TICK-BORN ENCEPHALITIS RISK ASSESSMENT BASED ON SATELLITE DATA

JAN KOLÁŘ, MARKÉTA POTŮČKOVÁ, EVA ŠTEFANOVÁ

Charles University, Faculty of Science, Department of Applied Geoinformatics and Cartography, Czech Republic

ABSTRACT

Tick-borne encephalitis (TBE) belongs among the dangerous vector-borne diseases. The number of TBE incidences has been permanently increasing in various geographical regions, including the Czech Republic. The presence of ticks and related diseases is driven by host-pathogen systems. The systems are rather complex and susceptible to environmental conditions represented in the first place by land cover/land use categories. The presented study looks for a possible relation between the types of forest vegetation specified in the Landsat 5 satellite data and relative TBE morbidity. First, supervised classification of forest areas into five vegetation classes predefined by a botanist was tested. Due to the spectral similarity of the classes, the resulting classification accuracy of Landsat scenes covering the entire area of the Czech Republic was quite low. Thus, an unsupervised approach was applied using nine spectral classes. Relative TBE morbidity data collected over 10 years for 206 administrative units covering the entire country presented field data that were correlated with the spectral classes. The TBE risk index (IRE) of a given spectral class was introduced at each satellite scene. To create a map of the TBE risk for the entire country, all IRE values were accumulated and divided into six risk categories. The disadvantages of the proposed method, especially regarding the accuracy of the final product with a nationwide cover age, are discussed. In addition, the correlation between the relative TBE morbidity and other environmental parameters, such as annual precipitation, average temperature, and number of hunted game were calculated, but they did not reveal any significant relationship.

Keywords: ticks, encephalitis, classification, Landsat

Received 15 December 2015; Accepted 13 June 2016

1. Introduction

1.1 Concepts of using remote sensing for encephalitis risk assessment

The surveillance of vectors and vector-borne diseases is essential for their control. Vector surveillance can be defined as the monitoring of arthropod populations responsible for the transmission of pathogens. In addition to other things, vector surveillance can be used to detect the presence/absence of a vector population and, subsequently, be used for disease risk assessment.

The spread of many vectors and related diseases is driven by host-pathogen systems. The distribution of their locations is susceptible to environmental conditions, represented in the first place by land cover/land use categories (e.g. Cortinas et al. 2002; Eisen et al. 2006; Estrada-Peña 2001). Determination and mapping of the main land cover categories like vegetation, water, or urban areas can be done effectively by means of remote sensing (RS) from satellites or airplanes.

RS and geographic information systems (GIS) have been widely used in health applications for several decades. A substantial proportion of the research papers published in this field deals with application to spatially delineate vector habitats and disease patterns (e.g. Cortinas et al. 2002; Rogers and Randolph 2003; Tatem et al. 2004). The incidence of tick-borne encephalitis (TBE) and Lyme borreliosis currently continues to increase in many European countries while the reasons for this are not yet fully understood.

It has been recognized that efficient extracting the information from RS images and applying it in studies of disease control requires the inclusion of several professional fields such as geography, RS, biology, ecology, computer science, etc., which in turn demands interdisciplinary cooperation and research teams.

RS adds qualified information for the identification of vulnerable ecosystems at a relatively low cost, thus providing an important ancillary tool for studying certain endemics and supporting surveillance and control activities. In contrast to point feature of field observations, satellite data continuously cover the entire area of interest, allowing a more complete picture of the environmental conditions. Moreover, RS data can be processed in time series to gather information about changes and trends in environmental conditions.

Among the various sensors used, there was a predominance of AVHRR (NOAA satellite) and TM (Landsat satellite), possibly because they have long historical series and they are easily accessible. By contrast, there is only a negligible number of case studies using very high-resolution images. Data from Landsat and NOAA satellites were used in vector habitats studies mostly to classify land

https://doi.org/10.14712/23361980.2016.13
Kolář, J. – Potůčková, M. – Štefanová, E. (2016):
Tick-born encephalitis risk assessment based on satellite data
AUC Geographica, 51, No. 2, pp. 155–167

cover categories, namely vegetation categories using the Normalized Difference Vegetation Index (NDVI) (Herbreteau et al. 2007).

The NDVI calculated from NOAA-AVHRR data was used to monitor the seasonal activity of larvae and nymphs of *Ixodes ricinus* together with the TBE virus in seven European countries (Randolph et al. 2000). Several studies (Brown et al. 2008; Lourenço et al. 2011) demonstrated a relation between the spatio-temporal variability of NDVI and the occurrence of vector-borne diseases. Other RS parameters have been applied as well. Altobelli et al. (2008) developed a prediction model for the abundance of infected ticks (Ixodes ricinus) in north-east Italy using the values of Land Surface Temperature (LST), Land Surface Water Index (LSWI), Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI). A model for the distribution of another tick, *Hyalomma marginatum*, over some parts of Europe was prepared by Estrada-Peña et al. (2014).

The number of projects has been focused on the development of reliable methods for creating prediction maps of tick occurrence and related disease risk distribution using RS. An overview of the most important papers is given in several articles; for example, Estrada-Peña (2001). A review of a study on the use of GIS and RS for the *Ixodes scapularis* and the spread of Lyme disease in the central part of the United States has been published by Cortinas et al. (2002), while Cromley (2003) mentions another application of satellite imagery for monitoring tick-borne and other diseases in the United States.

The relationship between tick occurrence and vegetation types was confirmed during an experiment in the Siebengebirge nature reservation in western Germany, as described by Schwarz et al. (2009). For seven months, ticks were regularly collected in 5 areas together with measurements of air temperature and humidity 5 cm above the ground, the water content in the soil, climatological data, soil types, and namely, vegetation types. The highest average number of ticks was found in vegetation with humid dense herbaceous and shrub undergrowth and a layer of leaves. Tick occurrence decreased where the undergrowth was less developed and was the lowest in xerophilous vegetation and underdeveloped herbaceous and shrub floors. Furthermore, the number of ticks increased with a growing temperature (up to 24 °C) and decreased when the air humidity increased. A very strong positive correlation was then found between the number of ticks and the water content in the soil.

An example of the application of satellite data for the creation of a prediction map is shown by a study in the Mendocino district of California, where the density of nymphs of *Ixodes Pacificus* was studied (Eisen et al. 2006). Images from Landsat 5 TM acquired in different periods of the year (May, July, November, and February) were processed to get the NDVI and spectral features derived from the Tasseled Cap transformation. Three vegetation classes were selected and defined in accordance with the level of tick appearance: dense forest with a high incidence of nymphs, permanent grassland and woodlands with a grassy understory with adult ticks, and agricultural land and water areas without the occurrence of *Ixodes Pacificus*. Supervised classification of Landsat data gave the best results (82.64% overall accuracy of three risk categories) when carried out using NDVI and one spectral feature from February and July combined in one dataset. The classification results of seven forest types were assessed as insufficient. Processed satellite data together with additional climate and topographic parameters were compared with the density of *Ixodes Pacificus* nymphs collected during a field survey in 62 areas of the Mendocino district. The resulting model gives the nymph density prediction with 72% accuracy.

