygon simplifying The maximum number of points in a certain polygon at the importation step of **GeoBUGS** is limited. In some cases the user can have maps with polygons with points over the limit (10,000 points). In **maps2WinBUGS**, the point number can be decreased by simplifying function. This procedure is based on Douglas-Peucker algorithm (Douglas and Peucker 1973). The user must set up the distance tolerance threshold for the method.

Buffering Our tool can also create a buffer around all of the geographical features (point, line, polygon) with a predefined radius. It can be useful if a user has a point map and one needs to calculate an adjacency matrix based on intersections.

Spatial prediction To make spatial prediction in GeoBUGS one should have a grid of th study region, and the centroids of the gridboxes. To prepare these data one can generate a grid over a certain region (polygons) and extract the coordinate of the central location of the gridboxes. The grid as any map can be converted to **GeoBUGS** format, and the centroid list

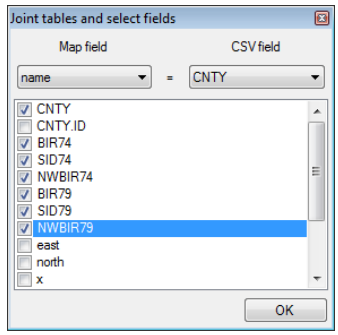


Figure 2: Setting up the key field for merging attribute and map dataset

is generated in S-Plus list format.

Attach data to map In some cases the attribute data is in a different order than the spatial feature data. For example, the attribute data was typed in the order of county name, but the polygons are in the order of their creation. If one puts together datasets ordered in a different manner without any record identification key, it may cause serious mistakes. By the function *Attach data* one can perform this linkage by a joining field (Figure 2). The spatial, and attribute data then can be exported in the same order to **WinBUGS**. Alternatively, one can attach **BUGS** Node Statistics to the map, in this case the linking function is based *not on any key field*, but on the order of the records.

Conversion

The map can be converted into **GeoBUGS** importable ArcInfo and S-Plus formats. The result of the conversion is a text what can be copied into an **WinBUGS** document and imported as a map.

From a map it is also possible to generate not just spatial information for **BUGS**, but to produce S-Plus formats (list, rectangular) from attribute data which is stored in the map. This can be done not just from the map, but from CSV files as well.

Script Wizard

For individuals not familiar with script language, the writing of a script may create many syntactic problems. During the writing of the **BUGS** code a difficult task is joining the source data to a code with suitable variable names. This task can create some problems.

The wizard in the case of implemented models (*Poisson-Gamma*, *Log-Normal*, *Besag-York-Mollié*) helps the user to generate scripts based on her own dataset (Figure 3). The generated scripts should be considered as starting code for modeling, and can be modified and re-parameterized by the user. In the case of two first model there is no need for maps, because they do not use spatial structure. But in the case of *Besag, York and Mollié* (BYM) model, a user must open a map before starting the wizard.

The resulting scripts can be produced for **BUGS** or **R**. In the latter case, the generated script call the **WinBUGS** by the **R2WinBUGS** package.

The inits are random samples from certain distributions according to the model (Gelman and Hill 2006). In the case of **R** script for the inits are pooled by **R**-functions. In the case of **BUGS** the inits are generated by the **Repast.NET** library.

4.2. Mapping

The result of analyses or other numerical variables stored in the attribute data of map can be visualized as a thematic map. Since the spatial reference feature must be polygons it is called choropleth mapping. To set up mapping variable, cut-points and colors for mapping the *Choropleth mapping* window (Figure 4) is used. From the *Classification field* combobox one can select a field for mapping. The *Class No.* option allows a user to set the number of categories in which to divide the selected variable values. The colors of categories can be set up one by one using colours name (the same as in **R**). Also, it is possible to use schemas from predefined liear gradient scales, or to create linear scales between two or among three colours.

4.3. Exporting maps

The map in the map window can be exported in different format to use in other tools or just for save the results. The ESRI shape file can be reused in different systems like GIS or **R**.

The map can be exported into a SQLite database as a spatial table, which is reusable in **maps2WinBUGS** or in **R** (Solymosi, Harnos, and Reiczigel 2008).

When exporting in Google Earth[9] format, one can easily share an interactive thematic map (Figure 5). In this case the user can set up layer name, labeling variable, polygon and line opacity and line color. Choosing this option generates a KML file and the legend PNG file is compressed into a KMZ file. When opened in Google Earth the map is only reasonable if the original map was in the Lat/Lon WGS84 projection system, otherwise the polygons will be positioned on an incorrect global location.

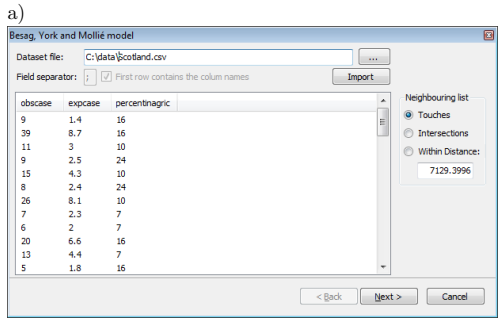
As pixel graphical image also exportable the map in BMP, GIF, JPEG, PNG, TIFF or WMF format.

5. Acknowledgements

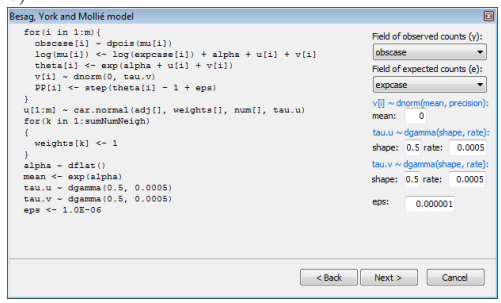
The development was supported by the Hungarian Scientific Research Fund (OTKAT049157).

[9] http://earth.google.com/

a)



b)



c)

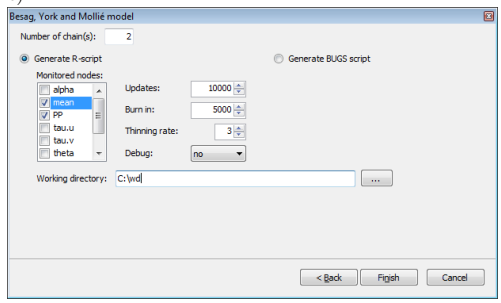


Figure 3: The script wizard for Besag, York and Mollié model (Lawson *et al.* 2003): setting up the source data and adjacency properties (a); the source variables and parameters (b); the script output (c)

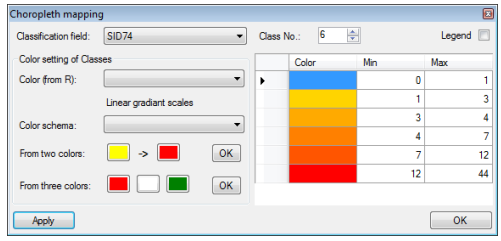


Figure 4: Form for setting up the classification variable, class number with cut-points and colors for thematic mapping

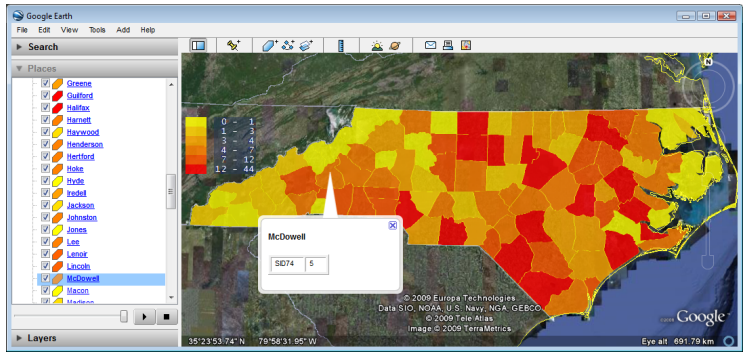


Figure 5: Exported map and legend in Google Earth

References

Douglas DH, Peucker TK (1973). "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature." *The Canadian Cartographer*, **10**(2), 112–122. URL http://utpjournals.metapress.com/content/fm576770u75u7727/.

