International Journal of Environmental Health Research

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cije20

Relating land cover and spatial distribution of nephropathia epidemica and Lyme borreliosis in Belgium

J.M. Barrios [a], W.W. Verstraeten [a b c], P. Maes [d], J.M. Aerts [a], J. Farifteh [a] & P. Coppin [a]

[a] Biosystems Department, M3-BIORES, Katholieke Universiteit Leuven, Willem de Croylaan 34, B3001 Heverlee, Belgium

[b] Climate Observations, Royal Netherlands Meteorological Institute, PO Box 201, NL-3730, AE De Bilt, The Netherlands

[c] Department of Applied Physics, Eindhoven University of Technology, PO Box 513, Eindhoven, 5600 MB, The Netherlands

[d] Laboratory of Clinical Virology, Rega Institute for Medical Research, Katholieke, Universiteit Leuven, Minderbroedersstraat 10, Leuven, B3000, Belgium

Version of record first published: 15 Aug 2012

To cite this article: J.M. Barrios, W.W. Verstraeten, P. Maes, J.M. Aerts, J. Farifteh & P. Coppin (2012): Relating land cover and spatial distribution of nephropathia epidemica and Lyme borreliosis in Belgium, International Journal of Environmental Health Research, DOI:10.1080/09603123.2012.708918

To link to this article: http://dx.doi.org/10.1080/09603123.2012.708918

International Journal of Environmental Health Research
2012, 1–23, i*First* article

Relating land cover and spatial distribution of nephropathia epidemica and Lyme borreliosis in Belgium

J.M. Barrios[a]*, W.W. Verstraeten[a,b,c], P. Maes[d], J.M. Aerts[a], J. Farifteh[a] and P. Coppin[a]

[a]*Biosystems Department, M3-BIORES, Katholieke Universiteit Leuven, Willem de Croylaan 34, B3001 Heverlee, Belgium;* [b]*Climate Observations, Royal Netherlands Meteorological Institute, PO Box 201, NL-3730 AE De Bilt, The Netherlands;* [c]*Department of Applied Physics, Eindhoven University of Technology, PO Box 513, 5600 MB, Eindhoven, The Netherlands;* [d]*Laboratory of Clinical Virology, Rega Institute for Medical Research, Katholieke, Universiteit Leuven, Minderbroedersstraat 10, B3000 Leuven, Belgium*

(*Received 3 January 2012; final version received 6 June 2012*)

Lyme borreliosis (LB) and nephropathia epidemica (NE) are zoonoses resulting from two different transmission mechanisms and the action of two different pathogens: the bacterium *Borrelia burgdorferi* and the *Puumala* virus, respectively. The landscape configuration is known to influence the spatial spread of both diseases by affecting vector demography and human exposure to infection. Yet, the connections between landscape and disease have rarely been quantified, thereby hampering the exploitation of land cover data sources to segment areas in function of risk. This study implemented a data-driven approach to relate land cover metrics and an indicator of NE/LB risk at different scales of observation of the landscape. Our results showed the suitability of the modeling approach ($r^2 > 0.75$, $\rho < 0.001$) and highlighted the relevance of the scale of observation in the set of landscape attributes found to influence disease risk as well as common and specific risk factors of NE and LB.

1. Introduction

Nephropathia epidemica (NE) and Lyme borreliosis (LB) are vector-borne diseases for which awareness has grown in Western Europe as consequence of remarkable outbreaks reported for both diseases in recent years (Mailles et al. 2005; Heyman and Vaheri 2008; Clement et al. 2011). Nephropathia epidemica is caused by the *Puumala* virus (PUUV) hosted by the bank vole *Myodes glareolus*. Humans get in contact with the virus by inhalation of aerosolized dry excreta. Lyme borreliosis is a bacterial infection caused by the *Borrelia burgdorferi* bacterium. This pathogen reaches humans by means of bites of the tick *Ixodes ricinus*. *Myodes glareolus* is one of the competent reservoirs of *B. burgdorferi*, among other rodents, birds, reptiles and big mammals (Rizzoli et al. 2011). The latter aspect is common for both diseases and points at the existence of habitat-related processes affecting the incidence of both diseases.

*Corresponding author. Email: miguel.barrios@biw.kuleuven.be

ISSN 0960-3123 print/ISSN 1369-1619 online

http://dx.doi.org/10.1080/09603123.2012.708918
http://www.tandfonline.com

The observed spatial spread of both diseases and the current knowledge on the transmission mechanisms support the idea of landscape configuration as a crucial determinant of the location of epidemiological risk (Patz et al. 2004; Lambin et al. 2010). Hence, the study of landscape in terms of spatial notions like adjacency and proximity of different landscape elements, density, cover fraction and other, may have important applications in epidemiology. For instance, knowledge on connections between landscape and disease may be the basis for detecting potential disease emergence as effect of landscape management measures and therefore, can support in targeting prevention policies.

Recent studies provide interesting references about the role of landscape configuration in NE/LB and/or in the dynamics of the vector populations, suggesting specific attributes like the presence of forest (Killilea et al. 2008), cover fraction (Jackson et al. 2006; Dearing and Dizney 2010; Guivier et al. 2011), fragmentation (Gerlach and Musolf 2000; Allan et al. 2003; Brownstein et al. 2005; Guivier et al. 2011), and edge effects (Despommier et al. 2006; Guivier et al. 2011). However, those findings are rarely expressed in terms of quantitative units and attempts to determine the values of those metrics that can be associated with levels of risk are even more seldom. The lack of association rules between landscape attributes and NE/LB infection risks constrains the full exploitation of spatial data sources in epidemiological studies as no criteria exist to segment land cover maps and/or space-borne data in function of disease risk.

The objective of this study was to describe the connections between landscape attributes, in terms of landscape metrics, and NE/LB risk. Specifically, we aimed at determining the most influential landscape metrics in the distribution of NE and LB risk and defining levels of disease risk as function of those metrics. In doing so, it was important to consider that landscape metrics exhibit different degrees of sensitivity to the scale at which the landscape is studied (Harrison and Dunn 1993; Wu 2004) and therefore the properties one ascribes to a landscape depend on the scale of observation and the overall sampling setting. Hence, the landscape attributes found to be relevant for epidemiological concerns may differ at different scales of observation. In this respect, the previously stated objective was enlarged to include an evaluation of different landscape sampling settings as they may highlight different aspects concerning the connections between landscape and the diseases under study.

2. Materials and methods

This study aimed at expressing the NE/LB risk as a function of landscape-related attributes. This objective can be represented by the following Equation (1)

$$y_{ds} = f(\{X_{cm} | c \in LCC \wedge m \in LM\}) \quad (1)$$

where *yds* is a vector containing the values of a risk estimator for disease *d* (NE or LB) in each of the sampling units comprising a certain spatial sampling setting *s*. The vector *yds* was tested as response variable of a set of landscape metrics (*LM*) computed for several land cover classes (*LCC*) at each sampling unit. *Xcm* is a subset of landscape metrics that were found to be determinant as explanatory variables of *yds*. Each sampling setting *s* is an array of rectangular sampling units covering the same study area (Belgium) but at different scales of observation.

This procedure is described in more detail in the following sections. Section 2.1 presents the description of the calculation of a disease risk estimator derived from official statistics on the number of NE/LB cases. The *LM* and *LCC* considered in the study are presented in Section 2.2. Section 2.3 explains the conformation of the sampling schemes and Section 2.4 describes the methods for the selection of relevant landscape-related variables and the derivation of the association rules between landscape metrics and disease risk.

2.1. Risk assessment

The graphs in Figure 1 show the aggregated number of reported cases of NE and LB in Belgium for the period 2000 and 2010 and illustrate the remarkable outbreaks recorded in this decade. Throughout this period, the spatial spread of the NE and LB incidence has exhibited a positive spatial correlation, as can be learned from the local Moran's *I* scatterplots presented in Figure 2. These plots show the incidence of both diseases expressed as deviations from the country's arithmetic mean plotted against the average value of the same parameter in the adjacent municipalities (spatial lag). The alignment of most municipalities along a positively sloped line indicates that municipalities are, in general, surrounded by other municipalities with similar disease incidence values. The latter points at the existence of municipality clusters and supports the assumption that risk factors follow a spatial pattern.

The plots shown in Figure 2 are derived from official records on NE and LB cases provided by the Belgian Institute of Public Health (IPH) (Ducoffre 2010) which are geocoded to the patient's municipality of residence. The possibility of being infected

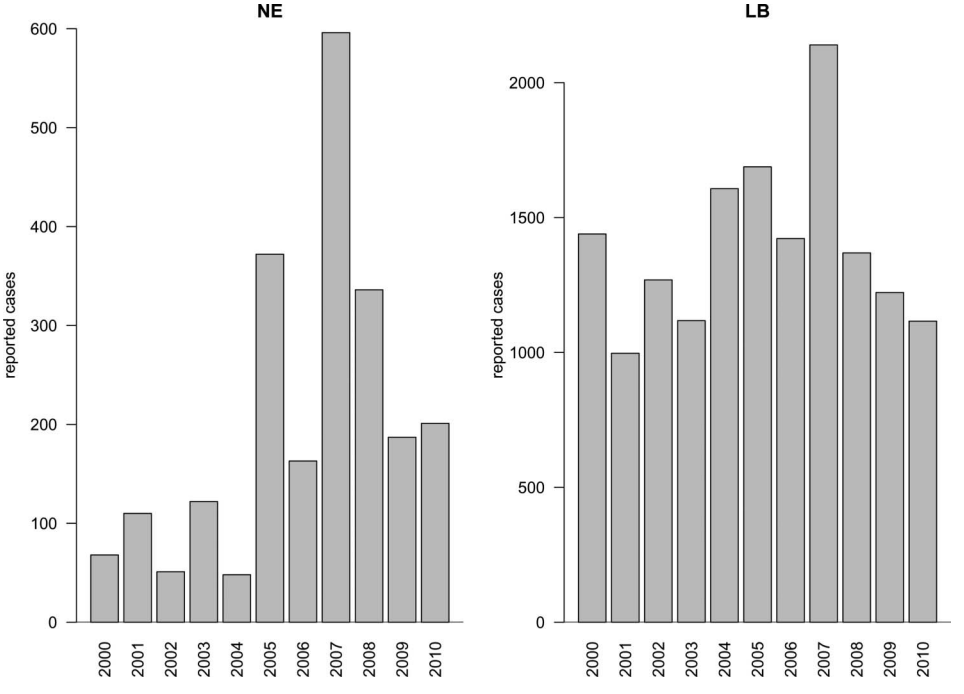


Figure 1. Annual number of NE and LB reported cases in Belgium for the period 2000–2010.