RS measurements and processing methods cannot identify the vectors themselves, but they can identify and characterize suitable vector habitats. However, they can be an important input in the development of disease risk maps and for monitoring changes over time. Maps showing seasonal risks of vector-borne diseases will be critical in monitoring the impacts of global climate changes on vectors. Satellites are unique tools for observation of the environmental influence on the spread of vectors and should be a part of any vector surveillance program (Martin et al. 2007).

Recently, environmental variables that are interpolated from meteorological stations or monthly estimates of remotely sensed features are included in correlation modelling. Estrada-Peña et al. (2014) produced a global dataset of variables derived from the monthly series of MODIS satellite data by Fourier transform. The dataset includes variables, such as day and night temperature or vegetation and water availability, which could potentially affect the physiological processes of the vectors.

1.2 Satellite data use for TBE risk determination in the Czech Republic

The most common type of tick in a wide geographical area of Europe, including the Czech Republic, is the *Ixodes Ricinus*. As everywhere, its specific occurrence depends on many different abiotic and biotic factors, such as temperature, humidity, vegetation density and potential host occurrence (Estrada-Peña 2001). The most serious diseases that this species of tick in Europe transfers are Lyme disease and tick-borne encephalitis.

Built on the heritage of traditional forms of tick research, testing of the possible use of satellite data started in the Czech Republic in 1990 with a pilot study demonstrating the use of RS data and methods to map the occurrence of *Ixodes ricinus* in the European context (Daniel, Kolar 1990). Image data from the Landsat 5 satellite of 1,600 km^2 test area was used for the localization of six land cover categories confirming the key influence of forests on tick occurrence.

The following research focused mainly on the selection and appropriate definition of vegetation types under consideration and on methods of correlation between TBE statistical data and classified satellite imagery.

Daniel, Kolar, Zeman (1995) evaluated using satellite data for predicting the risk of TBE morbidity caused by occurrence of *Ixodes Ricinus* on a 75 × 75 km^2 area south of Prague. Landsat 5 satellite scenes from the main part of the vegetation growth period (June–September) were classified under forest mask to get nine different categories of forests. An unsupervised classification method was used to get these nine classes. Botanical descriptions and definitions were given to the classes after a field trip to respective areas recognized in the classified images. Data obtained in the field created a training dataset for supervised classification of the whole area of interest. For the same test area a 0.5 km grid was created based on statistics of TBE morbidity in the region during the last twenty years expressing the expected incidence of the disease in eight levels. The grid was subsequently merged with classified satellite data for correlation analysis relating every forest class to a certain level of risk of the infection. Homogenous coniferous forest has shown the smallest risk, while very heterogeneous young deciduous trees have been indicated as a forest vegetation type with the highest risk to become infected by the TBE virus.

Geographical expansion of the processing procedure over the entire area of Bohemia was the objective of a follow-up project (Daniel, Kolar, Benes 1999). Its results demonstrate the advantage of RS for the prediction of sites with increased risk of tick-borne encephalitis infection over large areas. The study also includes a structural analysis of forest areas in addition to spectral classification in an effort to use the heterogeneity index as another indicator of tick presence in forests.

The GIS tools were used to select suitable sites for the field collection of ticks in southern Bohemia (Švec et al. 2009) followed by development of a disease risk prediction model, based on altitude, vegetation cover, population density, and recreational load. A significant correlation was demonstrated between the natural parameters and density of ticks as well as between the overall risk and the total number of diseases. Mixed and deciduous forests were reported as the places with an increased risk of TBE infection (Honig et al. 2011).

1.3 Concept of the research

The aim of the study described in this paper was to construct a map of a potential risk of TBE disease. In order to perform risk prediction of exposure to vector-borne diseases, both biotic data (e.g., tick and host abundance) and abiotic data (environmental constraints) are commonly employed.

The distribution of encephalitis is always indicated by tick presence. This correlation is rather fundamental, but the direct causal relationship linking habitat conditions to tick distribution or abundance still needs to be established. Additionally, tick presence numbers are general and the differentiation between infected ticks and non-infected ticks requires additional costly laboratory examination.

In our approach to develop a risk model for the geographical area of the entire country, we excluded a tick abundance input. Instead, long-term statistical data on encephalitis collected for smaller administrative units was used as an input about the disease distribution in the country.

The study was based on an assumption that the spatial distribution of vector-borne diseases follows certain habitat compositions providing suitable living conditions for vectors and pathogens. Our research position was that vegetation types in forests reflected inherently complex relations of both local micro-environmental variability and larger-scale climatic parameters.

Therefore, the suitable habitat for tick appearance was represented by forest vegetation categories obtained from processed Landsat satellite data depending on their spatial correlation with the morbidity data. As a novelty, vegetation categories have been defined by their spectral features only, not by botanical expression. On top of this, we included in our study the overlaying distributions of other variables, such as land-cover/land-use data, elevation, air temperature, and the occurrence of selected animals being potential tick hosts.

A suitable tick habitat is a necessary condition but still not a sufficient one to make another case of illness into statistics of TBE morbidity. For this, suitable tick habitats have to overlap with places in which human activities occur. Variables describing people's behavior are not easy to define, and they are less exact.

Our effort aimed at creating a statistical regression model estimating the relationship between one response variable (the TBE morbidity in our application) and a set of descriptive covariates. Rather large standard errors of estimated correlation just underline the complexity of the natural relations and objects under study. It provides uncertainty about the real significance of the examined variables for the risk prediction. In accordance with Dormann et al. (2013), a model may label some variables as not significant, even if they are truly influential for the disease incidence. Knowledge of the tick-habitat interactions needs to be further developed, particularly the scale of action of the environmental factors that are the most influential.

The application of the sample data for a modelling environment in different geographical or environmental landscapes can produce serious errors, because samples are likely to change. This is emphasized when satellite data from various dates and regions are to contribute to one resulting model. Bedia et al. (2013) give an example of statistical prediction models giving the impression of a well-fitted model which gives strongly insufficient

results in new geographical regions or changed climatic conditions.

Research in this project aimed to derive the spatial distribution of disease from satellite imagery and other sets of ground parameters. The final result has been presented in the thematic map of risk levels with regard to TBE over the entire country.

The map creation for a geographically large area in this study required testing of a newly developed process when combining results from several models in different geographical areas into one result for the entire country.