Gelman A, Hill J (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press. ISBN 052168689X. URL http://www.stat.columbia.edu/~gelman/arm/.

Lawson AB (2008). *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Interdisciplinary Statistics. Chapman & Hall/CRC, Boca Raton, Florida, USA. ISBN 978-1-58488-840-7.

Lawson AB, Browne WJ, Vidal Rodeiro CL (2003). *Disease Mapping with WinBUGS and MLwiN*. Statistics in Practice. Wiley, Chichester, England. ISBN 0-470-85604-1.

Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000). "WinBUGS – a Bayesian modelling framework: Concepts, structure, and extensibility." *Statistics and Computing*, **10**(4), 325–337. ISSN 0960-3174. URL http://dx.doi.org/10.1023/A:1008929526011.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Solymosi N, Harnos A, Reiczigel J (2008). "SQLiteMap: package to manage vector graphical maps using SQLite." In "The R User Conference 2008, Book of Abstracts," Technische Universität Dortmund, Germany.

Thomas A, Best N, Lunn DJ, Arnold R, Spiegelhalter D (2007). *GeoBUGS User Manual*. URL http://www.mrc-bsu.cam.ac.uk/bugs.

Thomas A, O'Hara B, Ligges U, Sturtz S (2006). "Making BUGS Open." *R News*, **6**(1), 12–17.

A. The menu of the maps2WinBUGS graphical user interface

This appendix describes the menu of the **maps2WinBUGS** GUI that operates as an executable under the Microsoft Windows operating system.

File

- *Import Map*
  - *Vector Graphical files*
  - PostGIS table
  - *SQLite table*
  - *S-Plus format map*
- *Open CSV file*
- *Save as*
  - *ESRI Shape file*
  - *Google Earth KML file*
  - *SQLite table*
  - *Image*
- *Exit*

Map

- *Neighboring List*
  - *Touches*
  - *Intersections*
  - *Within Distance*
- *Simplify Polygons*
- *Buffering*
- *Generate Grid*
- *Extract Centroids*
- *Attach data*

Convert Convert the map into **GeoBUGS** importable map formats, or into data table.

- *GeoBUGS*
  - *ArcInfo format* Convert the map into ArcInfo format.
  - *S-Plus format* Convert the map into S-Plus format.

- *Data table* All fields out of the spatial structure are extracted from the map into a table. In the table every record represents data belonging to one geographical feature. Table can be exported to CSV or S-Plus (list, rectangular) formats. Latter is useasble in **WinBUGS**.

Mapping

- *Choropleth* Produce a thematic map based on any field of map using arbitrary cut-points and color schemes.
- *Legend* Present the legend of thematic map.

BUGS

- *WinBUGS* Starts the previously by the settings given WinBUGS executable file.
- *OpenBUGS* Starts the previously by the settings given OpenBUGS executable file.
- *Import BUGS Node Statistics* By copy & paste import BUGS Node Statistics into a table what can be exported into CSV or S-Plus (list, rectangular) formats.
- *Settings* Set the paths to WinBUGS and OpenBUGS executable files.

Script Wizard

- *Poisson-Gamma model*
- *Log-Normal model*
- *Besag, York and Mollié model*

Help

- *Demos* For the main functions are Flash tutorials using simple examples.
  - *Import*
    - *ESRI shape file*
    - *MapInfo tab file*
    - *S-Plus format*
    - *CSV file*
    - *PostGIS table*
    - *SQLite table*
  - *Attach data to map*
    - *BUGS node statistics*
    - *CSV file*
  - *Convert map*
    - *GeoBUGS*
      - *ArcInfo format*

        - *S-Plus format*
      - *Data table*
  - *Mapping*
  - *Calculate neighbouring list*
  - *Generate buffer*
    - *Around points*
    - *Around polygons*
  - *Generate grid*
  - *Simlify polygons*
  - *Script Wizard*
    - *Besag, York and Mollié model*
  - *Save as*
    - *ESRI shape file*
    - *Image*
    - *Google Earth format*
    - *SQLite table*
- *About*

Affiliation:

Norbert Solymosi
Adaptation to Climate Change Research Group, Hungarian Academy of Science
Department of Biomathematics and Informatics
Faculty of Veterinary Science, Szent István University
1078 István u. 2, Budapest, Hungary
E-mail: Solymosi.Norbert@gmail.com
URL: http://klima.univet.hu/solymosi

Sara E. Wagner
Cancer Prevention and Control Program
University of South Carolina, Columbia, SC, USA

Ferenc Speiser
Department of Automation, University of Pannonia, Veszprém, Hungary

Alberto Allepuz Palau
Centre de Recerca en Sanitat Animal (CReSA), UAB-IRTA
Campus de la Universitat Autónoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume VV, Code Snippet II *Submitted:* 2009-02-11
MMMMMM YYYY *Accepted:* yyyy-mm-dd