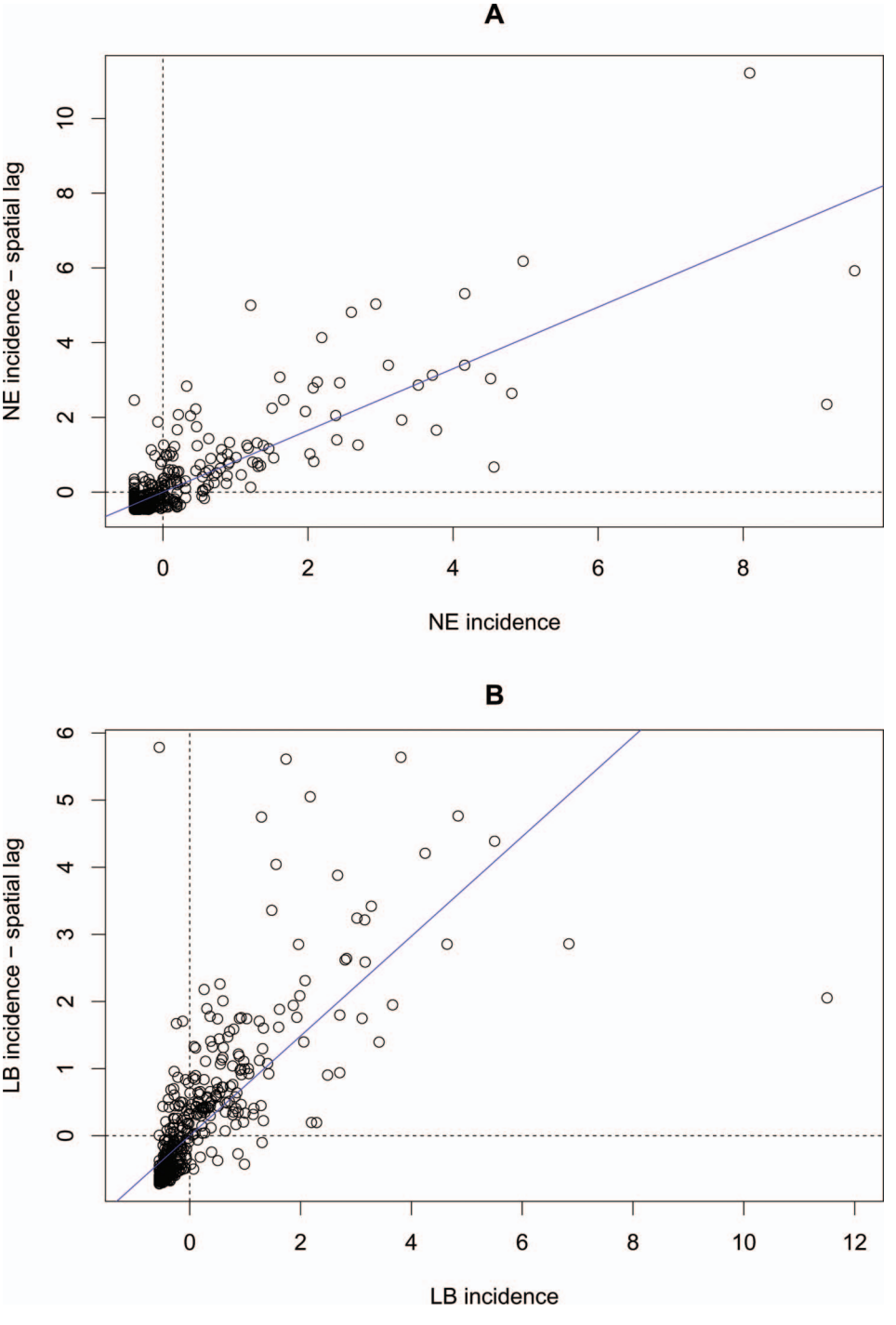


Figure 2. Moran's *I* scatterplot for NE and LB incidence for the period 2000–2010 calculated at the municipal level. The units of the axes are standard deviations from the incidence (number of cases/10,000 inhabitants) arithmetic mean.

outside the municipality of residence adds uncertainty to this analysis, where a link between these reports and the physical environment is explored. Nevertheless, case-control studies for NE and LB indicate that the majority of infected cases occur in the vicinity of homes (Zeman 1997; Crowcroft et al. 1999; Van Loock et al. 1999; Smith et al. 2001; Abu Sin et al. 2007).

The relative disease risk θ has been widely estimated in epidemiology by means of the *standardized morbidity ratio* (*SMR*). It results from the operation $\theta_i = O_i/E_i$; where i is the ith geographic entity (region, municipality, and district) and Oi and Ei are the observed and expected number of cases, respectively. *Standardized morbidity ratio* is a very intuitive and simple estimator but it presents important shortcomings. It is highly sensitive to contrasting population sizes across the geographic entities (Lawson et al. 2000; Leyland and Davies 2005) and the expected value Ei is derived from data of the whole study region (the whole country in our case). The former results in *SMR* values dominated by areas with small populations and the latter neglects the spatial specificities in risk grade.

In search for better risk estimators, local Bayesian methods have been proposed (Leyland and Davies 2005) whereby incidence records from nearby entities are considered as prior information in the computation of the posterior risk. Stated otherwise, the data of geographical entities are weighted in such a way that only the neighboring entities are considered in the estimation of risk. The risk estimator computed in this study is the local empirical Bayesian estimator (*EBE*) proposed by Marshall (1991). Marshall's method shrinks the value of the estimator toward a local mean by considering only the nearby geographical entities in its computation. Hereby, the spatial nature of the disease risk distribution for NE and LB was accounted for. Marshall's algorithm is the following (Marshall 1991):

$$\hat{\theta}_i = \hat{m}_i + (\theta_i - \hat{m}_i)\frac{\hat{a}_i}{\hat{a}_i + \frac{\hat{m}_i}{n_i}} \tag{2}$$

where, $\hat{\theta}_i$ is the *EBE* for municipality i in number of cases/10,000 inhabitants. θ_i was calculated as the ratio between the number of cases to 10,000 person-years at risk (n) in municipality i. $\hat{m}_i$ and $\hat{a}_i$ are the prior mean and variance of relative risk, respectively, calculated over municipalities adjacent to i.

The estimation of person-years at risk n_i in Equation (2) was based on demographic data per municipality concerning the number of inhabitants and the distribution in age and sex classes obtained from official statistical data sources (Belgian Federal Government 2010; European Commission 2011) and on the breakdown of reported cases per age and sex class provided by the IPH (Ducoffre 2010). The value of n_i resulted from a weighted summation based on the proportional contribution of each sex and age class to the overall number of reported cases. Neighboring municipalities were defined following a first order Queen contiguity criterion, i.e. two municipalities were considered neighbors when they had at least one vertex in common.

2.2. *Land cover classes and landscape metrics*

A set of *LM* was computed on patches presented in the CORINE land cover (*CLC*) map (European Environment Agency 2009). The *CLC* is a raster representation of

land cover classes with a cell size of 100 m. The choice of the *CLC* map for this study was founded on the Europe-wide view pursued in its elaboration. This attribute makes it an interesting option for studying landscape structure, as the borders of landscape features are not subjected to the administrative limits constraining national or sub-national land cover maps. The *CLC* map is also the result of a European scale standardization of criteria for the definition of land cover classes and field and processing procedures (Bossard et al. 2000) and therefore, it lends itself to future expansions of the extent of this study. Table 1 presents the land cover classes (*LCC*) and the computed metrics (*LM*) as well as the abbreviations used further on to refer to these metrics.

The calculation of *LM* $j = 1$ to 5 in Table 1 for each of the 10 *LCC* in addition to the metrics $j = 6$, 7, 8 and 9 rendered a matrix of 54 candidate covariates. The definition of *LCC* used in this study follows partially the Level 1 *CLC* legend but, due to the purposes of this study, a more detailed disaggregation of vegetation systems was considered. A brief description of these vegetation types was extracted from *CLC* technical guide (Bossard et al. 2000) and is presented in Table 2.

The selection of *LM* seeks to quantify patch attributes that may impact the vector and host populations and the human–vector interaction. Particularly, attention is given to the quantification of presence and configuration of *LCC* (metrics *COV, DEN, MPS*), the border zone between two adjacent *LCC* (*MEL, PER*) and forest fragmentation (*EMS*). The cover metric *COV* gives an indication of the prominence of a certain *LCC* as constitutive element of the landscape in the sampled area; it is computed as the fraction of the sampled area occupied by each *LCC*. The density *DEN* and the mean patch size *MPS*, expressed in number of cells, provide complementary insight on the configuration of the *LCC* cover. *DEN* is computed as the number of patches present in the sampled area. The edge of vegetated *LCC* is an important landscape feature in epidemiology as it is the location of important ecological processes involving disease vectors and represents spaces of interaction between humans and vector habitats (Despommier et al. 2006). A quantification of the presence of this important landscape

Table 1. Land cover classes (*LCC*) and landscape metrics (*LM*) evaluated as determinants of NE and LB incidence.

i	*LCC*	Abbreviation	*j*	*LM*	Abbreviation
1	Artificial surfaces	*art*	1	Cover	*COV*
2	Green urban areas	*gr.ur*	2	Density	*DEN*
3	Agricultural areas	*agr*	3	Mean patch size	*MPS*
4	Pastures	*past*	4	Mean edge length	*MEL*
5	Heterogeneous agricultural areas	*het.agr*	5	Sum of perimeters	*PER*
6	Broad-leaved forest	*blf*	6	Effective mesh size (*blf*)	*EMS.blf*
7	Coniferous forest	*con*	7	Effective mesh size (*conif*)	*EMS.con*
8	Mixed forest	*mix*	8	Effective mesh size (*mix*)	*EMS.mix*
9	Herbaceous vegetation	*herb*	9	Effective mesh size (*forest*)	*EMS.for*
10	Open spaces with little or no vegetation	*open*			

Table 2. Description of vegetation land cover classes (European Environment Agency 2009).

