ISEIS

Journal of Environmental Informatics 30(2) 107-118 (2017)

Journal of Environmental Informatics

www.iseis.org/jei

Modelling Land-Use Change with Dependence among Labels

H. Omrani[1,4*], F. Abdallah[2], A. Tayyebi[3], and B. Pijanowski[4]

[1]*Urban Development and Mobility Department, Luxembourg Institute of Socio-Economic Research, Esch-sur-Alzette/Belval, Luxembourg*
[2]*Faculty of Sciences, Lebanese University, Beirut, Beirut, 6573, Lebanon*
[3]*Center for Conservation Biology, University of California-Riverside, Saint Louis, MO 63146, USA*
[4]*Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907-2061, USA*

Received 5 September 2014; revised 9 July 2015; accepted 22 January 2016; published online Novermber 2 2017

ABSTRACT. Current literature on land use highlights the considerable methodological challenges in predicting how land will be used in the future. This paper addresses one of these challenges, namely the restrictive nature of the mono-class assignment, in which a spatial unit has only one elementary label at a time. We apply the multi-label concept in which a unit may have several elementary labels. For instance, a spatial unit may belong to residential and commercial classes at the same time. Classes in land use may be correlated, and taking into account their correlation may improve the land use changes prediction. For instance, a spatial unit has more chance to be, or to evolve to a residential unit if it is already commercial. The applied model achieves very promising results, indicated by values of 0.923 and 0.910 for precision and recall, respectively. The application described in this paper demonstrates the advantages of modelling the dependence among the labels for predicting the land use change.

Keywords: Bayes rule, dependence between labels, *k*-NN, land use, multi-label

1. Introduction

Land use change has been extensively studied over the last 40 years (Lee, 1994, 1973; Turner et al., 1995; Watson et al. 2000; Green et al., 2002; Foley et al., 2005; Verburg, 2006; Brown et al., 2013). Since land use change can affect critical environmental processes such as water quality (Tong and Chen, 2002; Tang et al., 2005), biodiversity (Reidsma et al., 2006), regional climate dynamics (Watson et al., 2000), food security (Ericksen et al. 2009; Tayyebi et al., 2016a, b), and hydrological processes (DeFries and Eshleman, 2004), to name a few, the key aim of a land use change model is to predict the evolution of landscape changes in the future to support decision making, like community land use planning (Veldkamp and Lambin 2001; Verburg et al. 2004; Platt 2004; Song et al., 2015). Experts from different disciplines (computer science, engineering, geography, landscape ecology, and others) have contributed to this field and have applied several methods (Pijanowski et al., 2002, 2010). In the last two decades, machine learning techniques have been introduced to predict changes in land use as the process is known to be complex, nonlinear and large databases now exist that assist in quantifying change. For example, several modelers, Li and GarOn Yeh (2004) and Basse et al. (2014), have discussed the advantages of using advanced data mining tools for land-use modelling. They all describe how to build a functional relationship (or decision function) that captures the pattern of the land-use changes which is then used to mimic the evolution of land use over time scales that matter. Traditionally, single-label learning methods were used to model such patterns and their dynamics. These methods assign a single label to each spatial unit. However, this is not practical because mixed use of land is currently quite common (Omrani et al., 2015a, b) and this information is often considered in land use planning decisions. For example, urban and industrial areas may intermingle, agricultural land may be adjacent to a forest, and the like (Tayyebi and Pijanowski, 2014). A more detailed classification would be desirable (Valipour et al., 2013). The multi-label concept allows us to assign to each cell (also known as "example", "observation" or "instance") a set of target labels. Unlike single-label learning (Schneider and Pontius, 2001), multi-label learning assumes that labels are not mutually exclusive, and a cell of land may be associated with several labels simultaneously. For instance, a cell may belong to residential and commercial classes at the same time. Several recent applications confirm that multi-label modelling is more relevant for complex phenomena by providing a finer and more realistic description.

Multi-label classification is also appropriate when the resolution of the data is not sufficiently high, when the cells are too large (Briassoulis, 2000), or when they contain borders of clearly delimited areas that are subject to different use. Multi-labeling approach has been used in other related disciplines. For example, Jones et al. (2011) applied multi-label classifica-

* Corresponding author. Tel.: +352 585855313; fax: +352 585855700.
E-mail address: hichem.omrani@liser.lu (H. Omrani).

ISSN: 1726-2135 print/1684-8799 online
 doi:10.3808/jei.201700379

tion to model the distribution of multiple species in a landscape. This distribution plays a key role in creating reserves for species conservation, predicting the effects of ecological change, and testing ecological theory. Yang et al. (2012) applied multi-label classification to environmental modelling in relation to sustainable flood retention basins.

In multi-label learning framework, classes are usually correlated: one label appears with another more or less frequently than what independent assignment of labels would imply. A key challenge for a multi-label learning method is to exploit the information about the correlation among different classes. For example, a cell is likely to be labeled as urban if it belongs to class "road".

Although the issue of dependence among classes has shown considerable merit in several applications (Herold et al., 2005), but until now, no studies have applied it in land use change science within a framework of multi-label learning. Therefore, this paper applies for the first time the multi-label concept to land-use change with dependence among labels adapted to make a prediction for land use change. In this paper, we study land use in the Grand Duchy of Luxembourg. The cells used are squares of 1 hectare each; there are 255,698 of them. Of these, 4385 cells (1.7%) cross the borders of the country and the classification is based on the land use in their parts belonging to the country. Each cell is classified into four classes of land use: agriculture (A), forest (F), industrial (I) and urban (U). Further, a set of 13 variables is defined for each cell. These variables and the classification are available for years 1999 and 2007. We relate these variables to the classification in 1999 and classify cells in 2007 assuming that the same association applies in 2007. Since the classification of the cells in 2007 is known, we can assess the quality of this method.

The results of our analysis demonstrate the effectiveness of the modelling framework to exploit the dependence among labels. It offers a new approach that is intuitively more realistic and attractive. The key objectives of this paper are threefold:

- Construct a multi-label dataset from a vector map;
- Improve class assignment in land-use modelling by applying the multi-label concept;
- Predict and estimate land-use change with dependence among labels.

Our solution entails both methodological and empirical novelties. The remainder of this paper is organized as follows. The next section discusses further the advantages of using multi-labels in the study of mixed land use. Section 3 describes the modelling framework used for predicting land-use changes. Section 4 presents the multi-label dataset construction from a vector map. Section 5 presents the application and the results. Finally, Section 6 presents conclusions, a discussion and proposes further directions for this area of research.

2. Multi-Label Concept for Class Assignment

To remove any ambiguity about the terms used for land-use modelling, we describe the binary, multi-class and multi-label learning concepts. The binary learning concept is used when every cell has one of two labels that are exclusive and complementary. For example, a cell is classified as either built-up or non-built-up land. The multi-class concept deals with more than two classes, e.g., to classify a cell of land as urban, industrial, agricultural or forest.

Both binary and multi-class learning techniques make the assumption that each cell has a single label; a cell can be either urban or industrial but not both at the same time. Both binary and the multi-class learning correspond to single-label or mono-label concept. Figure 1 displays an example of a classification issue with two classes that overlap in the feature space. In single-label learning illustrated in Figure 1(a), the overlapping classes cause classification errors, while in multi-label learning, Figure 1(b), the classes overlap (are not exclusive). For multi-label data, the membership of a cell to more than one class is not due to ambiguity (fuzzy membership), but to multiplicity (full membership) (Boutell et al., 2004). The traditional supervised learning (binary or multi-class) can be regarded as a special case of multi-label learning, in which each cell can have only a single label. The multi-label concept generalizes both binary and multi-class concepts. By allowing a cell to have more than one label at a time, one of the major drawbacks of the land-use models is overcome (Omrani et al., 2015a; Omrani et al., 2017), and areas of mixed use can be represented without a compromise.