B

Maps and table of relative risk of rabies

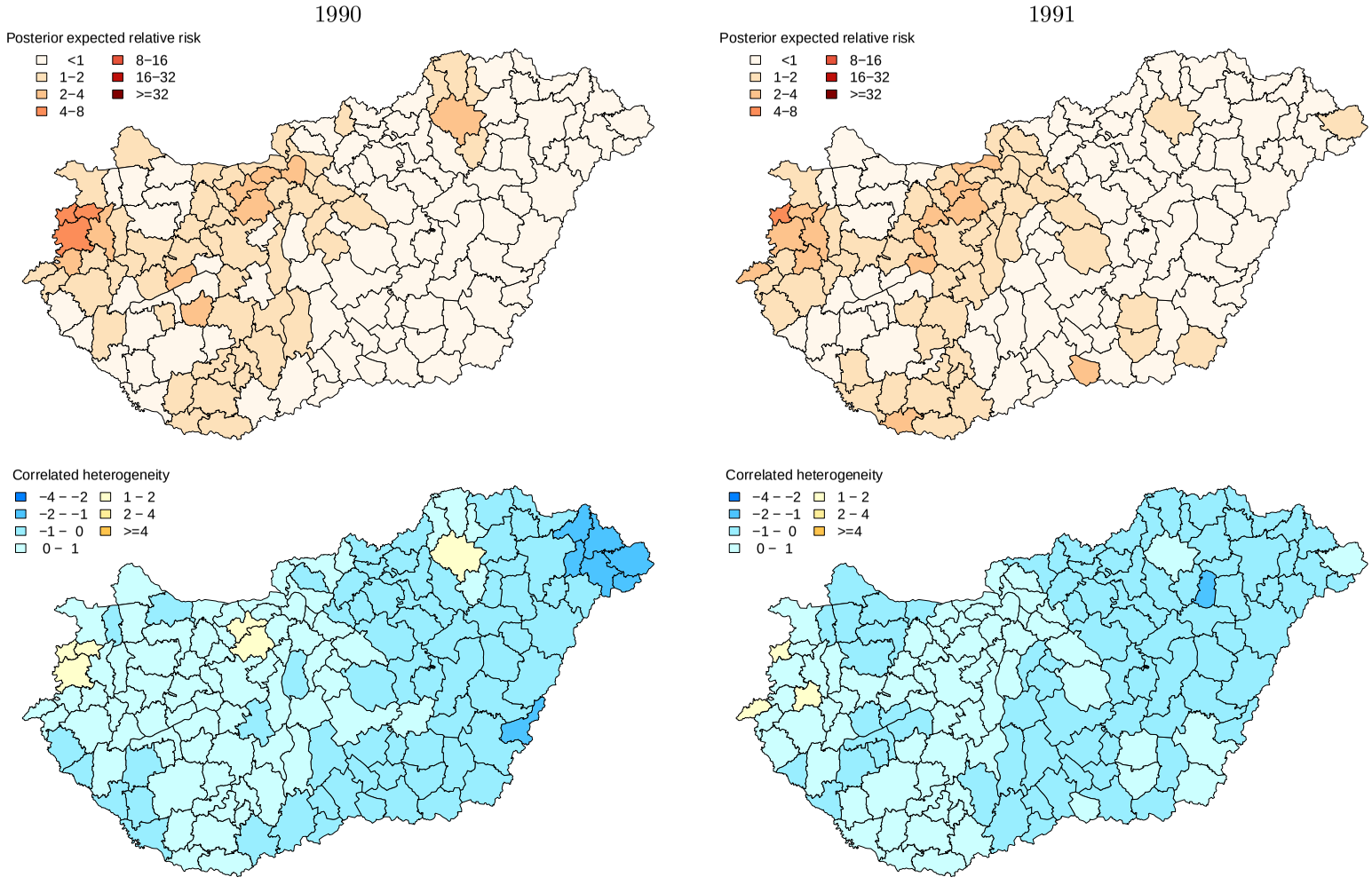


Figure B.1: Posterior relative risk and correlated heterogeneity (BYM model) for red fox rabies, Hungary (1990, 1991).

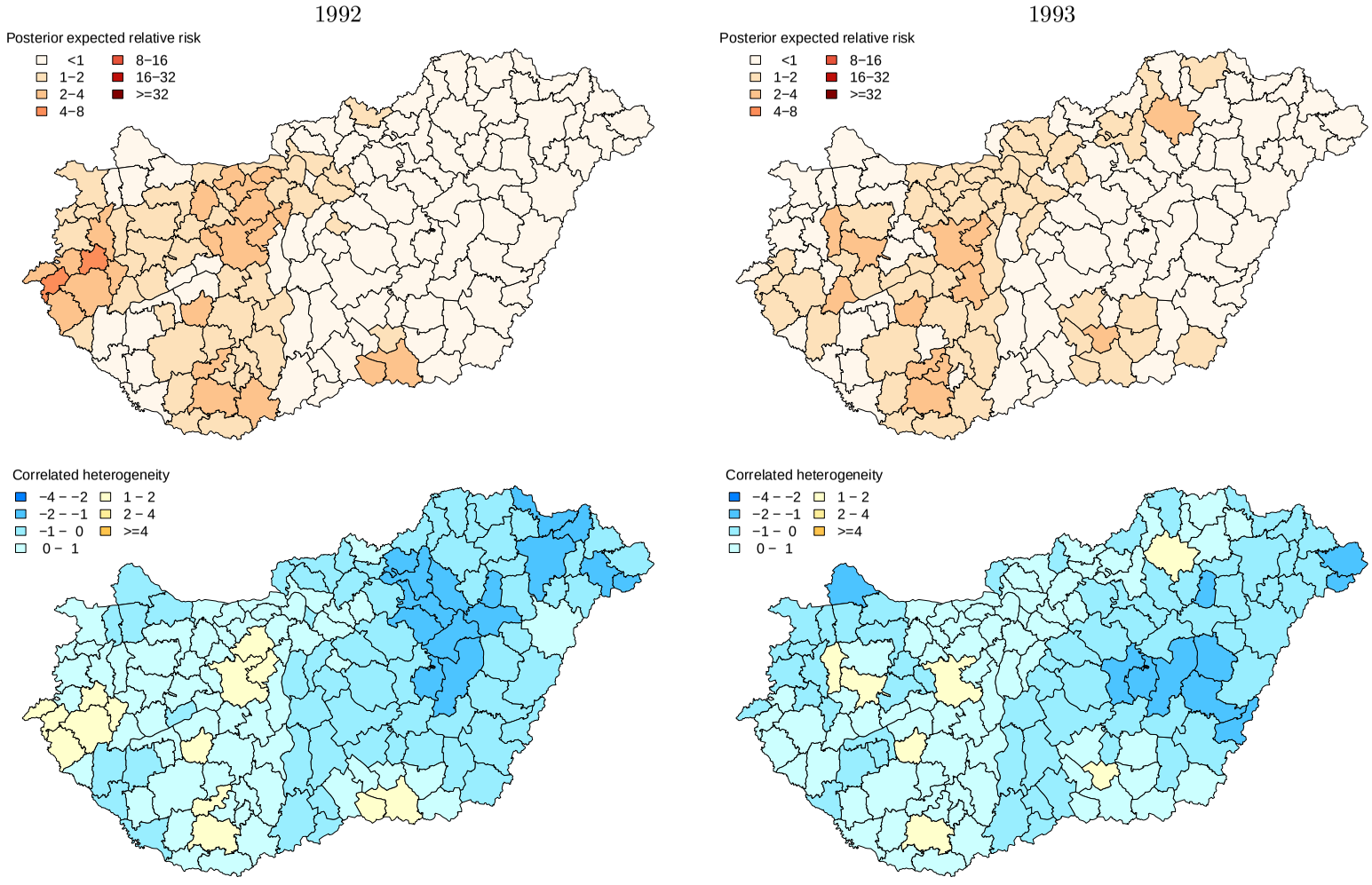


Figure B.2: Posterior relative risk and correlated heterogeneity (BYM model) for red fox rabies, Hungary (1992, 1993).