Class	Description
Green urban areas	Areas with vegetation within urban fabric, includes parks and cemeteries with vegetation and mansions and their grounds
Agricultural areas	Arable land and permanent crops
Pastures	Land permanently used for fodder production
Heterogeneous agricultural areas	Annual crops associated with permanent crops, forest trees, meadows or pastures
Broad-leaved forest	Vegetation formation composed principally of trees. Broad-leaved species predominate
Coniferous forest	Vegetation formation composed principally of trees. Coniferous species predominate
Mixed forest	Vegetation formation principally of trees. Neither broad-leaved nor coniferous species predominate

feature was obtained from the computation of the mean edge length and the sum of patches perimeter per *LCC*. An indicator of forest fragmentation is obtained from the computation of the effective mesh size (*EMS*), introduced by Jaeger (Jaeger 2000). The notion of *EMS* is derived from the calculation of the probability that two randomly chosen points in the studied area are situated in the same undissected patch. If the landscape element of interest is, for instance, coniferous forest, the aggregated probability is calculated as $\sum_i^n \frac{A_i^2}{A_{tot}}$ where A_i represents the size of the ith patch out of the n coniferous forest patches in the sampling unit (studied area) of size A_{tot}. The EMS is the size that subareas would have if the studied area were divided into equally sized areas provided that the probability mentioned earlier remains unchanged. Effective mesh size is thus computed as (Jaeger 2000):

$$\text{EMS} = \frac{1}{A_{\text{tot}}} \sum_{i=1}^{n} A_i^2 \tag{3}$$

This metric was computed separately for each type of forest identified in the *CLC*, i.e. broad-leaved, coniferous and mixed forest, and for a *forest* class that resulted from merging these three forest types.

2.3. Sampling scheme

The landscape metrics mentioned in the precedent section are essentially quantifications of spatial relationships among different *LCC*. Hence, the location and size of sampling units, grain size of *CLC* and the extent of the studied area are spatial attributes in the sampling design that potentially affect the modeled disease risk–physical environment relationship. Research in the field of landscape ecology has proven that the ecological process being observed and/or modeled greatly depends on these spatial attributes (Wiens 1989; Turner et al. 1991; Levin 1992; Harrison and Dunn 1993). Changing the scale of observation of the landscape provides insight on different levels of a hierarchical system and no single scale can be considered the adequate one to study or model the ecosystems (Levin 1992).

Previous works on the identification of landscape features as determinants of NE and/of LB did not elaborate on the implications of the scale of observation. Hence, in this study, sampling units of different sizes were tested as spatial boundaries for the computation of landscape metrics. Note that no variations in extent of the studied area or in grain size of the *CLC* were considered.

The sampling scheme consisted of square non-overlapping tiles covering the Belgian territory. The size of these tiles resulted from $n = 5$ consecutive subdivisions of a square containing the whole country and constrained by the following boundaries: to the north 51° 22′ lat N, to the south 49° 8′ 39″ lat N, to the west 2° 28′ 53″ long E, to the east 6° 32′ 28″ long E. Each subdivision rendered 2^{2n} tiles of equal size. Further subdivisions were judged inappropriate as the tiles would not exhibit patchiness. Moreover, smaller tiles would be completely contained in large municipalities and the metrics derived from them would represent the landscape conditions in only one section of the municipality. This would not be consistent with the computation of local risk estimators (*EBE*) at the municipal level. The first two subdivision levels ($n = 1$ and $n = 2$) were not considered in the analysis as much of the area covered by the resulting tiles was located outside Belgium. Figure 3 shows the location and extent of the tiles corresponding to subdivision levels for n equal to 3, 4, and 5. These arrays will be referred to as scales of observation A, B and C, respectively.

The three grids shown in Figure 3 were overlaid onto the *CLC* and the *LM* described in the previous section was computed for each tile. This resulted in a matrix of landscape-related covariates for each scale of observation. Likewise, the grids of sampling units were overlaid onto the mapped *EBE* values, described in Section 2.1, and the *EBE* value allocated to each tile was the average *EBE* of the overlapping municipalities.

2.4. Tree-structured regression

Modeling the relationship between NE/LB risk and features of the physical environment started with the identification of landscape-related variables that are associated with the spatial distribution of risk. To the authors' knowledge, no precedent research has explored the mathematical relation among the variables considered here and thus, no evidence can support the assumption of a unique mathematical relation across the physical space of a country exhibiting contrasting landscape configurations. The approach followed in this study to screen relevant variables and derive association rules among those variables was the construction of regression trees, according to the procedures described by Breiman et al. (1984) and implemented in the *rpart* package (Therneau and Atkinson 2011) running on the R software.

The basic idea behind the construction of regression trees is the sequential binary partition of the response variable vector into descendant subsets that are more homogeneous than the parent dataset and which can, in turn, be further partitioned until the resulting subsets reach a certain size or level of homogeneity. The dataset being partitioned is frequently referred to as parent node and the two resulting subsets as child nodes which, if further partitioned, become parent nodes of two new child nodes, and so on. For each partition, the algorithm searched across the array of 54 landscape-related covariates for the best variable and value that can serve as partition criterion.

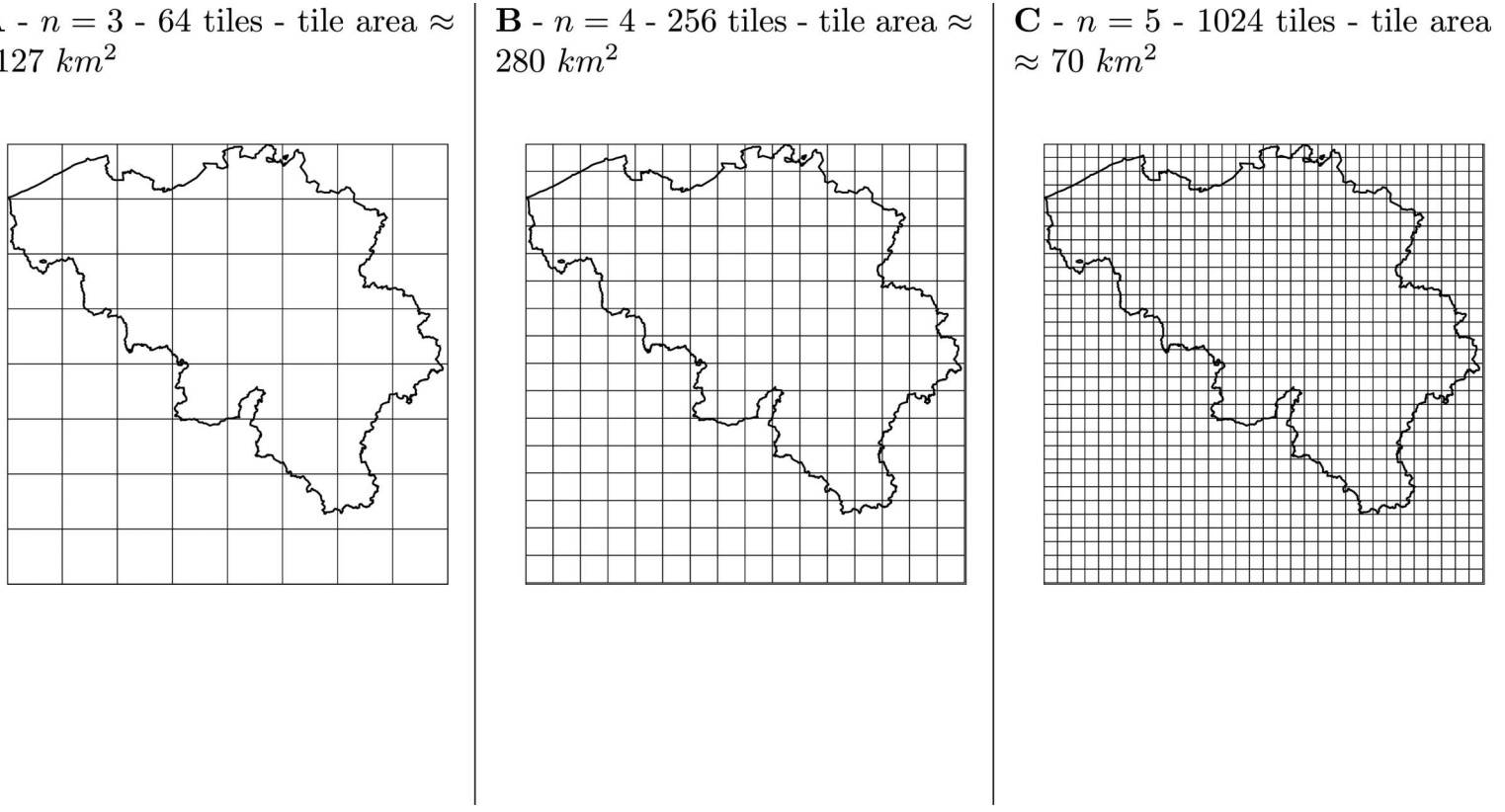


Figure 3. Sampling scheme resulting from partitioning a square covering the Belgian territory into 2^{2n} sampling units when $n = 3$(A), 4(B) and 5(C).