2.1. Relationships Between Labels

Furthermore, in multi-label learning, the possibility of joint membership of a cell to several classes implies the existence of information in the label space about the dependence (or correlation) among the labels. The assignment of a cell to a class may provide information about the membership of that cell in other classes. Labels A and B are said to be dependent if the probability (frequency) of a cell having both A and B as its labels is not equal to the product of the probabilities of having labels A and B at the same time: $P(AB) \neq P(A) \times P(B)$. Label dependence is present when the probability of a cell to have a given label depends on its having also the other label. For example, a cell with the label 'forest' is unlikely to be labeled also as 'urban', but the probability that the cell also has the label 'agriculture' is higher. Label correlation (dependence) can be represented in the form of a contingency matrix CM. An element of CM is the conditional probability P(AB). Since labels in land use tend to be correlated, it is judicious to use multi-label learning methods that take this correlation into account. One such method is DML*k*NN; it stands for Dependent Multi-Label *k*-nearest neighbour. It is described in the next section.

3. Method: Multiple Labels and DML*k*NN

This method is based on the Bayesian *k*-nearest neighborur (*k*NN) rule (Beyer, 1999). The *k*NN algorithm is a simple classification method. For each cell in the testing set we defi-

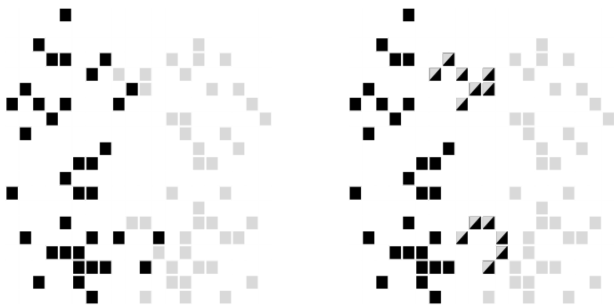

Figure 1. (a) Single-label classification with two overlapping classes (black and light-gray squares), and (b) multi-label classification with some cells belonging to both classes (squares split by the two colors).

ne its distances from all the cells in the learning set. For example, if the learning set contains 200 cells and the testing set has 5000 cells, then we have to evaluate a 200 × 5000 matrix of distances. We refer to a cell from the testing set as unseen. For each unseen cell u, we identify k cells in the learning set that are closest to u. These cells are called the (k nearest) neighbours (of u). For example, k may be set to 10. An unseen cell may be classified by the label that is in the majority among its k neighborurs. The method DML*k*NN is an advanced Bayesian version of the basic *k*NN method (Younes et al., 2011); details are given Section 3.2. It is a probabilistic multi-label classification method that exploits information about label (i.e., class) dependence. We give the main steps of the method in the next section.

3.1. The Algorithm

Let $X = \Re^d$ be the domain of d-dimensional cells of features and let $£ = \{l_1, ..., l_Q\}$ be the set of labels. The set $S = \{(x_1, Y_1), ..., (x_n, Y_n)\}$ is the multi-label dataset containing n cells. They are considered to be identically distributed and drawn independently from $X \times 2^£$ where $x_i \in X$ and $Y_i \in 2^£$. The cells of the set S are split into learning/training L (calibration) and test T (validation) sets. The split is accomplished by stratified random sampling with simple random sampling within each class of land use (stratum). We fit the model to set L and evaluate its performance on set T using several performance metrics. We replicate this process of splitting, fitting the model and evaluating its performance N times, obtaining N sets of the metrics. Their means (and standard deviations) are the overall evaluations. Figure 2 shows the procedure of modelling land-use change using the DML*k*NN method.

The method learns a multi-label classifier C: $X \rightarrow 2^£$ using the given learning dataset L, which predicts a set of labels for each unseen cell $x \in X$ (Zhang and Zhou, 2007; Spyromitros et al., 2008). In addition to C, DML*k*NN defines a scoring function g: $X \times £ \rightarrow \Re$, assigning a real number to each (cell, label) combination. For each label $l \in £$ the score $g(x, l)$ indicates the level of relevance of label l for the cell x.

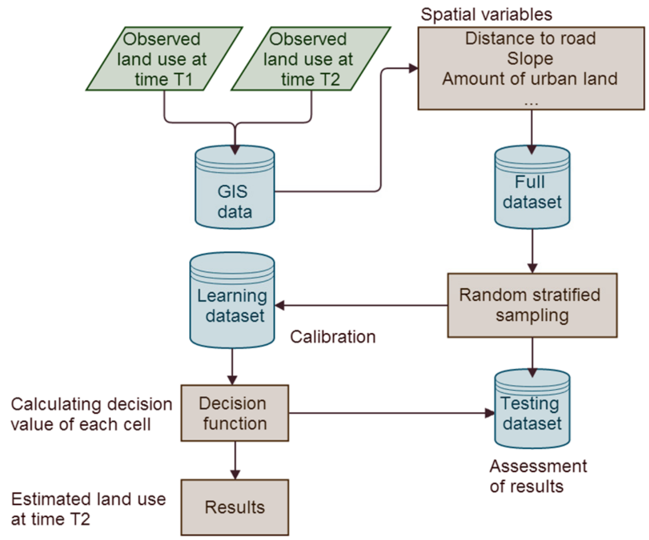

Figure 2. Flow chart of the procedure for modelling land-use change using the DML*k*NN method.

For a given threshold value t, the multi-label classifier $C(.)$ and the scoring function $g(.,.)$ are related by the following equation:

$$\mathcal{C}(x) = \{ l \in \mathcal{L} | g(\boldsymbol{x}, l) t \} \quad (1)$$

label l is associated with cell x if $g(x, l) > t$. The threshold value t is set by cross-validation or heuristically (Fan and Lin, 2007).

For a given cell x and its label set $Y \subseteq £$, denote by N_x^k the set of the k nearest training cells of x in $\mathcal{L}$ according to a specified distance function $d(...)$, and let y_x be the Q-dimensional indicator (0/1) vector of x; component q of y_x indicates whether or not x belongs to class l_q:

$$y_x(q) = \begin{cases} 1 & if\ l_q \in Y, \quad \forall q \in \{1, \cdots, Q\} \\ 0 & otherwise \end{cases} \quad (2)$$

Let c_x denotes the Q-dimensional *membership counting* vector of x. Component q of c_x is the number of cells amongst the *k*NNs of x that belong to class l_q:

$$c_x(q) = \sum_{x_i \in N_x^k} y_{x_i}(q), \quad \forall q \in \{1, \cdots, Q\} \quad (3)$$

3.2. Maximum-A-Posteriori (MAP) Rule

We first identify the set N_x^k of the k nearest neighbours in L of the test cell x and evaluate the counting vector c_x. Denote by E_1^q and E_0^q the respective hypotheses that the cell x belongs and does not belong to class l_q; $E_b^q : y_x$ for $b = 0$ and 1. Denote further by F_j^q, ($j \in \{0, 1, ..., k\}$) the event that exactly j cells in N_x^k belong to class l_q; $F_j^q : c_x(q) = j \in \{0, 1, ..., k\}$. The ML*k*NN method uses the following MAP (Zhang and Zhou, 2007): $\hat{y}_x = argmax\, P\left(E_b^q \mid E_{c_x(q)}^q\right)$, details are given in Omrani et al. (2015a). The DML*k*NN uses the following MAP rule for setting the indicator vector y_x see (Younes et al., 2011):

$$\begin{aligned} \hat{y}_x &= argmax\, P\left(E_b^q \mid \bigcap_{l_j \in \mathcal{L}} F_{c_x(l)}^l \right) \\ &= argmax\, P\left(E_b^q \mid F_{c_x(l)}^q, \bigcap_{l_j \in \mathcal{L}} F_{c_x(l)}^l \right) \end{aligned} \tag{4}$$

That is, the class of y_x is predicted by the componentwise maximisers of the conditional probability of belonging to class l_q, given the number of the nearest neighbours that belong to the class. The assignment of label l_q to the cell x depends conjointly on the event that exactly $c_x(q)$ cells have label l_q in N_x^k, i.e., $F_{C_x(q)}^q$, and that exactly $c_x(l)$ cells have label l_j in N_x^k, for each $l_j \in £\backslash\{l_q\}$. The criterion defined by Equation (4) takes the label correlation into account since all the components of the counting vector c_x are involved in the assignment (or not) of label l_q to the cell x.