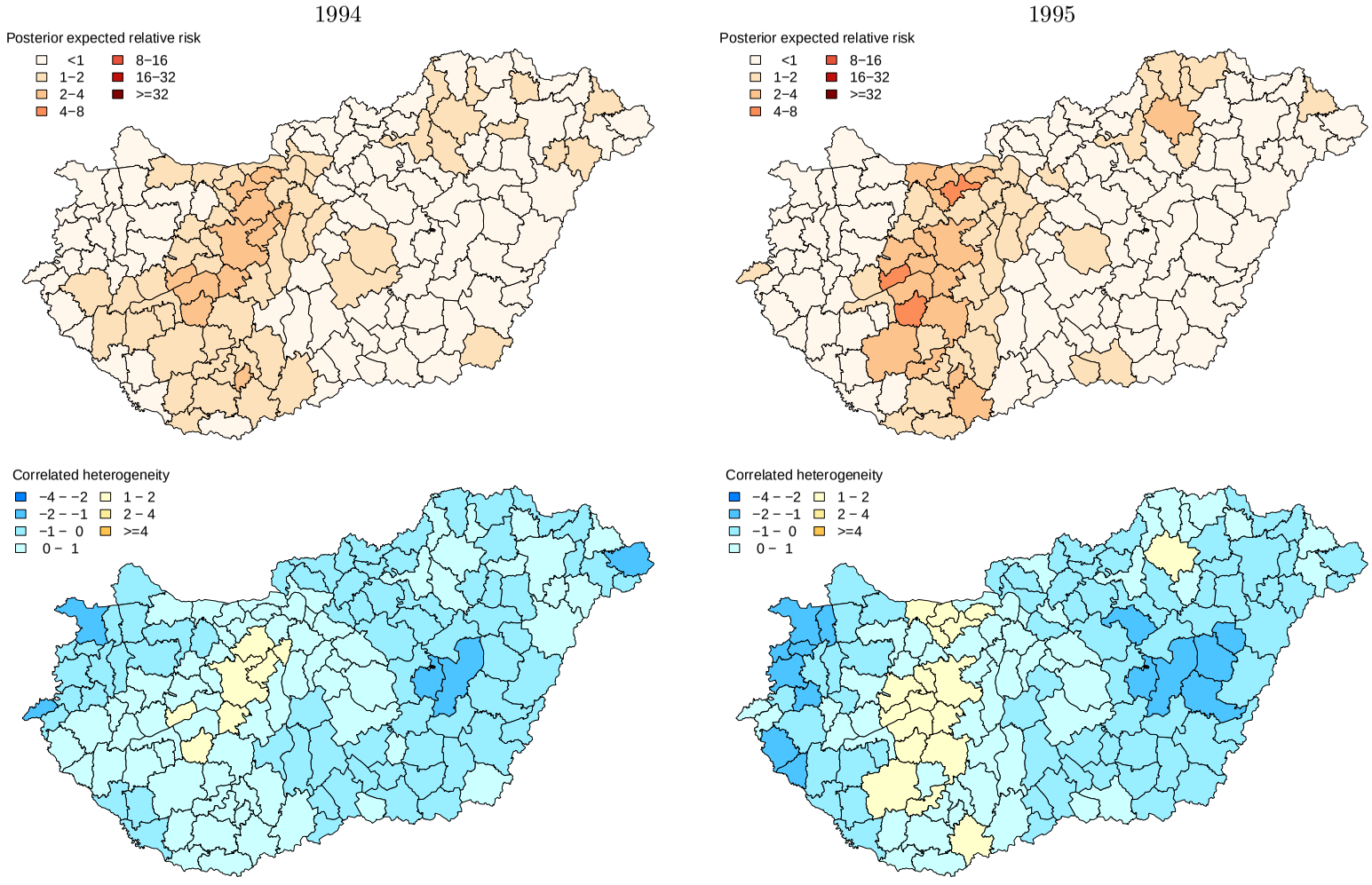


Figure B.3: Posterior relative risk and correlated heterogeneity (BYM model) for red fox rabies, Hungary (1994, 1995).

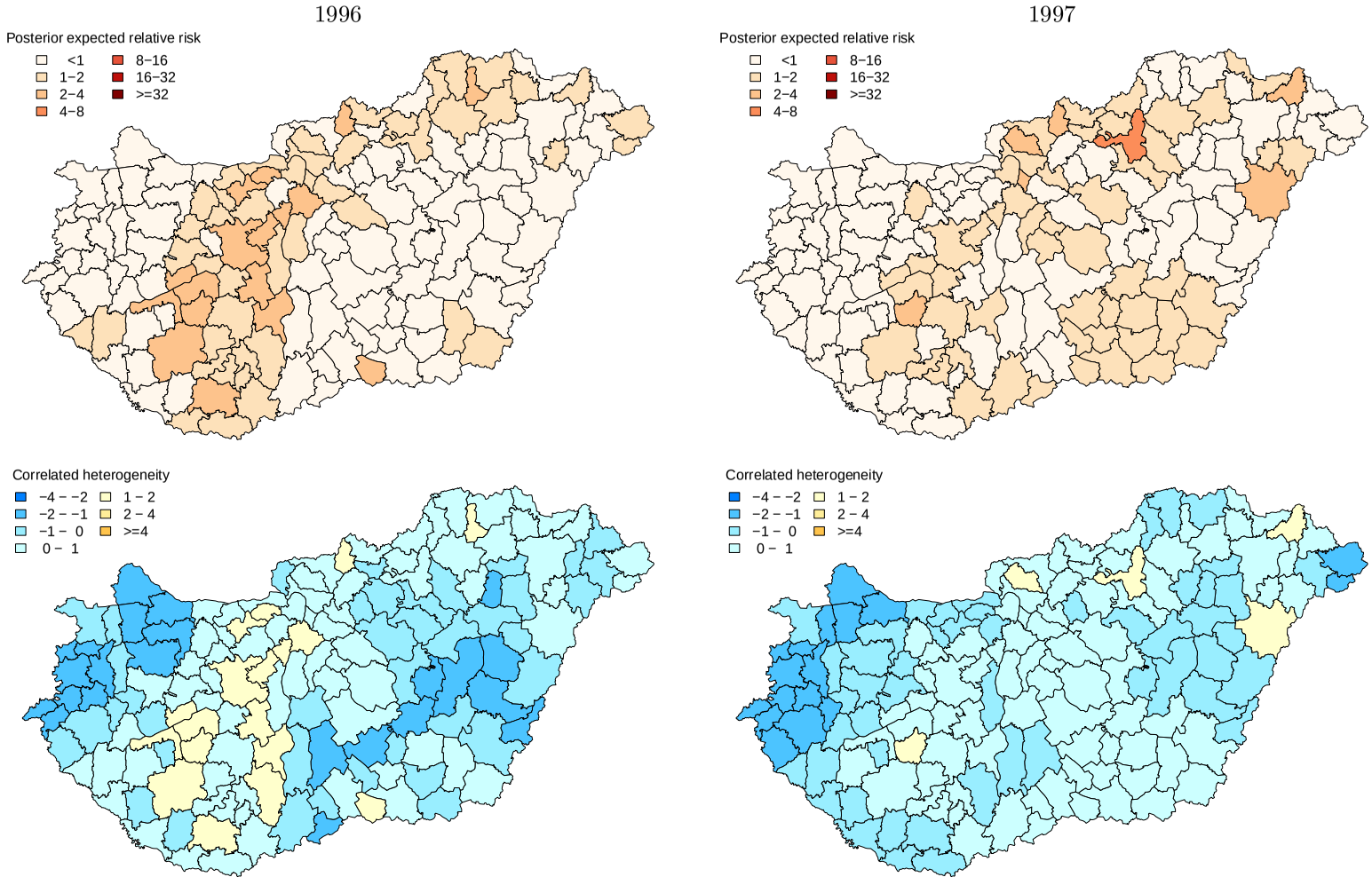


Figure B.4: Posterior relative risk and correlated heterogeneity (BYM model) for red fox rabies, Hungary (1996, 1997).