Throughout the fitting process, and at each partitioning node, a maximization of the following expression was sought: $DEVp - DEVch.1 - DEVch.2$; where *DEV* is the sum of the squared difference from the mean of all elements in a node. The subscript *p* refers to the parent node and *ch.1* and *ch.2* to the derived child nodes. The ratio of the former expression to *DEVp* is referred to as complexity parameter (*cp*) and indicates the improvement brought to the model by partitioning each specific node. At each step, the algorithm tests the performance of the model by cross-validation and the general pattern is that the error diminishes as the branching of the tree increases. However, in some cases, the splitting toward homogeneous subsets may lead to excessive branching which, in turn, increases the error as the model gets overfitted. In these cases, already splitted subsets were merged back – a procedure known as pruning – following the criterion of allowing a *cp* value that corresponds to the computed minimum cross-validation error.

3. Results

The maps in Figure 4 show the computed *EBE* values and illustrate the existence of well-defined zones of high incidence for both diseases surrounded by areas where the infection risk seems to diminish gradually with distance. For NE, the zone where disease risk is the highest is the south-western part of the country, in the vicinity of the Franco-Belgian border. In this zone, the LB risk is also high as well as the north-eastern region of the country. From the maps in Figure 4, it is clear that LB risk outside the previously mentioned *hot spots* is higher in a larger portion of the country as compared with NE.

The regression trees pictured in Figures 5 and 6 show the landscape-related variables that were found to determine the spatial spread of disease risk as well as the association rules between disease risk levels and landscape metrics. The association rules are splitting criteria at each node in the form of a statement that might be true or false. If the statement, for instance *COV blf* < 0.12 or *the cover fraction of broad-leaved forests is smaller than 0.12*, is true the direction left from the node should be followed; else, one has to continue to the node at the right hand side of the node. The

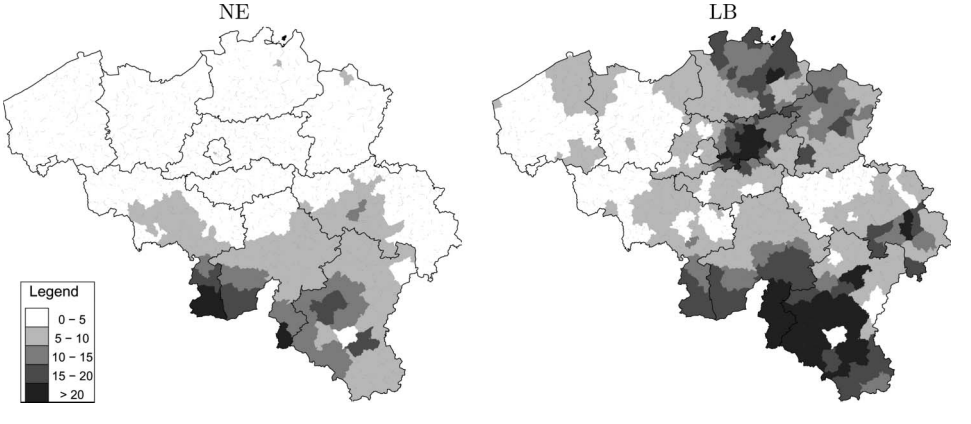


Figure 4. Map representation of *EBE* values for NE and LB in Belgium for the period 2000–2010. The *EBE* values are expressed in number of cases/10,000 inhabitants.

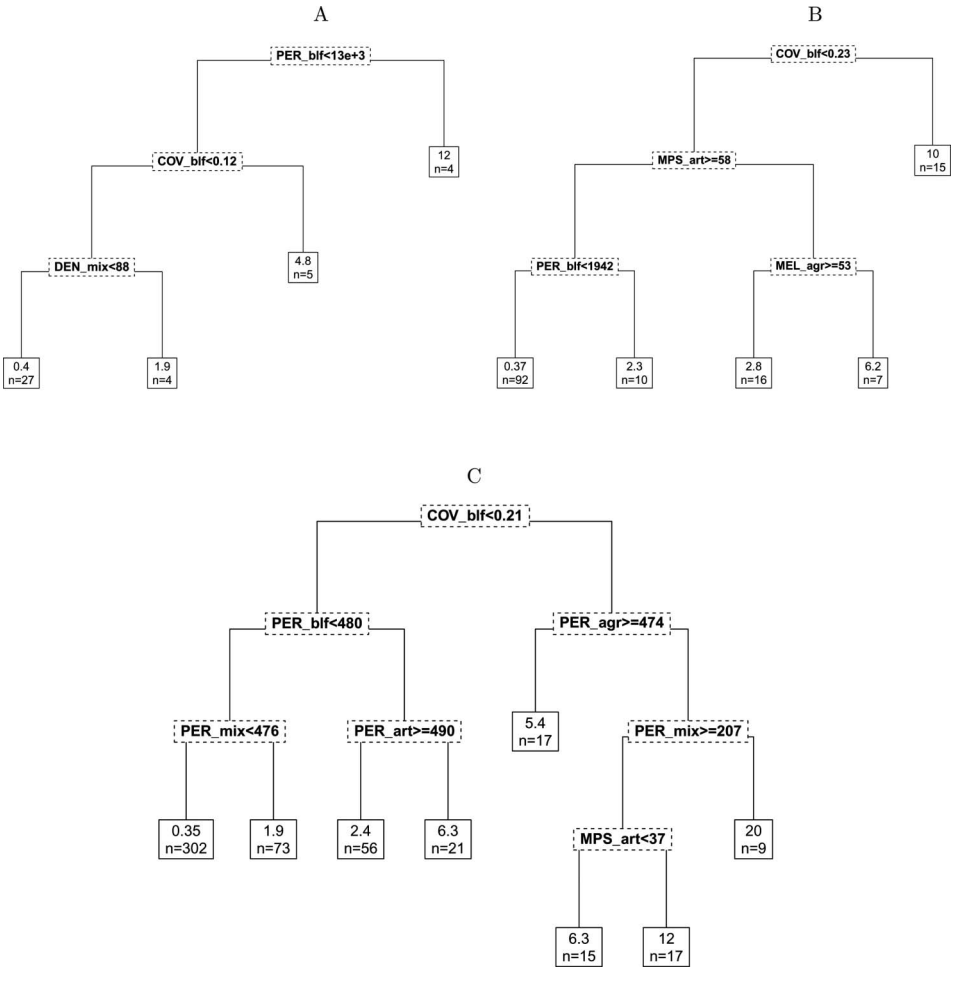


Figure 5. Nephropathia epidemica regression trees at scales of observation A, B and C correspond to the sampling schemes depicted in Figure 3. Labels are described in Table 1. The boxes with dashed lines contain the splitting criteria at each node and the boxes with solid lines are the terminal nodes with the associated *EBE* value (upper line) and the number of sampling units (lower line).

terminal nodes are represented in the regression trees as boxes with solid borders containing two lines with, above, the *EBE* value associated with the sampling units complying with the criteria leading to that terminal node; and, below, the number of sampling units in that subset. The association rules are written in boxes with dashed lines and labeled in accordance to the abbreviations presented in Table 1.

The regression trees in Figures 5 and 6 exhibit different levels of complexity in function of the scale of observation. This can be counted already as an effect of the sampling scheme in the obtained results.

As a general rule, the number of nodes in the regression trees augments as the size of the sampling units decreases. As the sampling scheme allows to focus on more local conditions, i.e. small sampling plots, more terminal nodes are needed and more