Younes et al. (2011) show that estimating the quantities in Equation (4) from the training set L is not accurate since the number of possible events $\bigcap_{l_j \in \mathcal{L}} F_{c_x(l)}^l$ is too large; its upper bound is $(k + 1)^Q$. This problem can be resolved by the following fuzzy approximation. Denote by G_j^l, $j \in \{0, 1, ..., k\}$, the event that the number of cells in N_x^k that belong to class l_j is in the interval $[j - \delta, j + \delta]$, for a suitable $\delta \in \{0, ..., k\}$; δ is the *fuzziness* parameter of the method. This results in a fuzzy MAP relation

$$\hat{y}_x = argmax\, P\left(E_b^q \mid E_{c_x(q)}^q, \bigcap_{l_j \in \mathcal{L}/\{l_q\}} G_{c_x(l)}^l \right) \tag{5}$$

The ML*k*NN method introduced by (Zhang and Zhou, 2007) is a special case of DML*k*NN in which $\delta = k$. In fact, if $\delta = k$, then $\bigcap_{j \in \mathcal{L}/\{l^q\}} G_{c_x(l)}^l$ is the *certain* event. The classification defined by Equation (4) corresponds to $\delta = 0$. When $\delta = k$, Equation (5) can be written using the Bayes rule, the denominator in the first line is equal to 1.0, because it does not involve b and relates to a certain event.

$$\begin{aligned} \hat{y}_x &= argmax \frac{P\left(E_b^q\right) P\left(F_{c_x(q)}^q, \bigcap_{l_j \in \mathcal{L}/\{l_q\}} G_{c_x(l)}^l \mid E_b^q \right)}{P\left(F_{c_x(q)}^q, \bigcap_{l_j \in \mathcal{L}/\{l_q\}} G_{c_x(l)}^l \right)} \\ &= argmax\, P\left(E_b^q\right) P\left(F_{c_x(q)}^q, \bigcap_{l_j \in \mathcal{L}/\{l_q\}} G_{c_x(l)}^l \mid_b^q \right) \end{aligned} \tag{6}$$

The probabilities in Equation (6) can be estimated from the training set. The label for an unseen cell with a covariate vector x is predicted as follows:

$$C(x) = \{l_q \in £ \mid \hat{y}_x(q) = 1\} \tag{7}$$

$C(x)$ contains all labels q for which the probability in Equation (6) is greater for $b = 1$.

3.3. Evaluation

Evaluation of the performance of an algorithm for multi-label learning is more complex than for single-label learning. The prediction for a cell obtained by a multi-label classifier may be correct fully, partly or not at all. We give next the performance measures (metrics) used in this paper. Other multi-label criteria can be found in Boutell et al. (2004) and Tsoumakas and Katakis (2007). Let Y_i be the set of the actual labels of x_i and let $Y_i = C(x_i)$ be the set of predicted labels for cell x_i. Our metrics depend on the following four counts:

$A_i = \left|Y_i \cap \hat{Y}_i\right|$, the number of labels that the actual and predicted classifications have in common;

$B_i = \left|Y_i \cup \hat{Y}_i\right|$, the number of labels that appear in at least one of the two classifications;

$C_i = \left|\hat{Y}_i\right|$ and $D_i = \left|Y_i\right|$, the numbers of labels in the actual and predicted classifications.

The metrics called *accuracy, precision* and *recall* and *F1* are defined as:

$$\begin{aligned} &Acc = \frac{1}{m}\sum_{i=1}^{m}\frac{A_i}{B_i},\ Prec = \frac{1}{m}\sum_{i=1}^{m}\frac{A_i}{C_i}, \\ &Rec = \frac{1}{m}\sum_{i=1}^{m}\frac{A_i}{D_i},\ F1 = \frac{2Prec.Rec}{Prec + Rec} = \frac{1}{m}\sum_{i=1}^{m}\frac{2A_i}{C_i + D_i} \end{aligned} \tag{8}$$

where m is the number of test cells. The measure *F1* is the harmonic average of *Prec* and *Rec* (Yang, 1999). Like *Acc*, it is a symmetric function of the actual and predicted labels.

The Hamming loss (Hamm.loss) is based on the symmetric difference between the observed and predicted labels, Y_i and $\hat{Y}_i$. This metric counts the errors of two kinds, incorrect inclusion and incorrect omission of a label. It is defined as:

$$Hamm.loss = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{Q}\left|\Delta\left(Y_i, \hat{Y}_i\right)\right| \tag{9}$$

where Δ stands for the symmetric difference between two sets and Q is the number of labels.

All five metrics are averages of fractions, and so their values are in the range [0, 1]. Larger values of *Acc*, *Prec*, *Rec* and *F1* correspond to better performance, while smaller value for Hamming loss corresponds to the better performance (Tsoumakas and Katakis, 2007; Yang, 1999).

4. Multi-Label Dataset

In the studied dataset generated for the Grand Duchy of Luxembourg, we have the following elementary labels (classes of land-use): urban, industrial, agriculture, forest, water and road. We assume that the water and road cells are static, subject to no change over time, and retain the other four classes for the model we develop. We define a multi-label classification for each cell in a given year, by an element of the set $\{0, 1\}^4$, indicating for each elementary label whether it is present in the cell or not. For example, the multi-label class (1, 1, 1, 1) indicates that all four elementary labels are present in the cell. We construct the dataset with multi-label classes and use it for classifying land use in Luxembourg (see Figure 3). In a vector map, features are represented by points, lines, and polygons. We rasterize the vector map of Luxembourg (i.e., convert the vector information into a raster format) by using a squared grid with a resolution of 100 × 100 meters (cell size of 1 hectare). This resolution is commonly used at a regional or national scale (White and Engelen, 2000). Such resolution is appropriate to capture variability that can be partly lost at lower resolutions (Himans et al., 2005). The Arc-GIS10 software and Python programming language were used for the rasterzation procedure (Figure 4). We obtain a raster map (i.e., rectangular grid of cells), in which a cell may be associated with more than one elementary label (called multi-label). In addition, we generated several variables, listed in Table 1, for explaining land use; they are based on the literature review (Verburg et al., 2004; Yang et al., 2008; Omrani et al., 2015a).

```
Algorithm 1: Multi-label dataset construction
  Input: map: vector land-use map, landUseClasses: land-use classes
         set, nbClasses: number of land-use classes
  Output: multilabelRaster: multi-label land-use dataset
  begin
    exist ← 1;
    nonExist ← 9;
    for i ← 1 to nbClasses do
      /* extract vector map for each land-use class          */
      extractedMap(i) ← extractMap(map, landUseClasses(i));
      /* convert extracted map into raster dataset           */
      rasterData(i) ← convertToRaster(extractedMap(i));
      /* recode i^th raster dataset so that a cell is coded
         with exist value if it has the i^th land-use and
         nonExist value if not                               */
      recodedRasterData(i) ←
      recodeRaster(rasterData(i), exist, nonExist);
      exist ← exist × 10;
      nonExist ← nonExist × 10;
    end
    /* sum all the raster dataset to obtain the multi-label
       dataset                                               */
    multilabelRaster ← sum(recodedRasterData(1...nbClasses));
    return (multilabelRaster);
  end
```

Figure 3. Multi-label dataset construction.