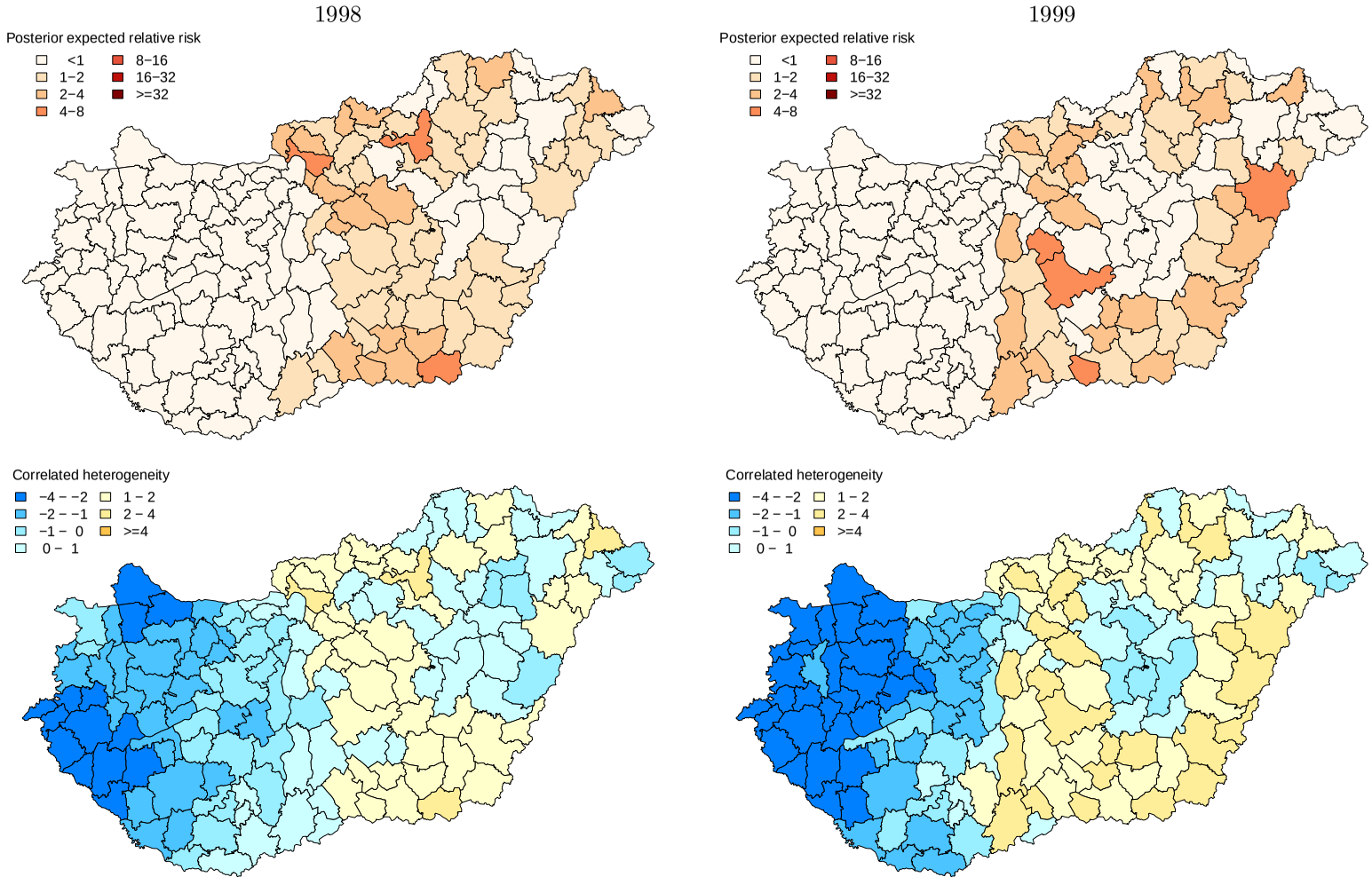


Figure B.5: Posterior relative risk and correlated heterogeneity (BYM model) for red fox rabies, Hungary (1998, 1999).

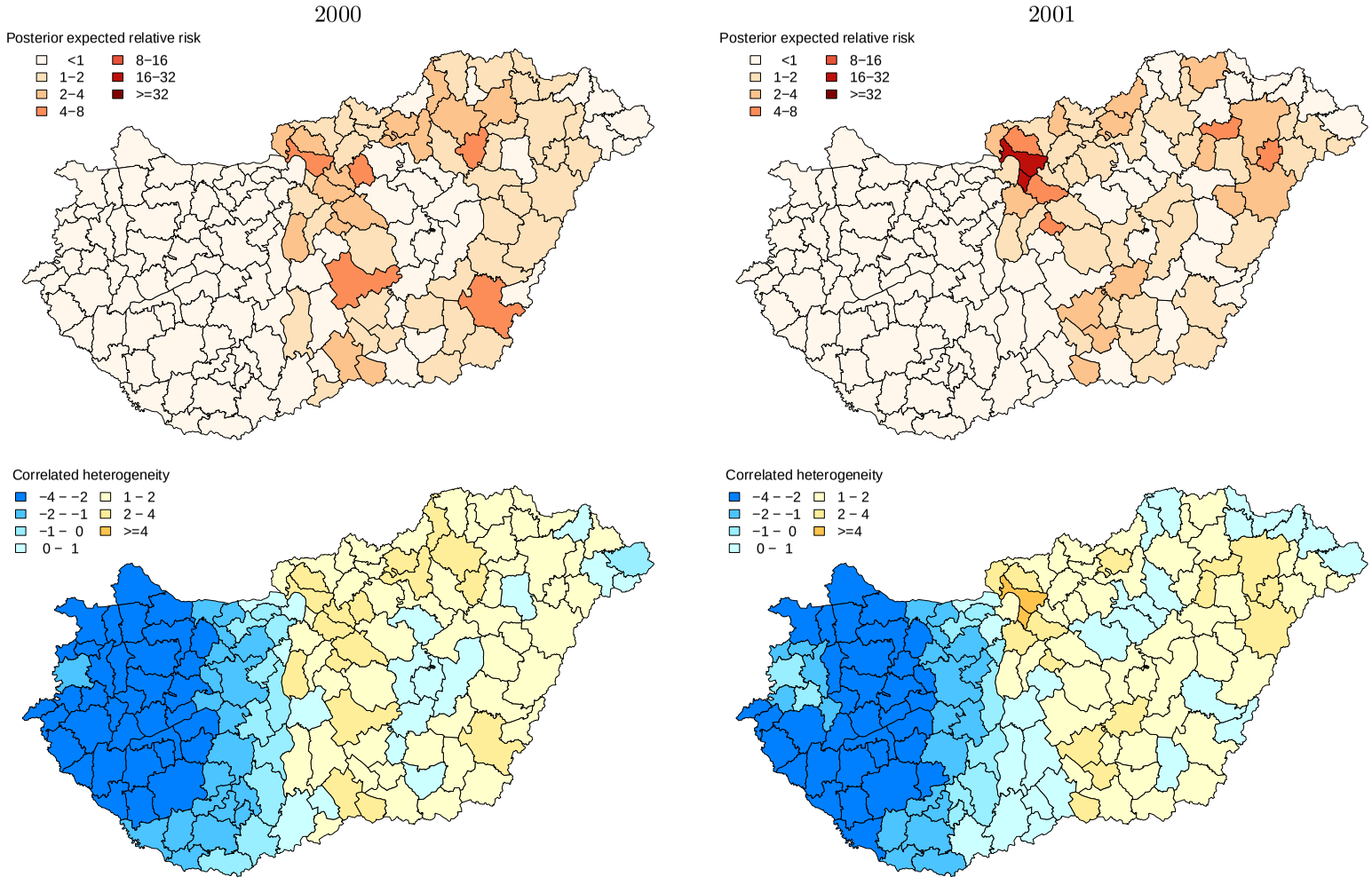


Figure B.6: Posterior relative risk and correlated heterogeneity (BYM model) for red fox rabies, Hungary (2000, 2001).