2. Data sets used in the study

2.1 Tick-borne encephalitis morbidity data

Nationwide dense sampling of ticks and laboratory testing of these ticks would be the most reliable but, in practice, barely achievable way to gain data about the true occurrence and geographical distribution of TBE. In respect to this, the only suitable and available source of data about the rate of illness and its spatial distribution is the register of TBE cases administered by the National Institute of Public Health. The provided data were related to the area of the municipalities with extended power (MEP). Thus, 206 MEP covering the entire area of the Czech Republic became the basic area units of the investigation. Specifically, the data represents an annual average of TBE morbidity in the last decade related to 100,000 inhabitants in the unit. Only those TBE records were included in further analysis where an infected person was an MEP resident and the probable site of infection was located inside the MEP administration boundaries. This approach eliminated cases when the location where the infection happened was different from the home administration unit of the infected person. The entire scope of TBE values in all 206 MEP was divided into six risk categories in such a way that approximately the same number of MEP was inside each category (Table 1).

Tab. 1 Distribution of the relative tick-borne encephalitis (TBE) morbidity in municipalities with extended power (MEP).

Relative TBE morbidity category	Relative morbidity per 100,000 inhabitants	Number of MEP in the category	Average MEP area (km^2)
A	<1.7	42	275
B	1.7–3.3	41	363
C	3.3–6.0	41	347
D	6.0–11.0	39	400
E	11.0–58.0	43	527

2.2 Satellite images and field data

Based on the assumption of an equal distribution of the infected individuals through the tick population, the risk of being infected is expressed by the morbidity. The spatial resolution of the morbidity is given by the respective MEP area, which is rather large and variable. Satellite imagery offers a meaningful alternative to obtain a nationwide thematic map of TBE risk with a substantially better level of detail.

After evaluating the spectral and spatial resolution, as well as the coverage and availability of imagery for the required time frames, data from the Landsat 5 satellite were chosen as the most suitable. It also allowed following the results of previous projects (Daniel, Kolar, Benes 1999; Daniel, Kolar, Zeman 1995) and improving the earlier applied methodology.

Only Landsat 5 scenes containing less than 10% cloud coverage were selected. The time period of the acquisition was restricted to between the years 2006 and 2010, and on dates in late summer (August, September).

The entire area of the Czech Republic was covered with nine Landsat scenes with about 30% and 10% overlap in the longitudinal and latitudinal directions, respectively. The selected satellite scenes (Table 2) were downloaded from the online USGS archive. The Standard Terrain Correction was applied to the data. The geometric quality of all the scenes was checked against the ZABAGED topographic database (ČÚZK 2015) using 12 check points evenly distributed in each scene. Moreover, the correspondence in geolocation was checked in the overlapping parts of the scenes. In both cases, RMSE smaller than one pixel were achieved. Atmospheric correction was not carried out because the classification results were primarily intended for evaluation inside one scene.

Seven three- to five-day field campaigns aiming at the collection of training and control polygons for the supervised classification were carried out from spring 2012 to autumn 2014 in selected forest areas distributed over the whole country. A botanist participating in the first field measurement described particularities of five defined forest classes (Section 3.1). Based on his input, a guideline for their discrimination in the field was created and used during the campaigns. In total, 676 training samples and 321 control points, including the photo documentation, were gathered in 128 different forest locations. All data were stored in a database.

2.3 Meteorological and environmental statistics and supporting data

Though the main focus of the study was on risk assessment of TBE based on forest vegetation categories, possible correlation between morbidity and other factors influencing the tick's life cycle were also studied. These factors comprised three groups of measures available for the entire area of the Czech Republic:

Tab. 2 Landsat scenes selected for the study and supervised classification accuracy (CA) for every scene.

Scene	Date	CA training dataset [%]	CA check points [%]
189-025	2007/06/11	92.48	78.18
	2010/09/23	79.98	60.81
	all	**79.09**	64.58
189-026	2007/08/14	41.22	35.85
	2007/06/11	58.75	41.98
	2009/09/20	53.07	31.58
	all	**80.91**	**53.09**
190-025	2007/05/01	80.80	54.88
	2007/08/01	87.01	73.13
	all	**93.29**	**73.42**
190-026	2007/05/01	50.36	31.92
	2007/08/01	62.32	55.39
	2009/09/27	74.85	57.69
	all	**79.37**	**59.52**
191-025	2009/08/01	63.45	45.80
	2010/09/21	67.19	37.90
	all	**71.50**	**46.28**
191-026	2009/06/14	62.19	38.64
	2009/08/01	52.32	28.00
	2010/09/21	63.70	41.82
	All	**79.37**	**37.50**
192-025	2007/04/29	60.90	38.69
	2009/08/24	63.70	58.94
	2010/10/30	60.33	48.29
	all	**75.87**	**57.36**
192-026	2006/06/13	86.23	63.49
	2010/07/10	85.20	76.81
	2009/09/09	88.75	69.57
	all	70.98	21.39
193-025	2006/07/22	74.8	37.4
	2006/09/24	68.1	41.0
	all	**77.2**	**37.4**

A) Climatic parameters in a 0.5 km grid format provided by the Czech Hydrometeorological Institute:
   – Annual average precipitation in the last decade;
   – Average number of days with temperatures over 10 °C in the last decade (the temperature over 10 °C is typical for tick activity);
   – Annual average temperatures in the last decade.
B) Numbers of hunted game (fallow, roe and red deer, mouflon, wild boar, duck, hare) in MPE areas provided by the Forestry and Game Management Research Institute.

In addition, two other data sets were collected in order to support our investigation and its results:

C) The European land cover database CORINE (EEA land cover 2006). Classes 311 (broad-leaved forest), 312 (coniferous forest), 313 (mixed forest), and 324 (transitional woodland-shrub) covering the majority of forest and shrub vegetation formed the mask of forestry areas utilized for the satellite data classification. The level of correlation between eight selected CORINE classes and values of TBE morbidity in every MEP unit was created to assess their possible mutual relationships (Table 8).
D) A vector layer with MEP borders was one of the most important inputs for the calculation of all statistics connected to the raster data (Landsat classification, CORINE, climate characteristics). It was also used for creating all final map products.

3. Satellite data processing

3.1 Supervised classification of satellite data

Based on the results of a previous project (Daniel, Kolar and Benes 1999), five forest vegetation categories were defined for the supervised maximum likelihood classification:

Class 1 – Coniferous forest
Class 2 – Mixed forest
Class 3 – Heterogeneous young deciduous forest and grassland
Class 4 – Homogeneous deciduous forest
Class 5 – Sparse deciduous forest

Single scenes as well as multitemporal datasets compiled from scenes acquired at different stages of the vegetation cycle were classified. The classification process consisted of the following steps:

1) creating a forest mask based on the CORINE database as mentioned in the Section 2.3,
2) detection of clouds and shadows and subtraction of respective pixels from the forest mask in each scene,
3) principal component transformation applied on single scenes and multitemporal composites; only components containing more than 0.5% information were used for the classification,
4) uploading training polygons and control points collected during the field campaigns,

5) maximum likelihood classification of pixels under the created forest mask; the classifier parameters were tuned for each scene in order to achieve the highest accuracy.