The following metrics summarizes 'label multiplicity' of a multi-label dataset $S = \{(x_i, Y_i), i = 1, \ldots, n\}$ with $x_i \in X$ and $Y_i \subseteq £$, (Tsoumakas and Katakis, 2007).

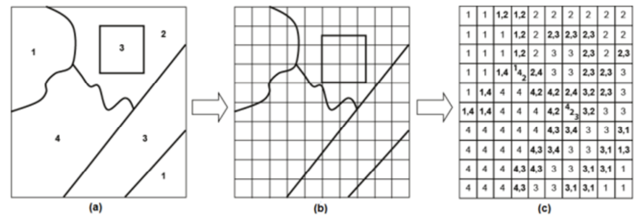

Figure 4. Conversion of a vector map to a raster: (a) vector dataset containing polygons with associated labels; (b) a grid with the desired cell size; (c) the values of the grid cells become the values of the labels of the polygons which contain them.

label cardinality, denoted by *LCard*, is defined as the average number of labels per cell:

$$LCard(S) = \frac{1}{n}\sum_{i=0}^{n} C_i \tag{10}$$

label density, denoted by *LDen(S)*, is defined as the average proportion of the labels in a cell:

$$LDen(S) = \frac{LCard(S)}{Q} \tag{11}$$

DL(S) counts the number of *distinct label* sets present in the dataset *S*:

$$DL(S) = |\{Y_i \subseteq \mathcal{L}; \exists x_i \in X : (x_i, Y_i) \in S\}| \tag{12}$$

If every combination of elementary labels is present in *S*, then $DL(S) = 2^Q$. For a multi-label dataset *S* with *Q* possible labels, the contingency matrix is defined by the elements $CM[q][r] = P(E_1^q \mid E_1^r)$, where q and $r \in \{1, \ldots, Q\}$. It is the proportion of cells in *S* that are assigned label l_q among those that are assigned also to l_r. The off-diagonal elements of *CM* are related to the label dependence, and $CM[q][q] = P(E_1^q)$ is the frequency of label l_q in *S*.

5. Application: Land-Use Change in Luxembourg

5.1. Study Area and Data

Despite its small area (approximately 2,600 square kilometers) and population (around 530,000 habitants), Luxembourg has a strategic geographic location and can be considered as an important contributor to decision making in the European Union. The City of Luxembourg is one of the most attractive metropolitan areas in Europe (Omrani et al., 2010). Its attraction is due to the socioeconomic development of the society and specifically to the strong economic sectors (financial and industrial) that have been developed since the end of 1970's. Surrounded by France, Germany and Belgium, Luxembourg

Table 1. Explanatory Variables

Category	Variable	Description
Physical	State of cell	Multi-label class with different labels
	Slope	Slope value of cell (%)
Spatial	Urban–neighbors	Amount of urban cells in the MN3×3
	Industrial –neighbors	Amount of industrial cells in the MN3×3
	Agriculture–neighbors	Amount of agricultural cells in the MN3×3
	Forest–neighbors	Amount of forest cells in the MN3×3
	Water–neighbors	Amount of water cells in the MN3×3
	Distance–border	Distance from the border of Luxembourg (meters)
Transport	Transport–neighbors	Number of transport cells in the MN3×3
	Distance–bus–station	Distance to the closest bus station (meters)
	Distance–train–station	Distance to the closest train station (meters)
	Distance–highway	Distance to the nearest highway access point (meters)
	Number–bus–station	Number of bus station within the distance of 2km from cell
	Number–train–station	Number of train station within the distance of 2km from cell

* MN 3 × 3 – 3 × 3 Moore neighbourhood comprises the eight cells surrounding a central cell on a two-dimensional square grid.

attracts a lot of well-qualified labour force to maintain its economic dynamism. That explains the importance of residential and daily mobility in Luxembourg and its bordering land (Omrani et al., 2010).

For predicting land-use changes in Luxembourg, we use a biophysical database for our land cover dataset. Features in this dataset are the land cover classes classified into 76 classes (OBS, 1999 and 2007). OBS − Occupation Biophysique du Sol - is also used; it is an inventory of land cover and biotopes in the Grand Duchy of Luxembourg within an inventory of 76 classes at the scale of 1:10.000 based on orthophotos.

These 76 classes distinguish between the variations of six main hierarchical categories: artificial, agriculture, forest and seminatural areas, wetlands and water. In this paper, we did not need to model the land use in that much detail. Thus, we reclassified the 76 into six major land cover classes (based on the 6 existing classes in the OBS): artificial (including urban, industry, transport), agriculture, forest, and water. We distinguish urban, industrial, and transport artificial land cover to study the attraction and repulsion between these three different land uses, and the influence of one on the others. The reclassification was conducted with the assistance of experts that have been working on the same data base for some time. We predict land use in 2007, based on the input variables in 1999. The input and output variables are described in Table 1. All these inputs are in line with what is used in the literature for prediction of land use (Yang et al., 2008). The ArcGIS software and Matlab tool were used for spatial data processing and modelling, respecttively.

5.2. Main Results

Table 2 shows a summary of land-use multi-label data in the observed periods (1999 and 2007). The dataset comprises 255,698 cells. Each cell is associated with a 13-dimensional feature vector. The input variables are classified as physical, spatial, and transport related, as shown in Table 1. The state of a cell is its multi-label classification, with labels from the set comprising agriculture (A), forest (F), industrial (I) and urban (U). In the dataset, the label cardinality (the average number of labels for the cells) and label density are respectively equal to 1.54 and 1.54/4 = 0.38. Table 2 displays the binary coding of the multi-label classes with the corresponding number of cells in the dataset. As shown in the table, 1399 cells in 2007 have all four elementary labels. The mono-labels F and A represent together 50.97% (22.41% + 28.56%) of the studied area. The mono-label U accounts for 1.48% and I for 0.50% of the land. However, 46.98% of cells have multiple labels.

Figure 5 displays the contingency matrix CM for the multi-label land use dataset. It confirms that prediction of land use is a multi-label issue. The matrix is asymmetric, CM[A][U] ≠ M[U][A], because its elements are conditional probabilities. In fact, CM[A][U]/CM[U][A] = P(A)/P(U). In urban areas, forest is rare, but agriculture is more common than what chance would imply; CM[F][U] = 0.098, so labels F and U are unlikely to occur together. In contrast, CM[A][U] = 0.63; label A is frequently present with label U.

5.2.1. Model Calibration and Validation

We split the dataset S to the learning (L) and the testing (T) subsets by stratified random sampling with simple random sampling within each class of land use. It assures that all classes are well represented in L. Calibration of the model is based on L and it is validated on T. The DMLkNN method has two parameters, the number of neighbours used, k, and and fuzziness parameter, δ; see Section 3. It is essential to select the optimal values for k and δ. For this purpose, we use cross-validation with 100 replications. We explore the (integer) values of the parameters k and $\delta \leq k$, 1, 2, ..., 10.