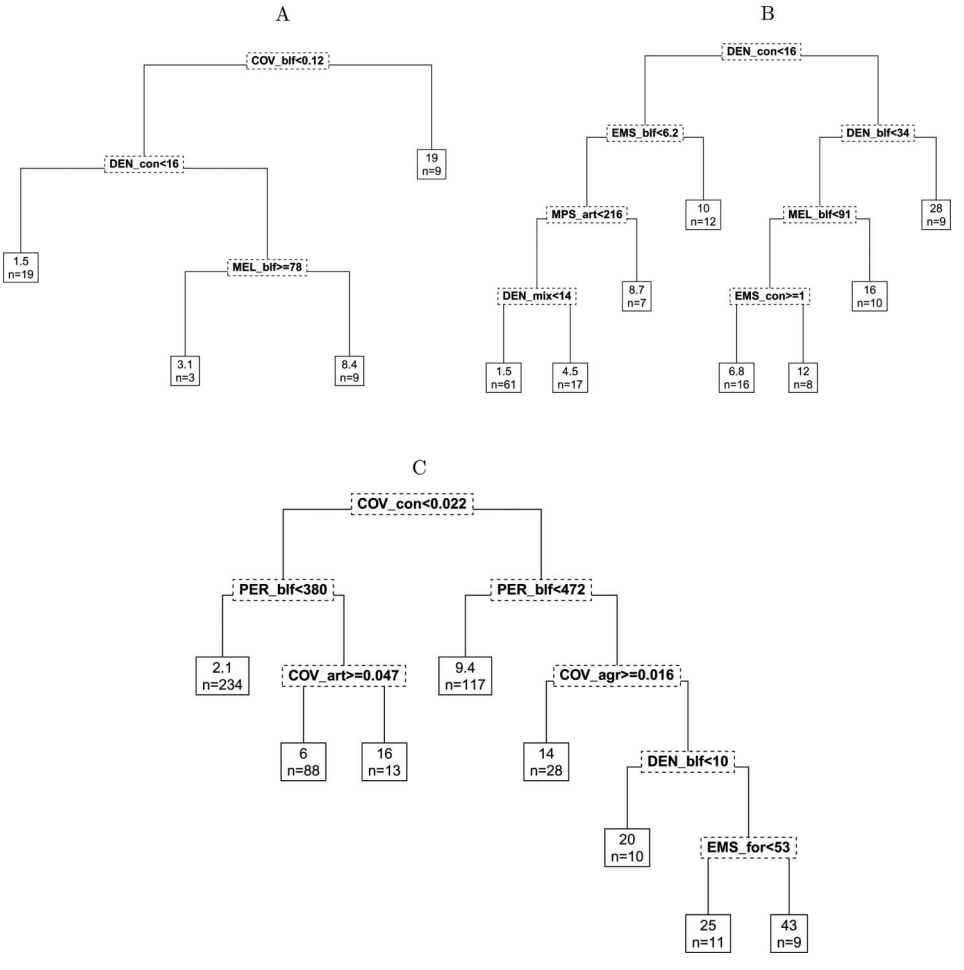
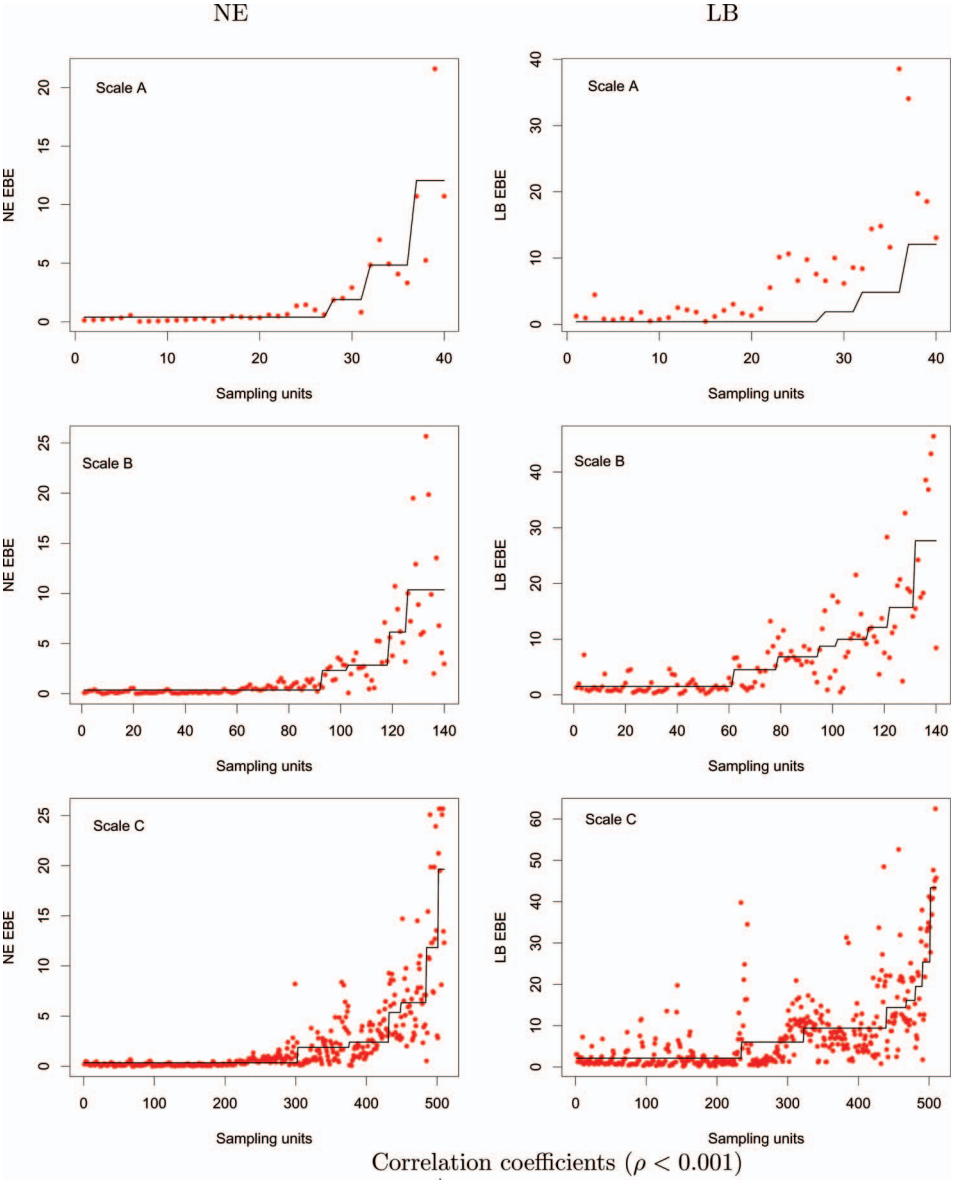


Figure 6. Lyme borreliosis regression trees at scales of observation A, B and C according to the sampling schemes in Figure 3. Labels are described in Table 1. The boxes with dashed lines contain the splitting criteria at each node and the boxes with solid lines are the terminal nodes with the associated *EBE* value (upper line) and the number of sampling units (lower line).

landscape features appear relevant to model the variability of disease risk conditions in the country.

With regard to the agreement between measured and modeled values, the plots in Figure 7 depict the modeled *EBE* values at terminal nodes as horizontal lines and the measured *EBE* values as dots. These plots show a general good agreement between measured and modeled values in most sampling units resulting in r^2 values of 0.76 ($\rho < 0.001$) and higher, also shown in Figure 7. The analysis of these plots also shows that the difference between measured and modeled values is larger in high disease risk zones as compared with zones where the *EBE* values are low. Figure 8 shows this in the form of within-node standard deviations. These disagreements might be partly explained by the fact that most of the variability in the number of cases reported for the period 2000–2010 (and depicted in Figure 1) occurred in zones of high disease incidence. The dynamism in this zone is not only related to land cover pattern but also to interannual variations in forest environmental conditions



Correlation coefficients ($\rho < 0.001$)

Scale	NE	LB
A	0.8766622	0.8289074
B	0.8061257	0.8137778
C	0.8594913	0.7595432

Figure 7. Graphical representation and correlation coefficients ($\rho < 0.001$) of measured (dots) and modeled (lines) *EBE* values at different scales of observation.

concerning food availability (Clement et al. 2010), forest phenology (Barrios et al. 2010) and humidity (Barrios et al. 2012), among others. Therefore, further splitting of these nodes would not add explanatory elements of disease spread and may lead to overfitting.

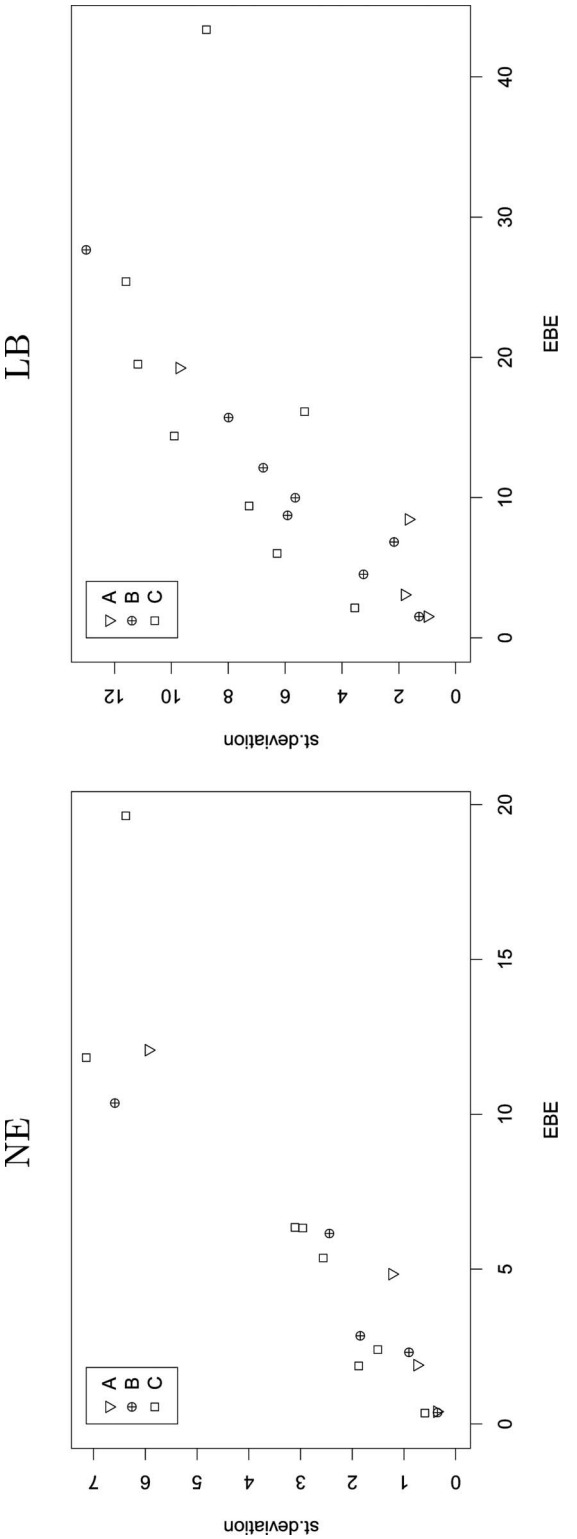


Figure 8. Standard deviations of the *EBE* values within each node at different scales of observation.

The regression trees concerning NE infection risk are consistent at the three studied scales of observation in showing that the most important feature defining spatial spread of risk is related to *blf*. Following the criteria showed in trees in Figure 5B and C, the areas of high NE infection risk are those where the *blf* cover exceeds, respectively, 23% and 21% of the sampled surface area. The root node in tree in Figure 5A indicates that at scale of observation A, the sum of perimeters of *blf* patches was the most important landscape feature to identify the high NE risk areas. This criterion can also be related to a dominance of *blf* cover fraction in areas with the highest infection risk as the tiles with *PER_blf* values larger than 13,000 also have a *blf* cover fraction that averages 25%. A map representation of these high NE infection risk areas at scales A, B and C are shown in Figure 9 (dark gray color). These maps are consistent in the identification of the south-western corner of Belgium as the area with the highest NE infection risk. More detailed definition of the area of highest NE infection risk is obtained with smaller sampling units. According to the maps in Figure 9*b* and, specially, 9*c*, this risk zone covers the total extension of the southern Franco-Belgian border as well as other areas off the border zone.

Areas with less prominent presence of *blf* were associated with a lower, though important, level of NE infection risk. The regression tree in Figure 5A presents an average risk value of 4.8 for areas with $COV\_blf > 0.12$; a similar rule is shown in regression trees 5B and 5C expressed in terms of *MPS_art* and *PER_blf*, respectively. The areas in which these criteria hold are illustrated with light gray colored tiles in the maps of Figure 9. Scales of observation B and C are coincident in allocating these levels of NE infection risk to areas located in the center of the country and a number of locations in the south (maps 9*e* and 9*f*).