The statistical evaluation of the classification results is summarized in Table 2. The classification accuracy varies from one scene to another. While the total accuracy differs up to 20%, the producer/user accuracy of separate classes changes up to 55%. The coniferous forest class achieved the highest accuracy. Misclassification was typical for heterogeneous young deciduous forests and sparse deciduous forests. Sparse deciduous forest was classified with the lowest accuracy, 20% to 35%.

Classification results of such low accuracy did not provide a reliable input for searching for a relationship between the spatial distribution of classified forest classes and morbidity categories. The results of classification differed considerably in the overlapping areas of the scenes. Changes in atmospheric conditions between acquisition dates and geographical differences across the scenes were two of the reasons causing these discrepancies. Nevertheless, the main reason for the low classification accuracy comes from the spectral variability inside the vegetation categories. Although the class definitions were created by a professional botanist, their verbal descriptions could not always be objectively applied when assigning every forest type to one of the defined classes in the field. Moreover, even when the selected vegetation classes were recognizable and separable in the terrain, their spectral features in the datasets acquired by the Thematic Mapper sensor did not differ sufficiently between one to another to obtain better discrimination. In this case, even an increase of training samples would not lead to much better results. Thus, supervised classification did not bring expected and further applicable results and an unsupervised approach was applied.

3.2 Unsupervised classification

The objective of the research was to develop a process that would allow obtaining the spatial distribution of different levels of risk of TBE. Therefore, it has not been necessary to perform classification of the satellite data into predefined botanical vegetation classes; however, such a legend would certainly facilitate reading of the final map.

The unsupervised approach to the classification of satellite data was also supported by the finding of low portability of spectral features describing defined classes in one scene to another location. Classification into the classes defined previously could not sufficiently reflect the heterogeneity of the forests themselves and their differences between regions throughout the country. Moreover, natural conditions provided by a botanically defined vegetation class may change in space and time, resulting in different occurrences of ticks in a given class in different geographical areas.

To exploit fully the information assets of multispectral satellite data, the approach was to determine recognizable classes based on their spectral differences, only without an effort to give them a botanical name. The spectral features used for supervised classification were the measured relative values in the respective spectral bands (DN) and the normalized vegetation index NDVI, also computed from relative radiometric values:

$$NDVI = (DN_{NIR} - DN_R)/(DN_{NIR} + DN_R),$$

where DN_{NIR} and DN_R indicate the value in the near infrared and red bands, respectively.

The unsupervised classification procedure takes the same first three steps used for the supervised classification (Section 3.1). The unsupervised classification method by Isodata (Jensen 1998) was applied as the fourth step. The preselected number of requested classes was set to nine. The same number of classes was also used for the classification of forest cover in the previous work (Daniel, Kolar and Benes 1999). The decision about this number of different forest categories has also been considered to be large enough for the expression of a rather wide interval of morbidity values.

Classification took place for nine satellite scenes. Thus, altogether, 81 different spectral classes were identified. Each resulting class contains woodland objects with similar spectral properties inside the given scene. Although the number of spectral classes was the same for all the scenes, their botanical compositions are not necessarily the same. This is because the spectral features determining the given class (e.g. class 4) may be different from the spectral features determining the class labeled with the same number in another scene. This approach respects the diversity of the conditions under which the satellite scenes were taken and the natural diversity of the forest vegetation at various scenes.

The classification results are presented in Table 3. The total number of classified values in each scene is given together with its distribution into nine key spectral classes KL_i (i = 1, …, 9). Similar statistics could be computed for every MEP.

4. Evaluation of the relationship between TBE morbidity and other parameters

4.1 Correlation analysis between relative TBE morbidity and environmental parameters

For the purpose of correlation analysis, climate parameters were averaged for each MPE area. The numbers of game hunted were normalized to the area of the forest and agricultural land within an MPE. The relation between relative TBE morbidity and climate and hunted game data was evaluated using Spearman's correlation coefficient r.

The climate parameters did not reveal an expected relation to TBE morbidity (e.g. r = −0.19 for annual

Tab. 3 The number of pixels assigned to classes KL1 through KL9 resulting from the unsupervised classification of the nine satellite scenes.

Scene	Total number of classified pixels	Number of pixels assigned to spectral classes								
		KL1	KL2	KL3	KL4	KL5	KL6	KL7	KL8	KL9
189-25	3 351 577	1 028 610	1 018 292	268 927	644 144	23 536	144 108	174 356	39 907	9 697
189-26	3 869 039	648 114	813 359	156 636	986 777	244 770	34 910	768 890	145 759	69 824
190-25	5 252 992	1 721 685	1 385 177	410 233	957 784	201 788	395 399	76 865	78 706	25 355
190-26	4 518 316	1 192 604	969 033	980 795	203 442	629 035	318 168	155 155	7 584	62 500
191-25	7 606 077	2 555 087	2 034 022	559 522	1 514 383	145 788	315 570	389 665	65 477	26 563
191-26	6 064 862	1 891 916	2 035 849	460 719	979 540	76 353	235 593	292 018	73 051	19 823
192-25	9 728 116	611 875	2 077 203	2 275 441	1 884 782	579 206	1 160 729	567 985	387 976	182 919
192-26	3 984 172	1 380 039	1 029 141	416 724	618 329	183 481	63 752	194 336	72 249	26 121
193-25	3 941 376	1 160 796	1 122 622	370 294	598 363	59 928	253 929	222 475	86 992	65 977

Tab. 4 The average number of hunted game per MPE and the relative morbidity risk categories.

Hunted game	Relative morbidity risk categories				
	A	B	C	D	E
Wild boar	19.9	27.8	34.7	41.1	33.1
Fallow deer	2.6	5.4	3.2	3.5	2.8
Red deer	5.1	2.7	2.3	2.8	2.1
Duck	33.0	56.6	50.6	37.2	121.8
Mouflon	1.4	2.2	1.6	2.7	1.5
Roe deer	29.1	36.7	34.0	27.6	28.3
Hare	101.9	225.8	133.6	84.92	57.7

precipitation, r = −0.15 for the number of days with a temperature above 10 °C). The main reason for such weak relationships is their high variance. Their usually strong heterogeneity within an MEP area was generalized into one value for the purpose of correlation. The distribution of these generalized parameters did not correspond to the normal distribution and their histograms differed from the histogram of the relative morbidity.

For hunted game, the highest correlation coefficient value (r = 0.39) was achieved in the case of wild boar. The absence of a relationship between the studied counts of hunted game normalized to the vegetated area within an MPE and the morbidity risk categories is evident in Table 4.

A correlation analysis between the CORINE land cover classes and the relative morbidity was also carried out. Table 5 summarizes the percentage of selected CORINE classes in an MPE area per morbidity risk category. Coniferous forest (312) is the only class showing a slight proportion of its area with a morbidity risk.

4.2 Relationship of the heterogeneity in a classified image to morbidity

The degree of heterogeneity was expressed for every pixel of the classified image. The number of different classes in the surrounding eight pixels was selected as a measure of heterogeneity. Depending on the number of adjacent pixels belonging to a different class, the central pixel was put into one of the categories of heterogeneity identified, H0 to H8, where H0 contains pixels with only the same class in their neighbourhood while category H8 includes pixels around which all eight surrounding pixels are from different classes.