Figure 6 displays the values of Hamming loss as function of δ, with k set to 10. It shows that the best results were obtained for values of δ around 5. The setting of k is explored similarly, and we concluded that $k = 10$ is optimal. All the res-

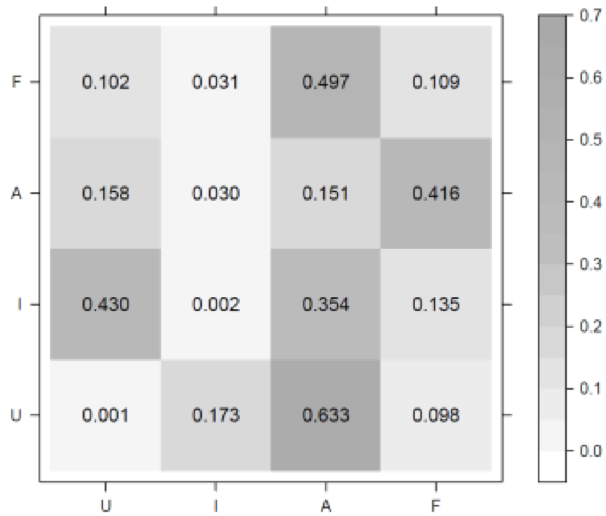

Figure 5. Contingency matrix for the multi-label land-use dataset.

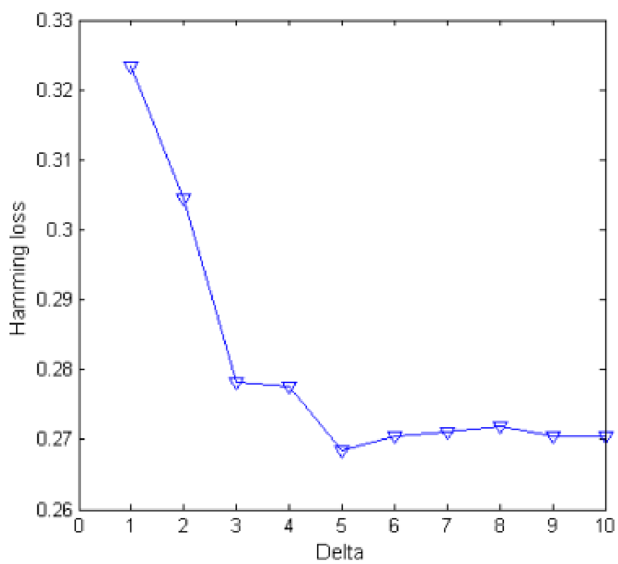

Figure 6. Values of Hamming loss for DMLkNN as a function of δ, where $k = 10$.

ults reported below relate to $k = 10$ and $\delta = 5$.

5.2.2. Results and Analysis

Figure 7 displays the transition matrix between 1999 and 2007. It shows that some industrial land has been converted to urban (example of new Esch-Belval in the south part of Luxembourg) and some forest or agriculture. Forest can be converted mainly to industrial land and agriculture to industrial, forest and urban lands. The transition matrix indicates that certain transitions tend not to occur. For example, urban land is not converted to agriculture or forest. The light-coloured squares

Table 2. Summary of Land-Use Multi-Label Data

Label set	Class				Observed label set			
					1999		2007	
	U	I	A	F	#	%	#	%
F	0	0	0	1	56,730	22.19	57,297	22.41
A	0	0	1	0	75,320	29.46	73,025	28.56
AF	0	0	1	1	77,330	30.24	76,777	30.03
I	0	1	0	0	1,305	0.51	1,282	0.50
FI	0	1	0	1	1,961	0.77	2,141	0.84
AI	0	1	1	0	2,558	1.00	3,001	1.17
AFI	0	1	1	1	2,277	0.89	2,609	1.02
U	1	0	0	0	3,933	1.54	3,796	1.48
FU	1	0	0	1	3,583	1.40	3,862	1.51
AU	1	0	1	0	17,024	6.66	15,565	6.09
AUF	1	0	1	1	9,342	3.65	9,375	3.67
IU	1	1	0	0	973	0.38	1,426	0.56
FIU	1	1	0	1	795	0.31	1,046	0.41
AIU	1	1	1	0	1,517	0.59	2,943	1.15
AFIU	1	1	1	1	892	0.35	1,399	0.55

* The output classes are agriculture (A), forest (F), industrial (I) and urban (U).
** The symbols # and % denote the number and percentage of cells. Classes are coded as 0 and 1.
*** The value 1 (0) means the label is present (absent).

above the diagonal in Figure 7 have all additions of label I.

From the numerical version of Figure 7, we compute the percentages of cells that have no change, small change and large change. The results are displayed in Table 3. Small change is defined as adding one or removing one elementary label (e.g., AF→A, A→AI), but not both, and all other changes are called large (e.g., A→F, F→A, AFI→F). The percentage of cells with no change (the same multi-label), small change and large change between the observed periods, 1999 and 2007, are equal to 90.89, 8.58 and 0.65%, respectively.

The results of predicting land use by the DMLkNN method are shown in Table 3. This table presents a summary of observed and predicted labelsets between 1999 and 2007. Figure 8 shows a map of the Hamming loss.

The map of Hamming loss shows the disagreement between observed (2007) and predicted (2007) cells labels. According to the results, the Hamming loss by cell is generally low and equal to (0, 1/4, 2/4, 3/4, 1) respectively with the following percentages (86, 12.663, 1.259, 0.076, and 0.002%). Overall, observed-predicted differences (2007 vs. 2007) are small, which highlight the model performance. However, in some areas the Hamming loss is a little bit high in some highly heterogeneous and dense urban areas, mostly in the city of Luxembourg, Eschsur-Alzette (the second municipality) and Differdange (the third municipality of the country), where a lot of constructions and other development took place in the 1999−2007 period.

A more detailed summary of predictions is obtained by the 15 × 15 confusion matrix for multi-label classification sh-

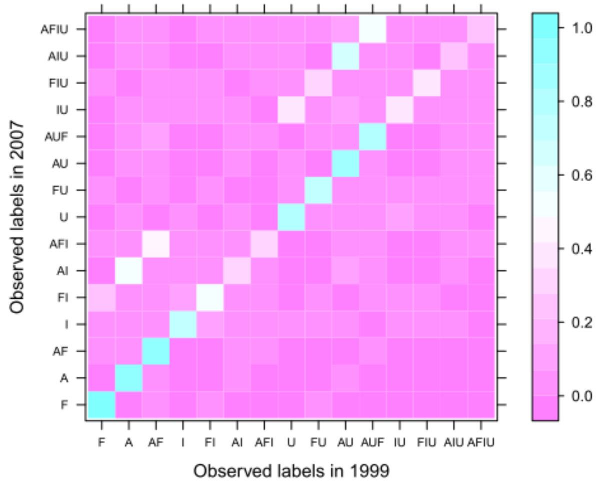

Figure 7. Transition matrix: percentage of cells having label class (*i*) in 1999 and label class (*j*) in 2007.