The regression trees in Figures 5B and C inform also over other *LCC* determining the spread of NE infection risk. Rules associated with the perimeter of mixed forest patches are components of two nodes in tree 5C and indicate that high values in *PER_mix* may rise the risk. These regression trees contain nodes that refer explicitly to artificial surfaces too. The nodes in the left-hand side of regression trees 5B and C confirm the rural character of NE risk factors by associating lower risk values to units where $MPS\_art > 58$ and $PER\_art > 598$, respectively. However, according to regression tree 5C, in a zone of high NE risk ($COV\_blf > 0.24$ and $PER\_mix < 270$), the size of artificial surface patches ($MPS\_art > 37$) can notably increase the risk. This element can be linked to the importance of human exposure as disease risk factor.

The regression trees that model the connections between LB and landscape are more complex than those of NE in terms of branching levels and diversity of *LCC* identified as determinants of disease. This is related to the large number of organisms acting as tick hosts and competent reservoirs of *B. burgdorferi* and the diversity of habitat conditions in which these organisms can be found. In contrast, bank voles are the exclusive reservoirs and vectors of PUUV.

At scale of observation A, the area of highest LB risk is associated with a coverage of *blf* of more than 12% and is presented in dark gray colored tiles in the map of Figure 10*a*. This risk criterion is similar to that of the highest NE risk and, in consequence, the area of highest risk for both diseases overlaps throughout a considerable geographical surface. The areas with the highest LB risk, as identified for scales of observation B and C, are also associated with an important presence of *blf* but include the coniferous forest as important landscape-related risk factor as

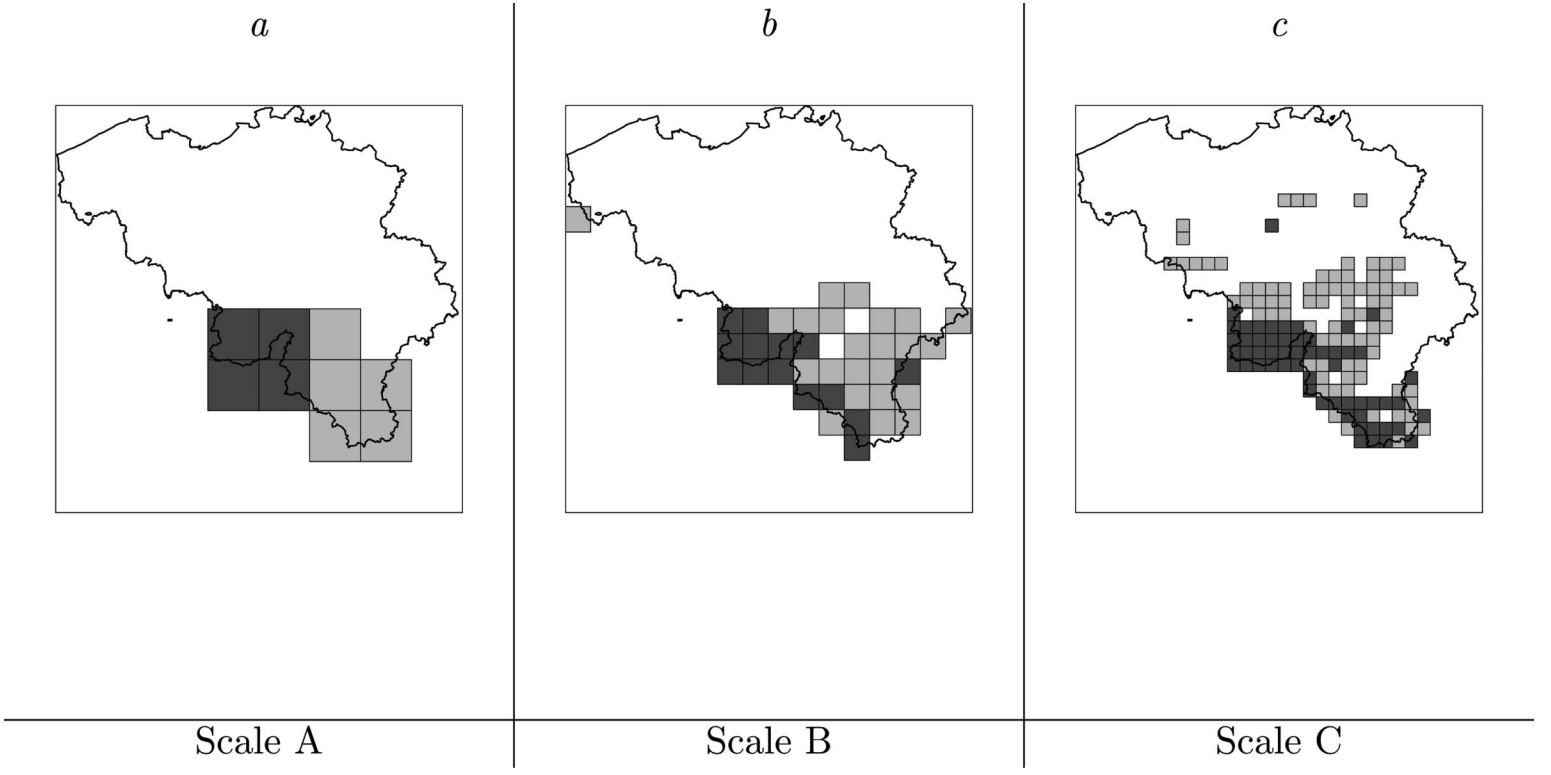


Figure 9. Map representation of the tiles associated with the highest (dark gray) NE *EBE* and moderate (light gray) NE *EBE* as modeled at scales of observation A(*a*), B(*b*) and C(*c*).

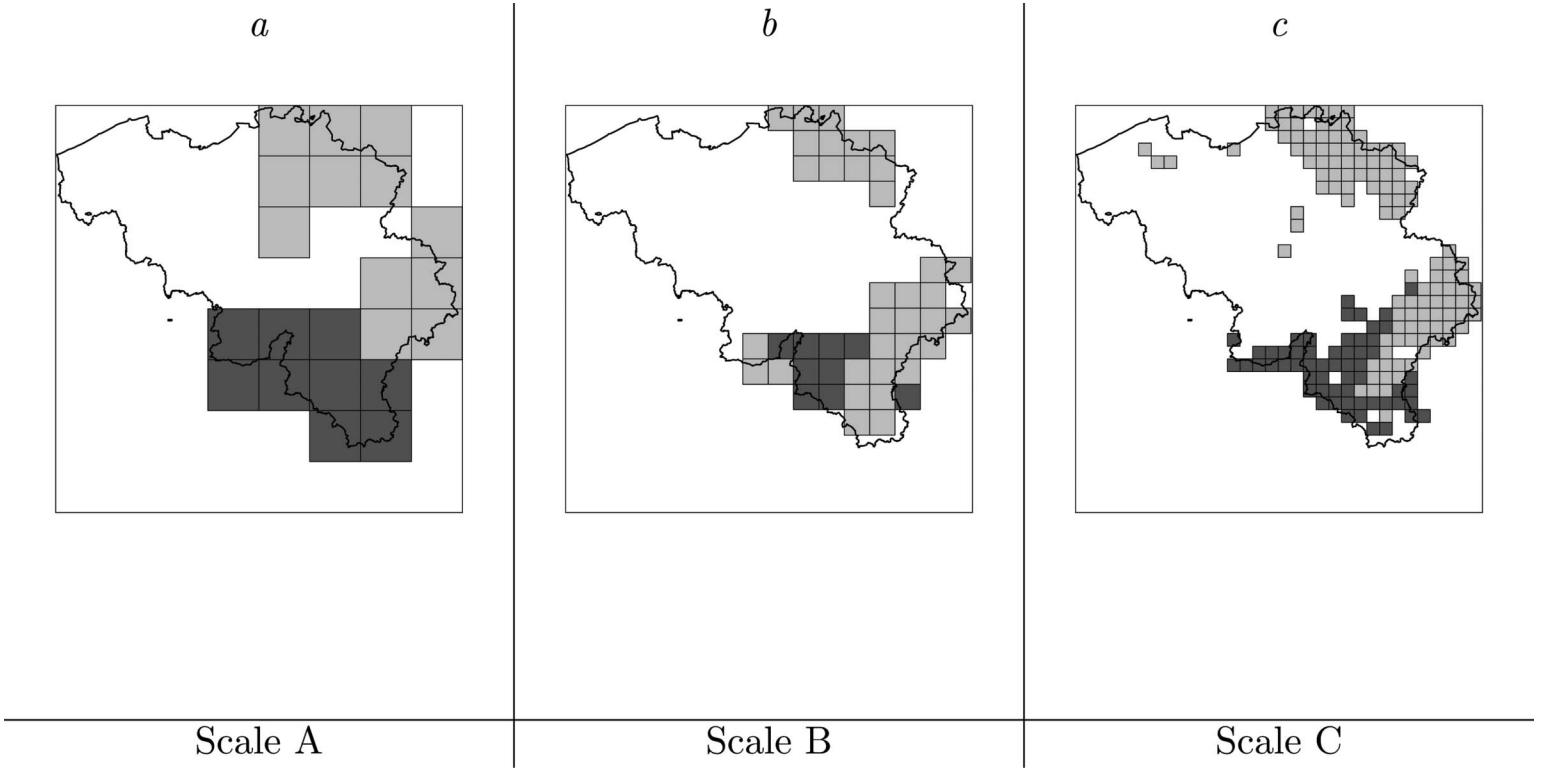


Figure 10. Map representation of the tiles associated with the highest (dark gray) LB *EBE* and moderate (light gray) LB *EBE* as modeled at scales of observation A(*a*), B(*b*) and C(*c*).

well. The maps in Figure 10*b* and *c* show the areas complying with these criteria. The coniferous forests are also an important criterion to separate the risk zone along the southern Franco-Belgian border from other important risk zones. The maps in Figure 10 show the areas in light gray where the LB risk is high (though not the highest in the country) and coniferous is the most outstanding forest type, as found at the three scales of observation.