Table B.1: District level posterior expected relative risk for red fox rabies, Hungary, 1990-2001.

District	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	$RR > 1$
Ajkai	1.22	1.07	1.84	2.61	0.99	0.83	0.82	0.37	0.09	0.04	0.03	0.05	4
Aszódi	0.76	1.44	1.28	0.91	0.72	1.03	0.87	1.15	1.82	1.09	2.81	1.16	8
Bácsalmási	0.40	0.48	0.74	0.79	0.42	0.80	0.31	1.74	0.94	0.35	1.18	0.44	2
Bajai	0.40	0.46	0.67	0.47	1.14	0.84	0.64	1.30	1.02	2.21	0.67	0.57	4
Baktalórántházai	0.26	0.81	0.34	0.50	0.70	0.57	0.43	1.00	1.49	0.32	1.29	1.37	3
Balassagyarmati	0.90	0.89	0.58	1.03	0.44	0.92	0.90	1.31	1.88	1.16	1.81	1.04	6
Balatonalmádi	1.00	2.20	0.91	1.34	1.50	3.45	1.23	1.26	0.28	0.07	0.05	0.06	6
Balatonfüredi	2.22	0.93	0.82	1.07	2.59	4.61	3.72	1.35	0.15	0.08	0.04	0.05	6
Balmazújvárosi	0.58	0.44	0.31	0.37	0.63	0.35	0.30	0.51	0.84	0.88	1.33	0.92	1
Barcsi	0.51	1.03	0.60	0.89	0.58	0.39	0.82	0.52	0.11	0.05	0.09	0.05	1
Bátonyterenyei	0.69	0.67	0.39	0.97	0.54	0.58	0.74	1.09	1.47	2.00	1.83	1.71	5
Berettyóújfalui	0.37	0.69	0.38	0.34	0.67	0.50	0.71	0.74	0.45	2.02	1.83	1.19	3
Bicskei	3.19	2.23	3.49	1.56	2.40	1.99	1.70	0.85	0.22	0.08	0.12	0.14	7
Bonyhádi	1.66	1.19	1.23	1.35	1.43	1.46	1.63	1.01	0.64	0.54	0.10	0.16	8
Budaörsi	1.37	1.28	2.22	1.80	3.40	1.31	3.19	1.25	0.38	0.74	0.36	0.40	8
Budapest	1.26	1.55	1.15	1.67	1.21	1.42	2.95	1.30	0.88	0.26	1.96	3.06	10
Ceglédi	0.82	1.25	0.38	0.45	1.78	1.40	0.94	1.79	1.32	1.00	1.88	1.84	7
Celldömölki	1.25	1.36	1.46	2.52	0.53	0.31	0.34	0.31	0.09	0.04	0.03	0.05	4
Csengeri	0.20	0.84	0.26	0.26	0.52	0.40	1.36	0.36	0.50	0.23	0.82	1.48	2
Csepregi	5.45	2.35	1.49	0.84	0.63	0.20	0.18	0.28	0.30	0.06	0.05	0.12	3
Csongrádi	0.74	0.80	0.82	0.69	0.94	0.44	0.49	1.36	1.01	2.57	0.82	1.08	4
Csornai	0.98	0.65	0.71	0.41	0.47	0.49	0.22	0.21	0.08	0.03	0.03	0.05	0
Csurgói	0.56	0.71	0.51	0.52	0.83	0.47	0.99	0.42	0.08	0.04	0.05	0.05	0
Dabasi	1.64	0.90	0.63	0.51	0.96	0.54	0.86	1.69	0.85	5.08	0.99	0.69	3
Debreceni	0.40	0.80	0.91	0.78	0.91	0.37	0.91	2.27	1.90	4.74	1.79	2.77	5
Dombóvári	0.96	1.13	1.37	0.97	1.13	1.65	1.19	0.97	0.17	0.36	0.06	0.07	5
Dorogi	2.09	1.83	2.27	1.63	2.21	2.13	3.17	0.58	0.68	0.14	0.28	0.44	7
Dunakeszi	1.13	1.31	0.89	1.77	1.77	1.04	1.69	2.02	2.02	0.47	1.35	17.58	10
Dunaújvárosi	1.18	1.44	0.88	1.92	1.20	1.98	1.66	0.72	0.25	0.17	0.21	0.19	6
Edelényi	1.02	0.48	0.39	0.61	0.81	1.38	1.08	0.77	1.07	0.28	1.22	0.50	5
Egri	0.78	0.51	0.30	1.18	1.34	1.58	1.16	4.10	4.08	0.86	3.12	1.42	8
Encsi	0.58	0.55	0.61	1.03	0.78	1.33	1.16	0.76	2.39	1.87	1.94	2.02	7
Enyingi	1.83	1.87	2.00	1.66	2.88	2.50	1.79	1.86	0.19	0.09	0.07	0.08	8
Ercsi	0.98	1.23	2.57	2.15	2.56	1.16	2.84	1.64	0.37	0.12	0.15	0.17	7
Esztergomi	1.87	2.55	1.35	1.76	1.63	1.90	1.29	0.63	0.89	0.18	0.40	0.67	7
Fehérgyarmati	0.23	1.18	0.43	0.21	0.34	0.64	1.25	0.31	0.42	0.75	0.32	0.72	2
Fonyódi	0.91	0.78	0.94	0.77	1.82	1.13	3.37	0.83	0.11	0.23	0.04	0.04	3
Füzesabonyi	0.68	0.39	0.21	0.49	0.70	0.51	0.32	1.46	1.57	0.50	0.90	0.46	2
Gödöllői	1.93	1.60	1.11	1.55	0.96	0.51	1.04	0.99	3.30	2.88	3.11	4.71	9
Gyáli	1.46	0.87	0.58	1.10	1.04	1.13	0.90	1.21	1.37	0.86	1.39	1.73	8
Gyöngyösi	0.55	0.70	0.33	0.61	0.53	0.36	0.29	0.78	0.84	0.26	0.95	1.01	1
Győri	0.75	0.53	0.45	0.43	1.02	0.41	0.20	0.20	0.08	0.05	0.04	0.05	1
Gyulai	0.39	0.98	0.36	0.77	0.54	0.37	0.49	1.41	1.43	2.09	4.49	1.73	5
Gyulai	0.27	0.77	0.44	0.25	0.92	0.73	0.27	0.59	1.76	2.73	0.89	0.52	2
Hajdúböszörményi	0.46	0.40	0.46	0.45	0.51	0.36	0.71	0.66	0.39	1.47	0.72	1.08	2
Hajdúszoboszlói	0.55	0.70	0.76	0.46	0.57	0.23	0.33	0.66	0.90	3.12	1.24	2.20	3
Hatvani	0.77	0.74	0.97	1.37	0.69	0.55	0.53	0.66	1.32	2.17	5.80	1.07	5
Hevesi	0.76	0.74	0.24	0.33	0.75	0.28	0.50	0.64	0.69	0.17	0.41	0.44	0
Hódmezővásárhelyi	0.78	1.19	1.00	1.72	0.57	0.43	0.45	1.49	3.33	1.41	0.50	1.62	6
Jánoshalmai	0.54	0.76	0.66	0.74	0.46	0.53	0.69	0.83	1.14	2.14	0.60	0.76	2