Table 3. Summary of Changes Between 1999 and 2007

Change	Observed 1999 – Observed 2007	Observed 1999 – Predicted 2007
None	90.89%	90.83%
Small	8.58%	8.55%
Large	0.65%	0.62%

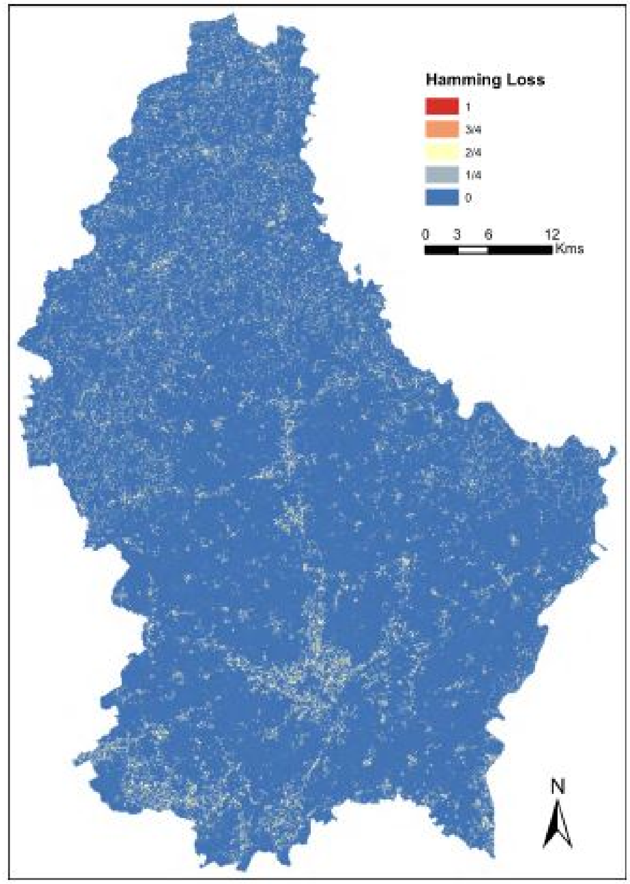

Figure 8. Map of Hamming loss.

own in Figure 9. The rows and columns of this matrix are the multi-labels (F, A, AF, ..., AFIU), and its entries are the percentages of cells. The diagonal of this matrix has very large entries because changes in land use over the eight years are not frequent. In Figure 9, we added a small dot in every square (element of the matrix) that represents a small change, and a (larger) circle in every square that is for a large change.

The results of land use predictions are summarized by the probability maps in Figure 10 for the four elementary considered labels. Probability map displays for each cell the probability that it belongs to a given class (Shahumyan et al., 2011; White et al., 2012; Tayyebi et al., 2015). The four panels show that prediction for most cells is easy; they are either very likely or very unlikely to have a given label.

Only a few small areas are difficult to predict. They correspond to probability values between 0.4 and 0.6. The general trend of spread of industrial areas is mainly in areas of low slope (flat landscape) and close to road and rail network. Similarly, the spread of the urban areas is significant around infrastructure (roads, railway and bus lines). The class of industry is the most difficult to predict. This can be explained by the dispersed nature of industrial class and the data structure. Indeed, the industry class plays an active role (the same as the urban class) with other classes of land use. Both urban and industrial classes are artificial areas, often connected to form contiguo-

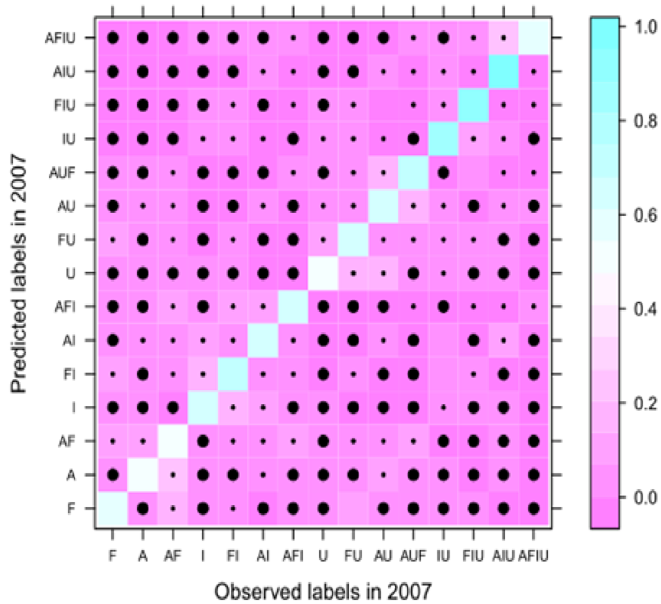

Figure 9. Confusion matrix for multi-label classification. Squares for small changes are marked by a black dot and squares for large changes by filled circles.

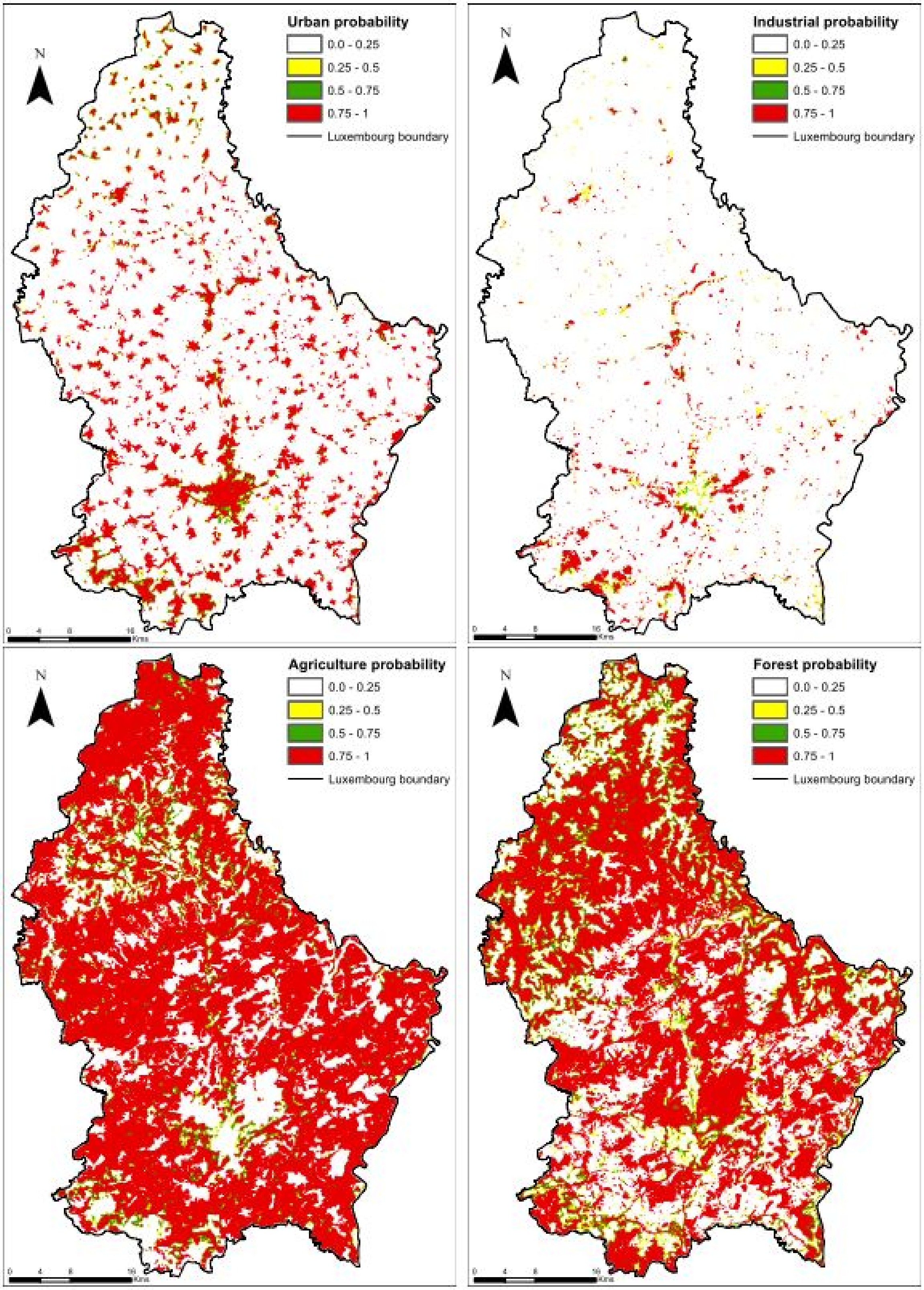

Figure 10. Probability maps.

us, discrete or mixed land use, which complicates spatial modelling. This problem may be solved by incorporating further variables related to socio-economics, population and employment.