Tab. 5 Average percentage of CORINE land cover classes in MPE and the morbidity risk categories. The sum of the areas of the selected classes corresponds at least to 94% of an MPE area.

CORINE land cover	Morbidity risk categories				
	A	B	C	D	E
Discontinuous urban fabric (112)	6.8%	5.5%	6.5%	5.3%	3.7%
Non-irrigated arable land (211)	27.1%	48.9%	40.9%	40.5%	36.7%
Pastures (231)	11.4%	5.9%	8.1%	7.3%	9.7%
Agriculture with natural vegetation (243)	12.0%	7.7%	9.0%	8.3%	10.1%
Broad-leaved forest (311)	6.1%	4.2%	4.4%	3.2%	1.8%
Coniferous forest (312)	18.7%	14.1%	15.3%	21.3%	27.2%
Mixed forest (313)	8.6%	6.9%	8.6%	9.6%	6.8%
Transitional woodland-shrub (324)	3.4%	1.9%	2.4%	1.2%	1.2%

Table 6 shows the number of pixels in six heterogeneity categories, H0 to H5 for different categories of relative morbidity. Categories H6 to H8 are no longer listed because they contain insignificantly small numbers of pixels. The highest number of pixels belongs to category H1, with a rapid decrease for other categories. The trend is almost identical for all different morbidity categories. Thus, heterogeneity as defined did not show a potential of being an indicator of relative morbidity.

4.3 Relationship of spectral classes to morbidity

To assess the possible correlation between certain spectral classes and morbidity in MEP areas, the results of the satellite data classification for every MEP were used. Then, a relationship between an area of spectral class occurrence and category of relative morbidity was created for each MEP.

Tab 6 Number of pixels in six heterogeneity categories, H0 to H5, for each category of relative morbidity. The correlation coefficient shows the strength of the relationship between local heterogeneity and relative morbidity.

Categories of relative morbidity	Number of pixels in heterogeneity categories					
	H0	H1	H2	H3	H4	H5
A	1 008 387	1 821 665	1 043 281	410 020	111 381	17 737
B	773 413	1 858 880	1 333 247	603 112	171 635	28 466
C	815 575	1 941 289	1 427 530	631 860	180 703	29 435
D	1 133 586	2 569 743	1 917 263	877 519	256 404	42 574
E	1 484 439	3 710 632	2 669 129	1 264 191	383 868	63 523
Correlation coefficient	0.22	0.33	0.33	0.34	0.34	0.32

Tab. 7 Distribution of spectral classes K1–K9 within the morbidity classes A–E, Landsat scene 189-26.

Scene 189-26									
Categories of relative morbidity	Number of pixels in spectral class								
	KL1	KL2	KL3	KL4	KL5	KL6	KL7	KL8	KL9
A	135105	144551	14752	238651	38738	6814	207544	27515	10786
B	293293	315161	83781	340694	84414	16023	246133	41197	23251
C	179004	279861	41521	284833	88184	8555	233133	55435	22078
D	23701	52194	12111	99596	26187	2675	72627	18763	11700
E	17011	21592	4471	23003	7247	843	9453	2849	2009

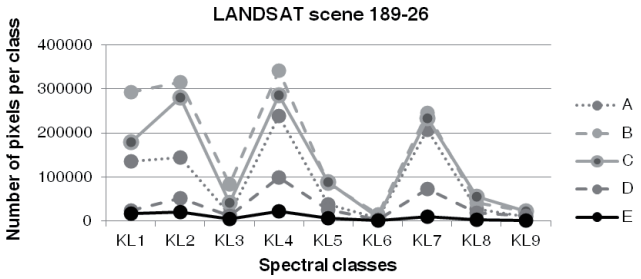

Fig. 1 Distribution of nine spectral classes within the morbidity classes A–E, Landsat scene 189-26.

Tab. 8 Risk associated with the relative morbidity categories expressed as weight coefficient W_j.

Categories of relative morbidity	W_j
A	1
B	2
C	4
D	8
E	16

The example of the relationship for Landsat scene 189-26 is presented in Table 7, giving a number of pixels P_{ij} in each spectral class i (i = 1, …, 9) belonging to certain categories of relative morbidity j (j = A, B, C, D, E). P_{ij} is a measure of the risk of potential disease associated with the given spectral class. The same relationship is presented in Figure 1.

To express the overall level of risk associated with a given class in the relevant scene more exactly, weights Wj were assigned to relative morbidity categories. The weights reflect the absolute values of relative morbidity in each category and are given in Table 8.

The index of the risk of the TBE – IRE was introduced for better expression of the relative risk connected with the specific spectral class inside the satellite scene in which the spectral class was classified. Using the parameters introduced above, the IRE for a given spectral class is defined by:

$IRE_i = (\Sigma P_{ij} W_j) / 10^{11}$, j = A, B, C, D, E

The IRE expresses the level of risk of disease associated with each of the nine classified classes for each satellite scene. For nine scenes there are 81 different values of relative risk indices (Table 9), which reflect the objective diversity of forests throughout the country and the difference between the external conditions under which the satellite data were taken. The IRE values determine which class indicates the areas of most risk; which one means less risky sites, and which class represents places where the risk of disease is almost zero. The relative order of risk is valid for the given single scene. Comparing the risk among scenes cannot be done without corrections for different air and ground conditions at the time of the data acquisition.

When the different conditions among satellite scenes are not taken into account, the IRE values can be presented over merged scene areas. The cost of this simplification is higher when differences in the scenes acquisition or in vegetation types within their territories are larger. Accepting this limitation, however, allows presenting unified incidence risk categories throughout the area of the entire country.

Using this concept, all 81 risk index values were categorized into six intervals of subjective selection (Table 10). Under this distribution there are two values in Category I of the highest risk, seven in Category II, four in

Tab. 9 Index of the TBE risk associated with the nine spectral classes KL1–KL9 in each scene.

Scene	Relative risk in spectral classes								
	KL1	KL2	KL3	KL4	KL5	KL6	KL7	KL8	KL9
189-25	101.8	117.3	4.6	37.9	0.0	2.0	3.9	0.2	0.0
189-26	0.5	1.4	0.1	2.9	0.2	0.0	0.9	0.1	0.0
190-25	13.6	10.6	0.8	5.5	0.3	1.2	0.1	0.1	0.0
190-26	237.5	85.9	31.0	4.4	10.5	7.1	2.0	0.0	0.4
191-25	229.0	150.4	14.7	67.2	1.0	4.8	5.7	0.2	0.0
191-26	192.3	257.4	12.5	70.9	0.4	3.8	6.4	0.4	0.0
192-25	33.9	355.1	442.5	232.1	31.8	70.1	16.2	6.6	1.7
192-26	384.0	248.0	53.7	99.6	8.8	1.0	10.2	1.3	0.2
193-25	10.1	34.8	1.7	4.9	0.0	0.2	0.2	0.0	0.0

Tab. 10 IRE values transformed into six risk categories.