District	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	$RR > 1$
Jászberényi	0.66	0.76	0.54	0.51	0.69	0.47	0.35	1.56	2.45	0.36	0.76	0.50	2
Kalocsai	1.06	0.77	0.48	0.56	0.64	0.81	0.53	0.73	0.87	3.68	1.02	0.39	3
Kaposvári	0.94	0.97	1.77	1.25	1.83	2.49	2.45	1.51	0.10	0.06	0.04	0.04	6
Kapuvári	0.71	0.75	0.71	0.42	0.39	0.27	0.42	0.24	0.11	0.04	0.03	0.06	0
Karcagi	0.40	0.39	0.26	0.30	0.30	0.21	0.17	0.48	0.97	0.14	0.62	1.35	1
Kazincbarcikai	1.33	0.50	0.38	1.08	1.33	1.18	1.10	0.78	0.66	2.82	2.34	0.61	7
Kecskeméti	0.97	0.69	0.34	0.44	1.15	0.83	0.84	0.88	1.79	4.96	4.28	0.93	4
Keszhely-Hévízi	0.90	0.97	1.14	2.11	1.65	0.89	0.80	0.35	0.07	0.04	0.03	0.05	3
Kisbéri	1.40	1.14	2.00	1.66	0.88	1.19	1.42	1.00	0.13	0.06	0.06	0.07	6
Kiskőrösi	0.65	0.51	0.43	0.70	0.63	0.54	0.27	0.71	0.81	1.41	0.84	0.78	1
Kiskunfélegyházi	0.62	0.67	0.75	1.45	0.68	0.47	0.23	1.08	1.05	0.86	1.16	2.60	5
Kiskunhalasi	0.39	0.93	0.83	0.73	0.47	0.76	0.79	0.77	2.38	1.13	2.07	0.71	3
Kiskunmajsai	0.41	0.93	0.96	1.00	0.67	0.68	0.56	1.07	1.46	0.94	1.28	1.81	4
Kisteleki	0.85	0.93	1.84	2.50	0.60	0.67	0.82	1.40	3.07	2.12	1.05	3.04	7
Kisvárdai	0.24	0.95	0.31	0.63	0.59	0.66	1.45	3.44	1.62	2.13	0.78	0.58	4
Komáromi	1.37	0.75	1.37	1.57	1.50	2.65	0.81	0.40	0.13	0.30	0.06	0.08	5
Komlói	1.58	1.06	1.58	2.25	1.06	1.74	1.29	1.52	0.36	0.18	0.09	0.14	8
Körmendi	2.03	1.67	2.17	1.01	0.47	0.33	0.15	0.17	0.07	0.03	0.04	0.10	4
Kőszegi	4.48	4.80	1.46	0.92	0.39	0.26	0.19	0.29	0.21	0.06	0.06	0.14	3
Kunszentmártoni	0.95	0.80	0.41	0.40	0.49	0.45	0.21	1.58	1.60	0.63	0.50	2.52	3
Kunszentmiklósi	0.95	0.88	0.44	0.44	0.78	0.63	0.59	0.75	0.42	1.77	0.72	0.37	1
Lengyeltóti	1.18	0.79	1.00	0.63	1.74	0.71	0.74	0.66	0.11	0.09	0.05	0.06	2
Lenti	0.55	1.65	2.38	1.49	0.93	0.21	0.31	0.18	0.06	0.03	0.03	0.07	3
Letenyei	0.71	0.85	1.22	0.98	0.92	0.26	1.10	0.36	0.06	0.03	0.04	0.07	2
Makói	0.53	0.50	0.85	0.93	0.58	0.33	0.70	1.44	5.36	2.98	1.02	1.58	5
Marcali	0.61	0.90	0.55	0.50	1.02	0.50	0.79	0.79	0.07	0.04	0.04	0.04	1
Mátészalkai	0.21	0.72	0.24	0.52	0.59	0.94	0.68	0.53	1.29	0.21	0.79	1.18	2
Mezőkovácsházi	0.55	1.07	0.67	1.39	1.46	0.35	1.40	1.04	1.57	1.93	1.34	1.28	9
Mezőkövesdi	0.80	0.42	0.26	0.80	1.00	0.73	0.74	1.95	1.23	1.10	1.63	0.51	4
Miskolci	2.97	1.27	0.79	2.94	1.41	3.15	1.75	1.88	1.78	1.19	3.62	1.56	11
Mohácsi	0.78	1.55	2.18	1.52	1.76	3.02	1.77	1.83	0.95	0.15	0.19	0.86	7
Monori	0.94	0.96	1.03	0.49	0.99	1.00	0.92	1.67	3.02	0.58	3.54	4.09	5
Mórahalomi	0.57	2.13	2.50	1.06	0.95	1.72	2.59	1.64	3.00	4.91	3.02	3.23	10
Móri	1.87	2.21	1.87	1.10	1.64	1.14	1.60	0.87	0.18	0.06	0.07	0.08	7
Mosonmagyaróvári	1.17	0.48	0.33	0.26	0.50	0.78	0.17	0.16	0.08	0.03	0.03	0.05	1
Nagyatádi	0.49	1.16	0.94	0.95	0.59	0.48	0.59	0.70	0.08	0.04	0.05	0.05	1
Nagykállói	0.45	0.79	0.42	0.51	1.82	0.67	1.19	1.26	1.69	0.32	1.08	4.70	6
Nagykanizsai	1.18	0.95	0.75	1.23	1.45	0.69	1.35	0.52	0.06	0.03	0.03	0.04	4
Nagykátai	1.02	1.05	0.70	0.75	0.95	0.95	1.27	0.88	2.44	2.15	3.00	1.97	7
Nyírbátori	0.37	0.81	0.33	0.61	1.42	0.35	0.49	1.38	1.62	1.63	1.86	1.43	6
Nyíregyházi	0.48	0.60	0.27	0.60	0.96	0.60	0.74	0.80	0.67	0.20	1.40	2.22	2
Orosházai	0.81	0.86	0.61	0.81	0.57	0.43	1.17	1.58	1.75	1.14	1.36	0.83	5
Oroszlányi	1.52	1.17	1.84	1.12	1.17	2.19	1.46	0.79	0.20	0.09	0.09	0.12	7
Ózdi	0.53	0.40	0.37	0.47	0.57	0.70	0.78	1.37	0.70	0.29	0.95	2.02	2
Őriszentpéteri	1.25	1.22	5.82	1.04	0.60	0.85	0.23	0.17	0.06	0.03	0.04	0.08	4
Paksi	1.40	0.95	1.59	1.68	0.80	1.49	2.81	1.01	0.24	0.42	0.19	0.15	6
Pápai	0.99	0.86	1.13	1.98	0.58	0.62	0.25	0.32	0.09	0.03	0.03	0.04	2
Pásztói	0.88	0.80	0.45	1.23	0.72	0.39	1.15	0.84	1.03	2.06	1.12	1.22	6
Pécsi	1.32	1.98	2.71	2.43	1.59	1.41	2.26	0.69	0.33	0.09	0.08	0.11	7
Pécsváradi	1.02	1.70	1.35	0.98	2.32	2.03	1.49	0.88	0.71	0.83	0.12	0.23	6
Pétervásárai	0.61	0.52	0.32	1.56	0.64	0.61	1.58	1.58	1.66	1.84	2.44	3.17	7
Pilisvörösvári	1.34	1.20	1.81	1.02	1.65	1.25	0.89	0.63	0.37	0.13	0.27	0.40	6