In Table 4, we present performance measures from the application of DML*k*NN and ML*k*NN methods on the testing dataset. Note that the ML*k*NN method is a special case of the DML*k*NN, in which dependence among the labels is ignored. Based on the results, the DML*k*NN outperforms the ML*k*NN method in terms of all five measures, but only on average. All the entries in the table are close to the ideal (1.0 or 0.0), because the classification has changed only for a small fraction of the cells. Arguably, the distance from 1.0 should be considered for the first four metrics, such as 0.0095 vs. 0.0125 for 1-*Accuracy*. The differences of the measures are not substantially greater than the standard deviations. The standard errors of the percentages are ten times smaller than the standard deviations. Notable is the substantially greater standard deviation of *Accuracy*, *F1* and *Hamming loss* for DML*k*NN than for ML*k*NN. These results clearly demonstrate the advantages of the DML*k*NN method in taking into account the intuitive hypothesis of label dependence and correlation in land use.

Table 4. Results (mean ± std) on the Land-Use Dataset by DMLkNN and MLkNN Methods (k = 10 and δ = 5) Based on 100 Replications

	DMLkNN	MLkNN
Accuracy+	0.9905±0.0151	0.9875±0.001
Precision+	0.9234±0.0170	0.8977±0.0164
Recall+	0.9105±0.0152	0.8935±0.0219
F1+	0.9166±0.0163	0.8952±0.0033
Hamming loss-	0.0584±0.0034	0.0795±0.0014

* +(-): the higher (smaller) the value, the better the performance.

6. Conclusions, Discussion and Future Work

In this paper, we proposed multi-label learning for land-use prediction, by considering dependence between labels, based on the DML*k*NN method. We illustrated and validated its use by modelling land-use changes in the Grand Duchy of Luxembourg. The performance of the DML*k*NN method is better than ML*k*NN using the established evaluation criteria of precision, recall and F-measure adapted to multi-label learning. Overall, the obtained probability maps show excellent agreement with the observed data.

In land change science, we cannot deny uncertainty or errors are not linked with land use class assignment (Pontius, 2000). First, the error might come from the interpretation of aerial photos (remote sensing field). This type of error might influence inputs and outputs of land use models. Second, errors could probably come from the scale of rasterization procedure from vector map to raster data. Third, since we use spatial data that contain errors to model land use change, errors can propagate during modeling as well (Brown et al., 2013). In the future, we can minimize the effect of these errors by taking the dominance of labels within each cell in order to gain more details and have a richer representation of reality. Furthermore, other multi-label learning methods could also be tested in order to overcome drawbacks of the nearest neighbor-based methods. In addition, handling the land use problem as a class imbalance one is not treated in this paper. At our best understanding, this class imbalance problem, even an important issue, is not yet studied in land use change science. These issues of (1) the class imbalance problem and (2) explicit understanding of sources of errors need to be explored more fully in future work.

Our next step is to investigate the problem of semi-supervised multi-label learning to manipulate both labeled and unlabeled cells at the same time. The problem of unsupervised multi-label learning is also important for handling totally unlabeled data including the special case when we have no prior knowledge about the target classes. Another interesting challenge is to combine several multi-label learning methods to enhance model performance. It has been shown in conventional mono-label learning that combined methods result in better generalization and higher accuracy than a single method. The combination of several methods is obtained by combining the posterior probabilities from each method using a specific rule, such as minimum, maximum, product, sum or majority voting. Future work can exploit the developments made in the DML*k*NN method to extend other methods for multi-label learning, such as those based on decision trees and neural networks. We can use specifically the artificial neural network (Valipour et al., 2013) with its extension to multi-label learning for land use change prediction. In particular, we are planning to integrate the new multi-label concept with dependence among labels into the well-known ANN-based land use model called Land Transformation Model-LTM (Pijanowski et al., 2002, 2014; Tayyebi et al., 2014). This future work would present to the community of geo-simulation a refinement of the LTM model that allows the storage of multi-class occupancy. Of course, other land-use models could benefit from including a multi-label learning framework if modelers desire to avoid over-simplification of land use classes in area assignment.

Acknowledgments. Funding to complete this research was obtained through the National Research Fund of Luxembourg (FNR, Smart-Boundary project, INTER/CNRS/12/02) and the Luxembourg Institute of Socio-Economic Research-LISER, Luxembourg.

References

Agarwal, C., Green, G.M., Grove, J.M., Evans, T.P., and Schweik, C. M. (2002). A review and assessment of land-use change models: dynamics of space, time, and human choice. *Gen. Tech. Rep. USDA For. Serv.,* 12-27.

Basse, R.M., Omrani, H., Charif, O., Gerber, P., and Bódis, K. (2014). Land use changes modelling using advanced methods: Cellular automata and artificial neural networks. The spatial and explicit representation of land cover dynamics at the cross-border region scale. *Appl. Geogr.,* 53, 160-171. https://doi.org/10.1016/j.apgeog.2014.06.016

Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999). When is "nearest neighbor" meaningful? *Proc. of the International Conference on Database Theory*, 217-235. Springer. https://doi.org/10.1007/3-540-49257-7_15

Boutell, M.R., Luo, J., Shen, X., and Brown, C.M. (2004). Learning multi-label scene classification. *Pattern Recogn.,* 37(9), 1757-1771. https://doi.org/10.1016/j.patcog.2004.03.009

Briassoulis, H. (2000). Analysis of land use change: theoretical and modeling approaches. *Gen. Inf.,* 485-514.

Brown, D.G., Band, L.E., Green, K.O., Irwin, E.G., Jain, A., Lambin, E.F., Jr, R.G.P., Seto, K.C., Turner Ii, B.L, and Verburg P.H. (2013). *Advancing Land Change Modeling: Opportunities and Research Requirements*, The National Academies Press, Washington DC.

DeFries, R., and Eshleman, K.N. (2004). Land-use change and hydrologic processes: a major focus for the future. *Hydrol. Process.,* 18(11), 2183-2186. https://doi.org/10.1002/hyp.5584

Ericksen, P.J., Ingram, J.S., and Liverman, D.M. (2009). Food security and global environmental change: emerging challenges. *Environ. Sci. & Policy,* 12(4), 373-377. https://doi.org/10.1016/j.envsci.2009.04.007

Fan, R.E., and Lin, C.J. (2007). *A Study on Threshold Selection for Multi-label Classification,* Department of Computer Science, National Taiwan University, pp. 1-23.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., and Carpenter, S.R., et al. (2005). Global consequences of land use. *Science,* 309(5734), 570-574.

Herold, M., Couclelis, H., and Clarke, K.C. (2005). The role of spatial metrics in the analysis and modeling of urban land use change. *Comput., Environ. Urban Syst.,* 29(4), 369-399. https://doi.org/10.1016/j.compenvurbsys.2003.12.001

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A., (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.,* 25(15), 1965-1978. https://doi.org/10.1002/joc.1276

Lee, D.B. (1994). Retrospective on large-scale urban models. *J. Am. Plann. Assoc.,* 60(1), 35-40. https://doi.org/10.1080/01944369408975549

Lee, D.B. (1973). Requiem for large-scale models. *J. Am. Inst. Plann.,* 39(3), 163-178. https://doi.org/10.1080/01944367308977851

Li, X., and Yeh, A.G. (2004). Data mining of cellular automata's transition rules. *Int. J. Geogr. Inf. Sci.,* 18, 723-744. https://doi.org/10.1080/13658810410001705325

OBS (1999, 2007). Land cove map at the scale of 1:10 000, Owned by ministry of environment Luxembourg.