4. Discussion

The results presented in the precedent section demonstrated the importance of the scale of observation when landscape is studied as condition for infection risk of vector-borne diseases. Only the most contrasting and dominant landscape features are identified as influential in disease spread when large sampling units are used. The complexity of the decision trees increases as the set of sampling units becomes more heterogeneous as a result of more and smaller sampling units. These results also highlighted the existence of common and specific risk determinants for NE and LB. These are important considerations to assess the adequacy and limitations of landscape data sources to be used in epidemiological studies. Forests are the most important landscape features determining the spatial spread of NE and LB. In this respect, the use of forest maps can provide valuable insights when categorizing areas in function of disease risk. However, the results can encounter limitations as other landscape elements, like artificial surfaces and agricultural areas, appeared to impact the infection risk estimations. Figure 11 is an example of landscape configurations associated with different levels of NE and LB infection risk. It was built by following the rules associated with three contrasting nodes in trees 5B and 6B and shows that *blf* is a major component of landscape in zones with high NE and LB infection risk. The mixed forest is also a determinant component of the landscape for NE as well as, for LB, the coniferous forests. The lowest *EBE* values shown in Figure 11 are, for both diseases, associated with higher cover of artificial surfaces and agricultural schemes. In these zones, the forests are not the dominant elements of the landscape.

These considerations can be applied when considering the use of remotely sensed data sources to study the landscape as determinant of epidemiological events as they can guide the selection of data sources and the season of image capture. In virtue of its dynamic nature, space-borne data are an interesting alternative to monitor possible expansions and shrinkages of infection risk areas as consequence of management policies or the natural dynamics of landscape. Previous knowledge on the *LCC* that are linked to disease risk can boost the efficiency in the use of space-borne data, as feature extraction will be targeted to a limited set of relevant landscape elements and, the phenology of vegetation types can guide the selection of the period in which the target landscape elements are accurately detectable.

As mentioned before, common and specific risk determinants for NE and LB were identified. The preeminence of *blf* cover fraction as common landscape-related risk factor was one of the most notorious elements in the regression trees at all tested scales of observations. This is not surprising since *blf* is known to be the preferred habitat of organisms involved in the transmission of NE and LB (Ostfeld et al. 2001; Escutenaire et al. 2002; Lindgren and Jaenson 2006; Clement et al. 2010).

Lyme borreliosis risk is also associated with the neighborhood of forested areas (Lindgren and Jaenson 2006; Killilea et al. 2008), although favorable habitat conditions for ticks and host organisms can also be found in other land cover classes

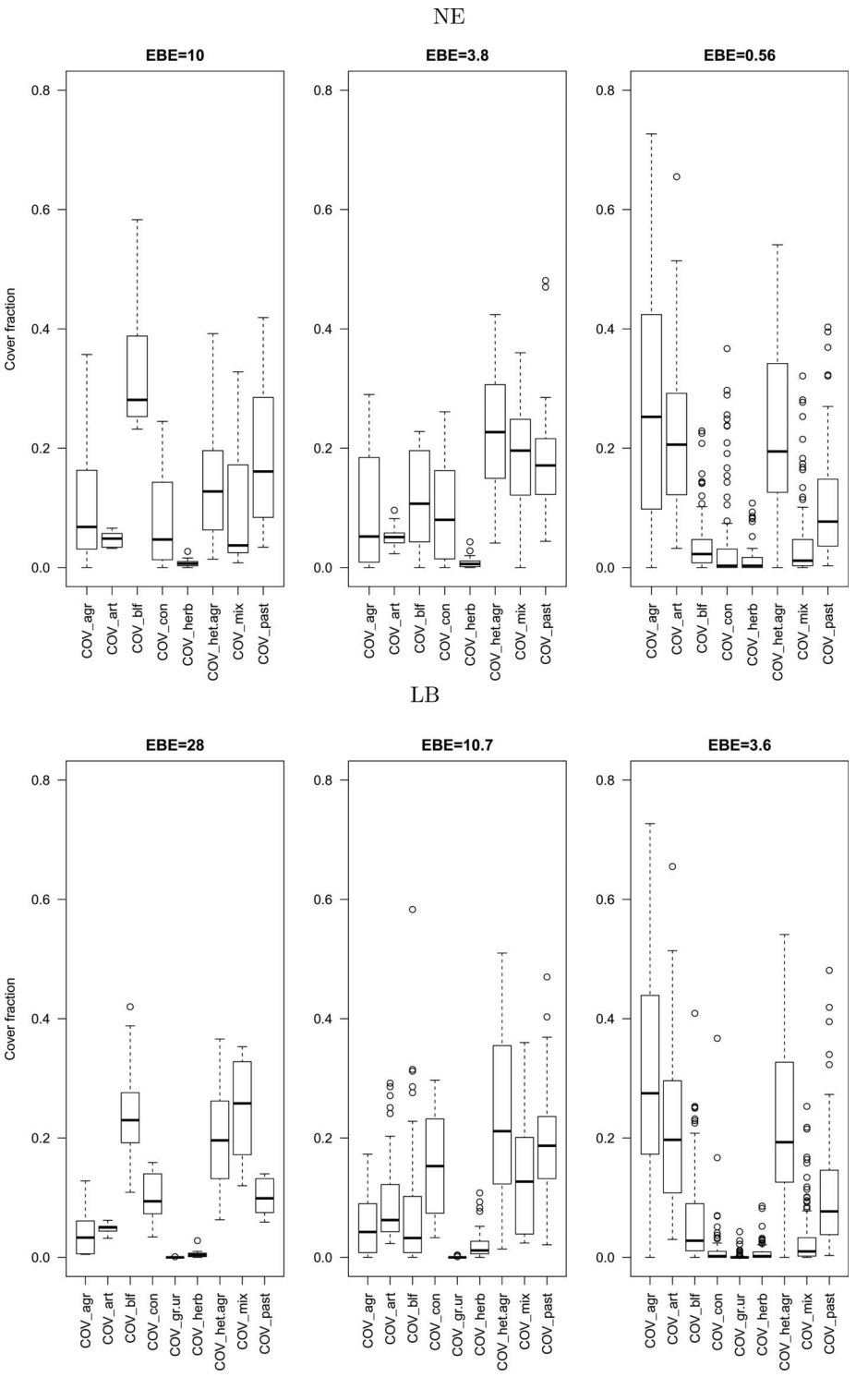


Figure 11. Cover fraction of different land cover classes in zones associated with three contrasting NE and LB infection risk levels.

(Ostfeld et al. 1995). This explains the variety of *LCC* found in the LB regression trees and the larger spread of LB incidence throughout the country. Coniferous forest and agricultural areas appear here as determinant landscape features in LB infection risk in addition to *blf* and mixed forest that are recurrent in NE regression trees.

The regression trees in Figures 6B and C refer to the forest fragmentation indicator (*EMS* or Effective Mesh Size) as a disease risk factor. In these regression trees, the *EMS* nodes are associated with moderate levels of disease risk. The association rules in these nodes can be related to findings by other studies where increased fragmentation in the landscape was related to higher tick-infection prevalence and tick density (Allan et al. 2003; Brownstein et al. 2005).

Several nodes in the regression trees of both NE and LB are built on metrics based on patch perimeters; especially at scale of observation C. The regression tree in Figure 6C, for instance, assigns a high infection risk value to areas where the sum of perimeter of *blf* patches exceeds 472. In the same regression tree, the highest category in the left hand side is assigned to areas with *PER_blf* larger than 380. This underlines the importance of these patch bordering areas as interaction sphere between humans and disease vectors in peri-urban areas. These ecotonal zones also host important ecological processes that can influence disease risk. The importance of edge environments in LB and other infectious diseases has been documented (Despommier et al. 2006) and is related to ecological effects in understorey vegetation (Hamberg et al. 2009) and the abundance of ticks (Lindgren and Jaenson 2006) and rodents (Escutenaire et al. 2002).

A noteworthy aspect of the regression trees at scale of observation B and C is the presence of artificial surfaces. The references to artificial surfaces as risk factors are related to human exposure to vector habitat as a condition of disease risk and this was reflected at scale of observation B and C. This notion is verified by case-control studies aimed at the detection of NE risk factors (Crowcroft et al. 1999; Van Loock et al. 1999); where living in the vicinity of forested area turned out to be a major risk factor. Thus, under similar vector habitat conditions, the vicinity of human dwellings to disease vector habitat can make an important difference in the categorization of areas in function of infection risk.

5. Conclusions

This study aimed at the derivation of association rules between landscape metrics and a computed NE and LB infection risk estimator. The main conclusions of this exercise are the following:

- The approach followed in this study results in association rules derived from the analysis of a large geographical area. Hence, specific conditions in rare landscape configurations may not be properly incorporated in the conformation of the regression trees. Nevertheless, the results were coherent with findings of other studies in which, from a theoretical point of view or based on field results, landscape features have been related to NE or LB incidence. One can therefore be confident about the usefulness of this modeling approach and, in general, the value of spatial datasets to categorize areas in function of epidemiological risk and/or foresee changes in infection risk as a consequence of landscape management policies.

- The use of a data-driven approach was principally motivated by the uncertainty about the nature of the mathematical relationships between landscape features and infection risk. The result was satisfactory and thus, the use of regression trees is recommended as an alternative to explore these connections. An additional advantage of regression trees is the simplicity to communicate and understand the association rules in multidisciplinary work environments.
- This study tested different scales of observation and the modeled results correlated well with the reference values. The major difference among scales of observations was the set of landscape metrics taken in the tree nodes and the complexity of the regression tree. These results should encourage the realization of future studies focusing on the effects of changing other scale-related features (like grain size, shape of sampling units, other sampling settings, etc.) in the study of landscape and disease.

Acknowledgments

The authors would like to thank Geneviève Ducoffre from the Belgian Scientific Institute for Public Health for providing the data on NE/LB cases in Belgium. This research has been supported by the Katholieke Universiteit Leuven (project IDO/07/005). Piet Maes is supported by a postdoctoral grant from the ''Fonds voor Wetenschappelijk Onderzoek (FWO)-Vlaanderen''. Willem W. Verstraeten is supported by a Vidi grant (864.09.001) from the Netherlands Organisation for Scientific Research (NWO).