District	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	$RR > 1$
Polgári	0.45	0.34	0.33	0.32	0.61	0.34	0.28	0.45	0.44	0.30	1.40	3.44	2
Püspökladányi	0.54	0.41	0.33	0.32	0.73	0.22	0.21	0.43	0.52	1.56	1.64	0.99	2
Ráckevei	0.61	1.15	0.46	0.97	1.03	1.38	1.28	0.84	0.79	3.42	2.25	1.38	7
Rétsági	0.78	1.18	0.54	1.04	0.53	0.48	1.00	3.57	3.14	2.69	2.29	4.55	7
Salgótarjáni	0.91	0.82	1.04	0.92	0.47	0.66	1.15	1.96	3.52	0.72	1.92	1.86	6
Sárbogárdi	0.84	1.89	1.49	2.06	1.84	1.93	2.47	0.92	0.18	0.09	0.16	0.10	6
Sárospataki	0.45	0.51	0.45	0.90	1.11	0.85	1.12	0.98	0.80	1.26	1.58	0.68	4
Sárvári	2.22	2.54	2.19	0.55	0.67	0.36	0.18	0.27	0.14	0.13	0.03	0.07	3
Sásdi	1.79	0.98	3.16	2.31	1.25	2.34	1.11	1.73	0.29	0.13	0.06	0.08	7
Sátoraljaújhelyi	0.51	0.83	0.25	0.79	0.70	0.48	0.94	1.44	0.65	0.54	1.12	0.80	2
Sellyei	1.39	2.12	1.81	1.92	1.77	1.73	1.29	0.46	0.34	0.08	0.10	0.09	7
Siklósi	1.36	1.63	1.55	1.03	0.92	1.22	1.50	0.73	0.59	0.08	0.20	0.16	6
Siófoki	0.92	0.83	0.94	1.28	2.21	2.29	2.41	1.58	0.23	0.26	0.05	0.06	5
Sopronfertődi	1.23	1.42	1.02	0.71	0.34	0.15	0.39	0.55	0.28	0.04	0.04	0.07	3
Sümegi	1.61	1.26	1.44	1.52	0.91	0.54	0.31	0.31	0.09	0.05	0.03	0.05	4
Szarvasi	0.73	0.62	0.33	0.75	0.40	0.32	0.31	1.08	1.49	0.27	1.05	0.88	3
Szécsényi	1.08	0.85	0.70	0.88	0.51	0.82	2.63	2.04	2.07	1.07	1.61	2.19	7
Szegedi	0.41	0.60	2.45	1.57	0.79	1.39	0.89	1.10	2.21	1.46	0.90	0.90	6
Szeghalomi	0.64	0.57	0.35	0.31	0.46	0.23	0.19	0.39	1.13	1.32	1.45	0.83	3
Székesfehérvári	1.34	1.94	2.92	2.57	2.99	3.68	3.78	1.51	0.27	0.05	0.08	0.08	8
Szekszárdi	1.36	0.82	1.25	1.37	1.20	1.32	1.99	0.72	0.38	0.83	0.16	0.19	6
Szentendrei	2.15	1.67	1.10	1.40	1.07	1.84	1.46	1.07	0.71	0.19	0.54	1.08	9
Szentesi	0.76	1.19	0.59	1.68	0.74	0.47	0.87	1.47	1.87	3.80	1.21	1.44	7
Szentgothárdi	1.37	2.75	2.85	0.70	0.39	1.27	0.15	0.18	0.08	0.04	0.05	0.12	4
Szerencsi	0.84	0.40	0.38	0.91	0.70	0.73	1.31	1.08	1.79	2.78	2.01	0.92	5
Szigetvári	1.75	1.59	1.39	1.18	1.62	0.96	0.56	0.61	0.17	0.25	0.07	0.07	5
Szikszói	1.58	0.59	0.65	0.92	1.44	1.60	2.24	0.92	1.18	3.54	0.88	1.26	7
Szobi	1.99	1.65	0.67	0.81	0.67	0.76	0.48	0.86	3.18	0.96	2.06	2.68	5
Szolnoki	0.75	0.54	0.57	0.28	0.51	0.33	0.28	0.76	1.32	0.22	0.62	1.25	2
Szombathelyi	6.11	2.33	1.92	0.60	0.43	0.18	0.15	0.19	0.11	0.04	0.07	0.18	3
Tabi	2.15	1.34	3.09	2.54	2.82	5.34	3.86	2.56	0.31	0.11	0.05	0.06	8
Tamási	1.81	1.14	1.50	1.29	1.17	2.12	1.61	1.47	0.19	0.12	0.07	0.07	8
Tapolcai	1.58	0.78	1.10	1.13	1.31	1.20	0.71	0.62	0.10	0.05	0.03	0.05	5
Tatabányai	2.84	2.18	2.20	1.32	2.38	4.33	3.25	0.66	0.32	0.10	0.16	0.23	7
Tatai	1.01	1.27	2.27	1.42	1.44	3.18	1.24	0.65	0.41	0.12	0.13	0.19	7
Téti	0.93	0.66	1.09	0.73	0.71	0.78	0.24	0.35	0.09	0.04	0.04	0.05	1
Tiszafüredi	0.56	0.43	0.25	0.47	0.53	0.30	0.40	0.56	0.74	0.62	0.97	1.12	1
Tiszaújvárosi	1.96	0.39	0.71	0.77	0.84	1.53	0.88	0.71	1.14	1.11	4.40	1.13	6
Tiszavasvári	0.55	0.42	0.40	0.38	1.11	0.58	0.36	0.53	0.51	0.32	1.81	4.97	3
Törökszentmiklósi	0.68	0.44	0.25	0.24	0.32	0.27	0.16	0.76	1.14	0.17	0.36	0.98	1
Váci	1.09	1.65	1.03	1.56	1.01	0.67	1.70	1.61	5.33	1.15	4.08	18.21	11
Várpalotai	1.29	2.46	2.17	0.88	1.17	2.32	0.84	0.89	0.29	0.06	0.06	0.07	5
Vásárosnaményi	0.17	0.92	0.46	0.57	1.16	1.56	0.37	0.95	3.65	0.91	1.00	0.76	3
Vasvári	1.49	2.90	4.86	0.49	0.82	0.28	0.16	0.20	0.08	0.04	0.04	0.20	3
Veszprémi	1.67	1.49	1.28	0.74	1.22	2.09	1.14	0.55	0.16	0.04	0.04	0.05	6
Zalaegerszegi	1.23	1.29	2.30	1.13	1.25	0.47	0.31	0.23	0.06	0.02	0.03	0.05	5
Zalaszentgróti	1.10	1.21	2.49	0.74	0.96	0.84	0.31	0.24	0.08	0.05	0.03	0.07	3
Zirci	1.01	1.30	1.95	0.77	1.19	1.56	1.18	0.61	0.13	0.05	0.05	0.06	6