Risk categories	IRE
I – the highest risk	>360
II – high risk	180–360
III – medium risk	90–180
IV – moderate risk	45–90
V – low risk	22–45
VI – no risk	0–22

Category III, five in Category IV, five in Category V, and the remaining 58 values in Category VI, meaning almost no risk.

Spatial distribution of the risk categories according to this model on the area of the Czech Republic is shown as a thematic map in Figure 2. An example of a more detailed view produced on the national topographical map background is presented in Figure 3.

5. Conclusion

Two processing approaches were tested in order to determine the different vegetation classes existing in Czech forests. Supervised classification was applied for five classes defined on a biological basis. The respective training and control datasets were obtained during field surveys in all main forest regions of the country. However, the final classification result did not achieve a useful degree of accuracy in practice. The reason was that even when significantly different from the biological standpoint, the spectral distinction of selected forest classes was too slight in the Landsat spectral bands. Other reasons include the variability inside each class in the field and the diversity of forest vegetation throughout the entire country, as well as the different conditions under which each satellite scene was taken.

As giving a botanical name and description to identified different forest types was not the final goal of the research work, the unsupervised classification approach was applied. In this approach, forest vegetation variations were discriminated only in respect to their spectral variety recorded in the given Landsat dataset. This processing concept respects both the diversity of forest and environment and climate conditions in different places and on different dates of the satellite data acquisition. Treated in this way, the satellite data can be used to assign spectrally various forest areas to degrees of relative risk of TBE. Through the classification of forest vegetation, the one known value of relative morbidity for the whole MEP territory has been decomposed into the hundreds or thousands of smaller areas across the surface of the MEP.

The practical outcome of the project was a thematic map of risk expressed in the relative six-level scale with the spatial resolution of the Landsat Tematic Mapper. The level of spatial detail allows producing maps of the risk distribution at the local or regional level in medium and large scales. No similar information is accessible from the currently provided statistical data. The credibility of the resulting relative levels of risk and their localization in the landscape is better when computed for a smaller area inside one satellite scene.

When one output for an area larger than one satellite scene is requested, a cost is paid for merging different conditions in several scenes and one generalized scale of risk categories. The local conditions can be determined only by direct comparison in the field. Ideally, the number of ticks picked up on the spot would be compared, which is not possible in practice. Therefore, the resulting map content provides a basic overview of the distribution of risk areas in forests over a larger area with reduced accuracy on local level. The index of disease risk (IRE) has been newly introduced for expressing this relationship.

Conclusions about the relationship between spectral classes classified in satellite data and the index of disease risk are also influenced by inaccuracies in morbidity assignment to the class. The statistical data of relative morbidity give a single value for an entire MEP territory of several hundred square kilometers. By contrast, the Landsat data give spectral features of area of dozens of square meters. Very high-resolution satellites even have pixel size of a few square meters.

Further progress in the use of remote sensing methods to determine risk areas for TBE can be expected in

two fields. One involves deeper knowledge about specific vegetation types and other natural parameters influencing the tick occurrence. The other deals with the development of advanced remote sensing measurement equipment required to recognize the vegetation types of interest.

In addition to the printed format, the map of TBE morbidity risk has been published in electronic form on the web site www.access. Based on the Grifinor platform, the interactive map allows browsing the map at a selected level of detail, even on mobile devices.

Acknowledgements

The work was accomplished thanks to a grant from the Ministry of Health of the Czech Republic No. NT11425-5/2010.

REFERENCES

ALTOBELLI, A., BOEMO, B., MIGNOZZI, K., BANDI, M., FLORIS, R., MENARDI, G., CINCO, M. (2008): Spatial Lyme borreliosis risk assessment in north-eastern Italy. Int. J. Med. Microbiol. 298, 125–128. http://dx.doi.org/10.1016/j.ijmm.2008.05.005

BEDIA, J., HERRERA, S., GUTIERREZ, J. M. (2013): Dangers of using global bioclimatic datasets for ecological niche modeling. Limitations for future climate projections. Global Planetary Change 107, 1–12. http://dx.doi.org/10.1016/j.gloplacha.2013.04.005

BROWN, H., DIUK-WASSER, M., ANDREADIS, T., FISH, D. (2008): Remotely-Sensed Vegetation Indices Identify Mosquito Clusters of West Nile Virus Vectors in an Urban Landscape in the Northeastern United States. Vector-Borne Zoonotic Dis. 8, 197–206. http://dx.doi.org/10.1089/vbz.2007.0154

CORTINAS, M. R., MARTA A., GUERRA, M. A., JONES, C. J., KITRON, U. (2002): Detection, characterization, and prediction of tick-borne disease foci. International Journal of Medical Microbiology 291(Suppl. 33), 11–20. http://dx.doi.org/10.1016/S1438-4221(02)80003-0

CROMLEY, E. K. (2003): GIS and Disease. Annual Review Of Public Health [online] 24(1), 7–24. http://dx.doi.org/10.1146/annurev.publhealth.24.012902.141019

ČÚZK (2015): http://geoportal.cuzk.cz

DANIEL, M., KOLÁŘ, J. (1990): Using satellite data to forecast the occurrence of the common tick Ixodes ricinus (L.). Journal of hygiene, epidemiology, microbiology, and immunology 34(3), 243–252.

DANIEL, M., KOLÁŘ, J., BENEŠ, Č. (1999): Predikce míst zvýšeného rizika napadení klíštětem Ixodes Ricinus a nákazy klíšťovou encefalitidou na území Čech. Závěrečná zpráva o řešení grantu Interní grantové agentury Ministerstva zdravotnictví ČR. 30 s.

DANIEL, M., KOLÁŘ, J., ZEMAN, P. (1995): Využití družicových dat pro prognózu výskytu klíštěte Ixodes Ricinus jako základ prevence jím přenášených chorob. Závěrečná zpráva o řešení grantu Interní grantové agentury Ministerstva zdravotnictví ČR. 32 s.

DORMANN, C. F., ELITH, J., BACHER, S., BUCHMANN, C., CARL, G., CARRÉ, G., LAUTENBACH, S. (2013): Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46. http://dx doi.org/10.1111/j.1600-0587.2012.07348.x

EEA land cover (2006). http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster [cit. 2015_10_02]

EISEN, R. J., EISEN, L., LANE, R. S. (2006): Predicting density of Ixodes Pacificus nymphs in dense woodlands in Mendocino county, California, based on geographic information systems and remote sensing versus field-derived data. The American Society of Tropical Medicine and Hygiene [online] 74(4), 632–640.