References

Abu Sin M, Stark K, van Treeck U, Dieckmann H, Uphoff H, Hautmann W, Bornhofen B, Jensen E, Pfaff G, Koch J. 2007. Risk factors for hantavirus infection in Germany, 2005. Emerg Infect Dis. 13(9): 1364–1366.

Allan BF, Keesing F, Ostfeld RS. 2003. Effect of forest fragmentation on Lyme disease risk. Conserv Biol. 17(1): 267–272.

Barrios JM, Verstraeten WW, Maes P, Clement J, Aerts JM, Amirpour Haredasht S, Wambacq J, Lagrou K, Ducoffre G, Van Ranst M, et al. 2010. Satellite derived forest phenology and its relation with nephropathia epidemica in Belgium. Int J Environ Res Public Health. 7(6): 2486–2500.

Barrios JM, Verstraeten WW, Maes P, Clement J, Aerst J, Farifteh J, Lagrou K, Van Ranst M, Coppin P. 2012. Remotely sensed vegetation moisture as explanatory variable of Lyme borreliosis incidence. Int J App Earth Observ Geoinf. 18: 1–12.

Belgian Federal Government [Internet]. June 2010. Structuur van de bevolking [cited 2012 Jul 20]. Available from: http://statbel.fgov.be/nl/statistieken/cijfers/bevolking/index.jsp

Bossard M, Feranec J, Otahel J [Internet]. 2000. CORINE land cover technical guide – Addendum 2000. European Environment Agency. Technical Report 40; [cited 2012 Aug 9]. Available from: http://image2000.jrc.ec.europa.eu/reports/corine_tech_guide_add.pdf

Breiman L, Friedman J, Olshen R, Stone C. 1984. Classification and regression trees. California (USA): Wadsworth, Inc.

Brownstein JS, Skelly DK, Holford TR, Fish D. 2005. Forest fragmentation predicts local scale heterogeneity of Lyme disease risk. Oecologia. 146: 469–475.

Clement J, Maes P, Barrios J, Verstraeten W, Amirpour Haredasht S., Ducoffre G, Aerts J, Van Ranst M. 2011. *Global warming and epidemic trends of an emerging viral disease in Western-Europe: the Nephropathia epidemica case*. In: Casalegno S, editor. Global warming impacts – Case studies on the economy, human health and on urban and natural environments. InTech. p. 39–52; [cited 2012 Aug 9]. Available from: http://www.intechopen.com/books/global-warming-impacts-case-studies-on-the-economy-human-health-and-on-urban-and-natural-environments/global-warming-and-epidemic-trends-of-an-emerging-viral-disease-in-western-europe-the-nephropathia-e

Clement J, Maes P, van Ypersele de Strihou C, van der Groen G, Barrios JM, Verstraeten WW, van Ranst M. 2010. Beechnuts and outbreaks of nephropathia epidemica (NE): of mast, mice and men. Nephrol Dial Transplant. 25(6): 1740–1746.

Crowcroft N, Infuso A, Ilef D, Le Guenno B, Desenclos JC, Van Loock F, Clement J. 1999. Risk factors for human hantavirus infection: Franco-Belgian collaborative case-control study during 1995-6 epidemic. BMJ. 318: 1737–1738.

Dearing MD, Dizney L. 2010. Ecology of hantavirus in a changing world. Ann New York Acad Sci. 1195(1): 99–112.

Despommier D, Ellis B, Wilcox B. 2006. The role of ecotones in emerging infectious diseases. EcoHealth. 3(4): 281–289.

Ducoffre G. 2010. Surveillance des maladies infectieuses par un réseau de laboratoires de microbiologie 2009. Tendances epidemiologiques 1983–2008 [Internet] [cited 2012 Jul 20]. Available from: http://www.iph.fgov.be/epidemio/epifr/plabfr/plabanfr/index09.htm

Escutenaire S, Chalon P, De Jaegere F, Karelle-Bui L, Mees G, Brochier B, Rozenfeld F, Pastoret PP. 2002. Behavioral, physiologic, and habitat influences on the dynamics of puumala virus infection in bank voles (clethrionomys glareolus). Emerg Infect Dis. 8(9): 930–936.

European Commission [Internet]. February 2011. Eurostat Population [cited 2012 Jul 20]. Available from: http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/database

European Environment Agency [Internet]. CORINE land cover 2000. [cited 2012 Jul 20]. Available from: http://www.eea.europa.eu/themes/landuse/clc-download, 2009.

Gerlach G, Musolf K. 2000. Fragmentation of landscape as a cause for genetic subdivision in bank voles. Conserv Biol. 14: 1066–1074.

Guivier E, Galan M, Chaval Y, Xuereb A, Ribas A Salvador, Poulle M, Voutilainen L, Henttonen H, Charbonnel N, Cosson J. 2011. Landscape genetics highlights the role of bank vole metapopulation dynamics in the epidemiology of Puumala hantavirus. Mol Ecol. 20(17): 3569–3583.

Hamberg L, Lehvävirta S, Kotze J. 2009. Forest edge structure as a shaping factor of understorey vegetation in urban forests in Finland. Forest Ecol Manage. 257: 712–722.

Harrison A, Dunn R. 1993. Problems of sampling the landscape. In: Haines-Young R, Green D, Cousins S, editors. Landscape ecology and GIS. London (UK): CRC Press. p. 101–110.

Heyman P, Vaheri A, the ENIVD members [Internet]. 2008. Situation of hantavirus infections and haemorrhagic fever with renal syndrome in European countries as of December 2006. Euro Surveill. [cited 2012 Aug 9]; 13(28):pii=18925. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=18925

Jackson LE, Hilborn ED, Thomas JC. April 2006. Towards landscape design guidelines for reducing Lyme disease risk. Int J Epidemiol. 35(2): 315–322.

Jaeger J. 2000. Landscape division, splitting index, and effective mesh size: new measures of landscape fragmentation. Landsc Ecol. 15(16): 115–130.

Killilea M, Swei A, Lane R, Briggs C, Ostfeld R. 2008. Spatial dynamics of Lyme disease: A review. EcoHealth. 5(2), 167–195.

Lambin E, Tran A, Vanwambeke S, Linard C, Soti V. 2010. Pathogenic landscapes: interactions between land, people, disease vectors, and their animal hosts. Int J Health Geogr. 9(1): 54.

Lawson A, Biggeri A, Boehning D, Lesaffre E, Viel J, Clark A, Schlattmann P, Divino F. 2000. Disease mapping models: an empirical evaluation. Stat Med. 19: 2217–2241.

Levin S. 1992. The problem of pattern and scale in ecology. Ecology. 73(6): 1943–1967.

Leyland A, Davies C. 2005. Empirical Bayes methods for disease mapping. Stat Methods Med Res. 14: 17–34.

Lindgren E, Jaenson T [Internet]. 2006. Lyme borreliosis in Europe. Influences of climate and climate change, epidemiology, ecology and adaptation measures. World Health Organization. Technical Report; [cited 2012 Aug 9]. Available from: http://www.euro.who.int/document/E89522.pdf

Mailles A, Abu Sin M, Ducoffre G, Heyman P, Koch J, Zeller H. 2005. Larger than usual increase in cases of hantavirus infections in Belgium, France and Germany, [Internet] June [cited 2012 Jul 20]. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=2754.

Marshall RJ. 1991. Mapping disease and mortality rates using empirical Bayes estimators. J R Stat Soc Series C, App Stat. 40(2): 283–294.
Ostfeld R, Cepeda O, Hazler K, Miller M. 1995. Ecology of Lyme disease: habitat associations of ticks (Ixodes scapularis) in a rural landscape. Ecol Appl. 5(2): 353–361.
Ostfeld R, Schauber E, Canham C, Keesing F, Jones C, Wolff J. 2001. Effects of acorn production and mouse abundance on abundance and Borrelia burgdorferi infection prevalence of nymphal Ixodes scapularis ticks. Vector Borne Zoonotic Dis. 1(1): 55–63.
Patz J, Daszak P, Tabor G, Aguirre A, Pearl M, Epstein J, Wolfe N, Kilpatrick A, Foufopoulos J, Molyneaux D, et al. 2004. Unhealthy landscapes: policy recommendations on land use change and infectious disease emergence. Environ Health Perspect. 112(10): 1092–1098.
Rizzoli A, Hauffe H, Carpi G, Vourc'h G, Neteler M, Rosa R [Internet]. 2011. Lyme borreliosis in Europe. Euro Surveill. [cited 2012 Aug 9]; 16(27):pii=19906. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19906
Smith G, Wileyto EP, Hopkins RB, Cherry BR, and Maher PJ. 2001. Risk factors for Lyme disease in Chester County, Pennsylvania. Pub Health Rep. 116: 146–156.
Therneau TM and Atkinson B [Internet]. 2011. R. port by Brian Ripley. rpart: recursive partitioning. R package version 3.1–50; [cited 2012 Aug 9]. Available from: http://CRAN.R-project.org/package=rpart
Turner S, O'Neill R, Conley W, Conley M, Humphries H. 1991. Pattern and scale: statistics for landscape ecology. In: Turner M, Gardner R, editors. Quantitative methods in landscape ecology. New York (NY): Springer-Verlag. p. 17–50.
Van Loock F, Thomas I, Clement J, Ghoos S, Colson P. 1999. A case-control study after a hantavirus infection outbreak in the south of Belgium: Who is at risk? Clin Infect Dis. 28: 834–839.
Wiens J. 1989. Spatial scaling in ecology. Func Ecol. 3(4): 385–397.
Wu J. 2004. Effects of changing scale on landscape pattern analysis: scaling relations. Landsc Ecol. 19: 125–138.
Zeman P. 1997. Objective assessment of risk maps of tick-borne encephalitis and Lyme borreliosis based on spatial patterns of located cases. Int J Epidemiol. 26(5): 1121–1130.