ESTRADA-PEÑA, A. (2001): Forecasting habitat suitability for ticks and prevention of tick- borne diseases. Veterinary Parasitology 98(1–3), 111–132. http://dx.doi.org/10.1016/s0304-4017(01)00426-5

ESTRADA-PEÑA, A., AYLLÓN, N., DE LA FUENTE, J. (2012): Impact of climate trends on tick-borne pathogen transmission. Frontiers in Physiology 3, 64. http://doi.org/10.3389/fphys.2012.00064

ESTRADA-PEÑA, A., ESTRADA-SÁNCHEZ, A., DE LA FUENTE, J. (2014): A global set of Fourier-transformed remotely sensed covariates for the description of abiotic niche in epidemiological studies of tick vector species. Parasites & Vectors 7, 302. http://dx.doi.org/10.1186/1756-3305-7-302

HERBRETEAU, V. et al. (2007): Thirty years of use and improvement of remote sensing, applied to epidemiology: From early promises to lasting frustration. Health & Place [online] 13(2), 400–403. http://dx.doi.org/10.1016/j.healthplace.2006.03.003

HÖNIG V., ŠVEC P., MASAŘ O., GRUBHOFFER L. (2011): Tick-borne diseases risk model for South Bohemia (Czech Republic). Symposium GIS Ostrava 2011 – Proceeding, VSB – Technical University of Ostrava, 23.–26. 1. 2011, Ostrava.

JENSEN, J. R. (1998): Introductory digital image processing: a remote sensing perspective. 2nd Ed. New Jersey: Prentice-Hall, 1996.

LOURENÇO, P. M., SOUSA, C. A., SEIXAS, J., LOPES, P., NOVO, M. T., ALMEIDA, P. G. (2011): Anopheles atroparvus density modeling using MODIS NDVI in a former malarious area in Portugal. J. Vector Ecol. 36, 279–291. http://dx.doi.org/10.1111/j.1948-7134.2011.00168.x

RANDOLPH, S. E., GREEN, R. M., PEACEY, M. F., et al. (2000): Seasonal synchrony: the key to tick-borne encephalitis foci identified by satellite data. Parasitology [online] 121(1), 15–23. http://ora.ox.ac.uk/objects/uuid:3f09131b-7a72-4910-b7a1-4d5cd17f06db

ROGERS, D. J., RANDOLPH, S. E. (2003): Studying the global distribution of infectious diseases using GIS and RS. Nat. Rev. Microbiol. 1, 231–237. http://dx.doi.org/10.1038/nrmicro776

SCHWARZ, A., MAIER, W. A., KISTEMANN, T., KAMPEN, H. (2009): Analysis of the distribution of the tick Ixodes ricinus L. (Acari: Ixodidae) in a nature reserve of western Germany using Geographic Information Systems. International Journal of Hygiene and Environmental Health 212(1), 87–96. http://dx.doi.org/10.1016/j.ijheh.2007.12.001

ŠVEC, P., HÖNIG, V., DANIEL, M., DANIELOVÁ, V., GRUBHOFFER, L. (2009): Využití GIS pro mapování klíšťat a klíšťaty přenášených patogenů v Jihočeském kraji. Geografie – Sborník české geografické společnosti [online] 114(3), 157–168.

TATEM, A. J., GOETZ, S. J., HAY, S. I. (2004): Terra and Aqua: new data for epidemiology and public health. Int. J. Appl. Earth Obs. Geoinformation 6, 33–46. http://dx.doi.org/10.1016/j.jag.2004.07.001

RESUMÉ

Stanovení míry rizika nákazy klíšťové encefalitidy pomocí metod DPZ

Klíšťová encefalitida (KE) je jedním z vážných virových onemocnění přenášených klíšťaty druhu *Ixodes ricinus*. V posledních letech je dokumentován nárůst tohoto onemocnění jak v České republice, tak v celé Evropě. Výskyt klíštěte je vázán na přírodní podmínky vhodné pro jeho existenci, které lze charakterizovat zejména typem vegetace, dále pak přítomností vhodného hostitele nebo nadmořskou výškou. Cílem prezentovaného výzkumu bylo najít vztah mezi prostorovým rozložením vegetačních tříd lesního porostu a relativní nemocností KE na úrovni obce s rozšířenou působností (ORP). Pro určení výskytu relevantních druhů lesního porostu byla použita data pořízená zobrazujícím radiometrem Thematic Mapper družice LANDSAT-5. Nejprve byla testována řízená klasifikace pro rozpoznání pěti botanikem popsaných tříd lesního porostu. Po porovnání výsledků s pozemními daty získanými během terénního šetření se tato metoda ukázala jako nevhodná vzhledem ke spektrální příbuznosti tříd, jejich odlišnosti na rozloze celého státu a také vzhledem k rozdílným podmínkám, za jakých byly pořízeny jednotlivé družicové scény. Proto byl sestaven metodický postup založený na neřízené klasifikaci, kdy bylo v každé scéně nalezeno devět tříd pouze na základě jejich spektrálního projevu v družicových datech. Výsledkem byl model respektující rozmanitost jak lesního porostu, tak i vnějších podmínek při pořizování družicových dat na různých místech v rozdílnou dobu. Na základě vztahu mezi četností výskytu spektrálních tříd a hodnotami relativní nemocnosti v ORP byl empiricky odvozen index míry rizika onemocnění (IRE), který je exaktním vyjádřením rizikovosti dané spektrální třídy na území příslušné družicové scény. Jako praktický výstup byla vytvořena tematická mapa zobrazující hodnoty IRE pro celé území České republiky s prostorovým rozlišením odpovídajícím datům Landsat, tj. 30 m. Pomocí platformy Grifinor byla mapa zpřístupněna v prostředí Internetu.

Jan Kolář, Markéta Potůčková, Eva Štefanová
Charles University, Faculty of Science
Department of Applied Geoinformatics and Cartography
Albertov 6, 128 43 Prague 2
Czech Republic
E-mails: jan.kolar@natur.cuni.cz,
marketa.potuckova@natur.cuni.cz,
eva.stefanova@natur.cuni.cz

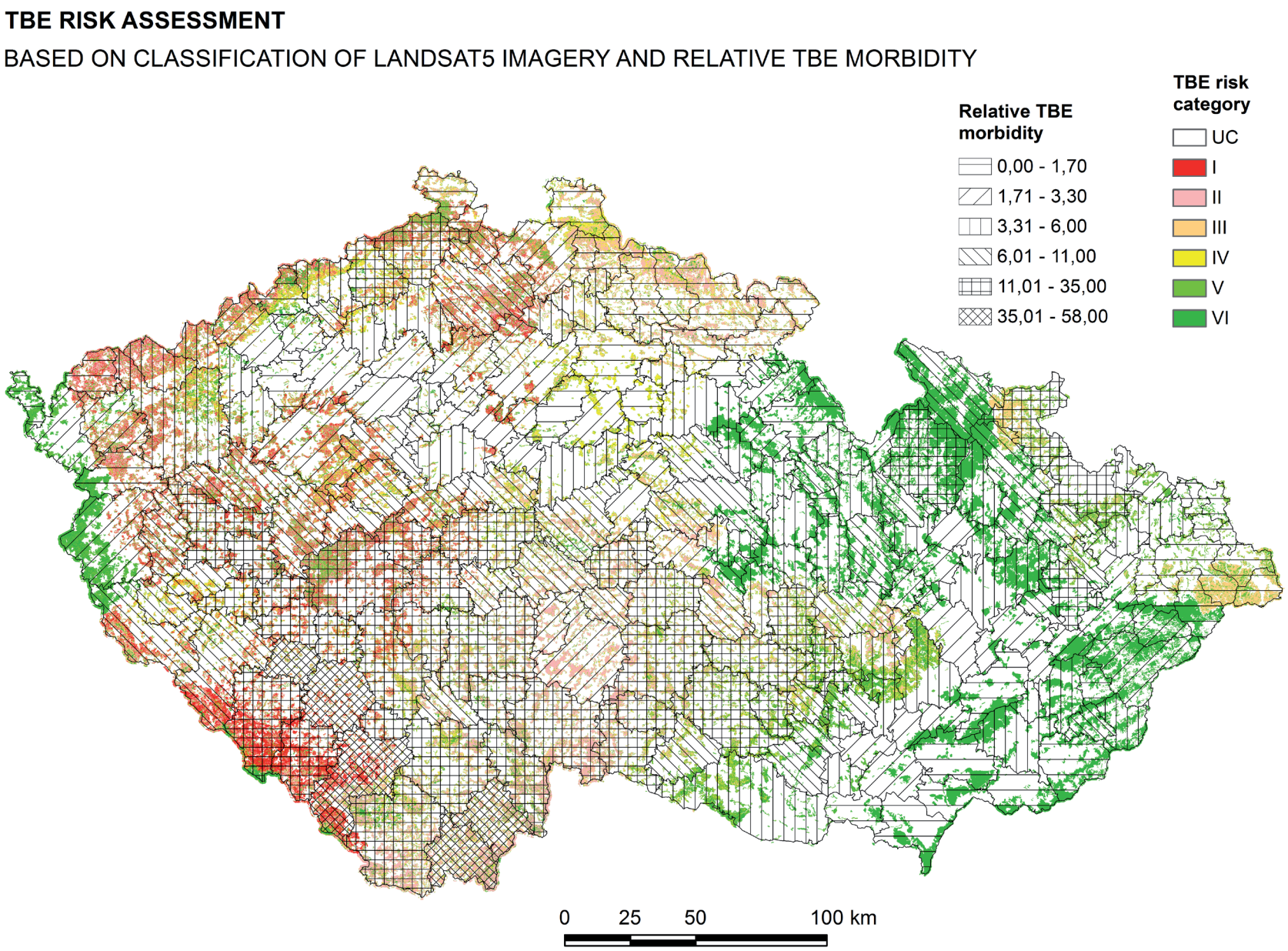

Fig. 2 Relative TBE morbidity and the TBE risk categories distribution in the Czech Republic.

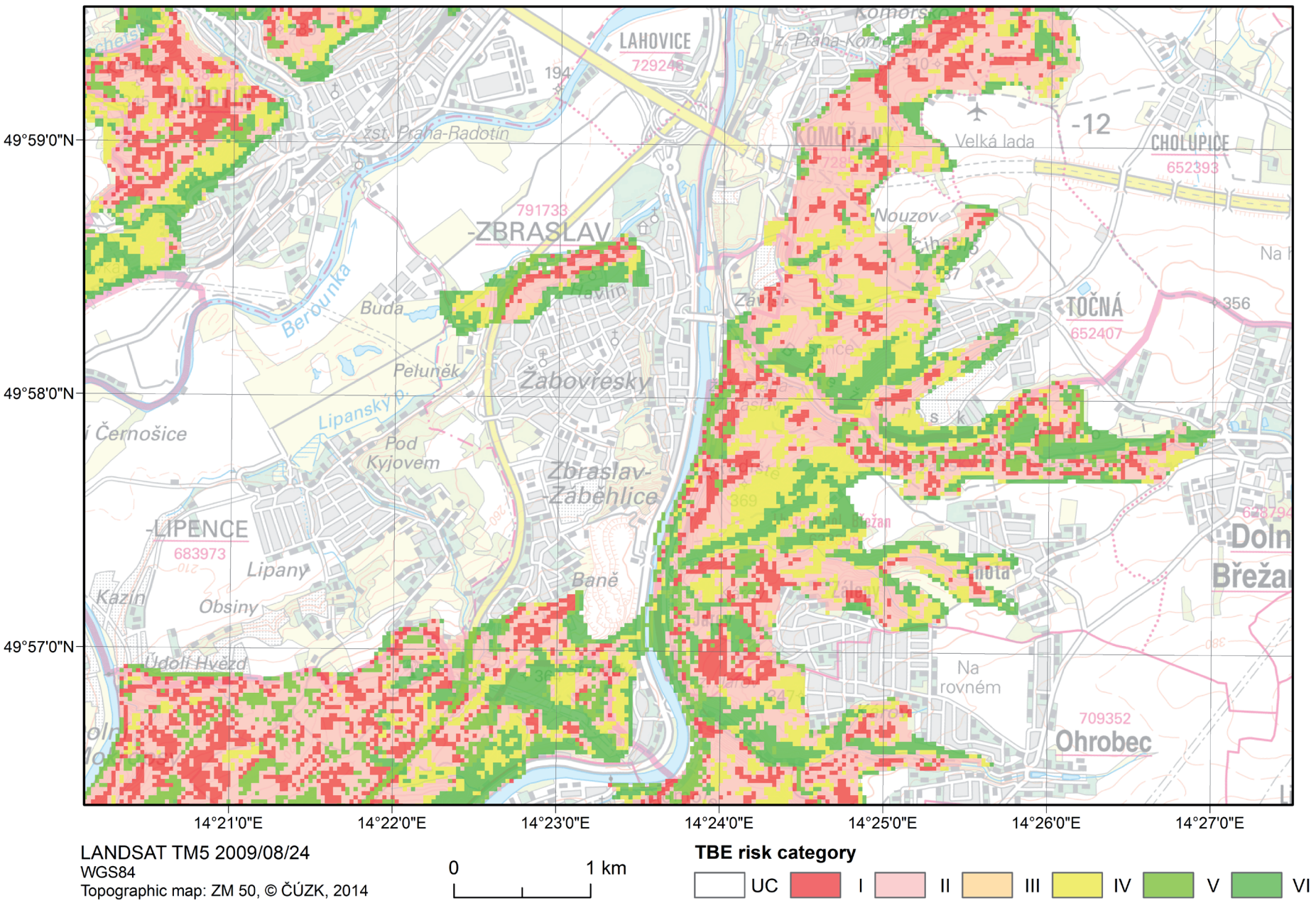

Fig. 3 TBE risk categories distribution in the model area Točná.