Omrani, H., Charif, O., Klein, O., Gerber, P., and Trigano, P. (2010). An approach for spatial and temporal data analysis: Application for mobility modeling of workers in Luxembourg and its bordering areas. *Proc. of the IEEE International Conference on Systems Man and Cybernetics (SMC)*, 1437-1442. https://doi.org/10.1109/icsmc.2010.5642474

Omrani, H., Abdallah, F., Charif, O., and Longford, N.T. (2015a). Multi-label class assignment in land-use modelling. *Int. J. Geogr. Inf. Sci.,* 29(6), 1023-1041. https://doi.org/10.1080/13658816.2015.1008004

Omrani, H., Tayyebi, A., and Pijanowski, B. (2015b). Integrating the multi-label land use concept and cellular automata with the ANN-based Land Transformation Model. *Proc. of the Geocomputing conference*, Dallas, Texas, USA, pp. 20-23.

Omrani, H., Tayyebi, A., and Pijanowski, B. (2017). Integrating the multi-label land-use concept and cellular automata with the artificial neural network-based Land Transformation Model: An integrated ML-CA-LTM modeling framework. *Gisci. Remote Sens.*, 54 (3), 283-304. https://dx.doi.org/10.1080/15481603.2016.1265706

Pijanowski, B., Brown, D.G. Shellitoc, B.A. and Manikd, G.A. (2002). Using neural networks and GIS to forecast land use changes: a land transformation model. *Comput., Environ. Urban Syst.,* 26, 553-575. https://doi.org/10.1016/S0198-9715(01)00015-1

Pijanowski, B.C., Tayyebi, A., Delavar, M.R., and Yazdanpanah, M. J. (2010). Urban expansion simulation using geographic information systems and artificial neural networks. *Int. J. Environ. Res.,* 3(4), 493-502.

Pijanowski, B.C., Tayyebia, A., Doucettea, J., Pekina, B.K., Braund, David., and Plourdea, J. (2014). A big data urban growth simulation at a national scale: configuring the GIS and neural network based Land Transformation Model to run in a High Performance Computing (HPC) environment. *Environ. Model. Softw.,* 51, 250-268. https://doi.org/10.1016/j.envsoft.2013.09.015

Platt, R.H. (2004). Geography, law, and public policy, *Land Use and Society (Revised Edition*), Island Press.

Pontius, R.G. (2000). Quantification error versus location error in comparison of categorical maps. *Photogramm. Eng. Remote Sensing,* 66(8), 1011-1016.

Reidsma, P., Tekelenburg, T., Van den Berg, M., and Alkemade, R. (2006). Impacts of land-use change on biodiversity: an assessment of agricultural biodiversity in the European Union. *Agric., Ecosyst. Environ.,* 114(1), 86-102. https://doi.org/10.1016/j.agee.2005.11.026

Shahumyan, H., White, R., Petrov, L., Williams, B., Convery, S., and Brennan, M. (2011). Urban development scenarios and probability mapping for greater dublin region: the moland model applications. *Proc. of the International Conference on Computational Science and Its Applications*, 119-134. Springer. https://doi.org/10.1007/978-3-642-21928-39

Schneider, L.C., and Pontius, R.G. (2001). Modeling land-use change in the Ipswich watershed, Massachusetts, USA. *Agric., Ecosyst. Environ.,* 85(1-3), 83-94. https://doi.org/10.1016/S0167-8809(01)00189-X

Song, W., Pijanowski, B.C., and Tayyebi, A. (2015). Urban expansion and its consumption of high-quality farmland in Beijing, China. *Ecol. Indicators,* 54, 60-70. https://doi.org/10.1016/j.ecolind.2015.02.015

Spyromitros, E., Tsoumakas, G., and Vlahavas, I. (2008). An empirical study of lazy multilabel classification algorithms, *Artificial Intelligence: Theories, Models and Applications*, Springer, pp. 401-406. https://doi.org/10.1007/978-3-540-87881-0_40

Tang, Z., Engel, B.A., Pijanowski, B.C., and Lim, K.J. (2005). Forecasting land use change and its environmental impact at a watershed scale. *J. Environ. Manage.,* 76(1), 35-45. https://doi.org/10.1016/j.jenvman.2005.01.006

Tayyebi, A., and Pijanowski, B.C. (2014). Modeling multiple land use changes using ANN, CART and MARS: Comparing tradeoffs in goodness of fit and explanatory power of data mining tools. *Int. J. Appl. Earth Obs. Geoinf.,* 28, 102-116. https://doi.org/10.1016/j.jag.2013.11.008

Tayyebi, A.H., Tayyebi, A., and Khanna, N. (2014). Assessing uncertainty dimensions in land-use change models: using swap and multiplicative error models for injecting attribute and positional errors in spatial data. *Int. J. Remote Sens.,* 35(1), 149-170. https://doi.org/10.1080/01431161.2013.866293

Tayyebi, A., Pijanowski, B.C., and Pekin, B.K. (2015). Land use legacies of the Ohio River Basin: Using a spatially explicit land use change model to assess past and future impacts on aquatic resources. *Appl. Geogr.,* 57, 100-111. https://doi.org/10.1016/j.apgeog.2014.12.020

Tayyebi, A., Meehan, T.D., Dischler, J., Radloff, G., Ferris, M., and Gratton, C. (2016a). SmartScape™: A web-based decision support system for assessing the tradeoffs among multiple ecosystem services under crop-change scenarios. *Comput. Electron. Agric.,* 121, 108-121. https://doi.org/10.1016/j.compag.2015.12.003

Tayyebi, A., Tayyebi, A., Vaz, E., Arsanjani, J.J., and Helbich, M. (2016b). Analyzing crop change scenario with the SmartScape™ spatial decision support system. *Land Use Policy,* 51, 41-53. https://doi.org/10.1016/j.landusepol.2015.11.002

Tong, S.T.Y., and Chen, W. (2002). Modeling the relationship between land use and surface water quality. *J. Environ. Manage.,* 66(4), 377-393. https://doi.org/10.1006/jema.2002.0593

Tsoumakas, G., and Katakis, I. (2007). Multi-label classification: An overview. *Int. J. Data Warehous. Min (IJDWM)*, 3, 1-13. https://doi.org/10.4018/jdwm.2007070101

Turner, B.L., Skole, D.L., Sanderson, S., Fischer, G., Fresco, L., and Leemans, R. (1995). *Land-Use and Land-Cover Change. Science/Research Plan*, Global Change Report (Sweden).

Valipour, M., Banihabib, M.E., and Behbahani, S.M.R. (2013). Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *J. Hydrol.,* 476, 433-441. https://doi.org/10.1016/j.jhydrol.2012.11.017

Veldkamp, A., and Lambin, E.F. (2001). Predicting land-use change. *Agric., Ecosyst. Environ.,* 85(1-3), 1-6. https://doi.org/10.1016/S0167-8809(01)00199-2

Verburg, P.H., Kok, K., Pontius Jr,R.G., and Veldkamp, A. (2006). Modeling land-use and land-cover change, *Land-use and Land-cover Change*, Springer Berlin Heidelberg, pp. 117-135. https://doi.org/10.1007/3-540-32202-7_5

Verburg, P.H., Schot, P.P., Dijst, M.J., and Veldkamp, A. (2004). Land use change modelling: current practice and research priorities. *GeoJournal,* 61(4), 309-324. https://doi.org/10.1007/s10708-004-4946-y

Watson, R.T., Noble, I.R., Bolin, B., Ravindranath, N.H., Verardo, D. J., and Dokken, D.J. (2000). *Land Use, Land-Use Change and Forestry: A special Report of the Intergovernmental Panel on Climate Change,* Cambridge University Press.

White, R. and Engelen, G. (2000). High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Comput., Environ. Urban Syst.,* 24(5), 383-400. https://doi.org/10.1016/S0198-9715(00)00012-0

White, R., Uljee, I., and Engelen, G. (2012). Integrated modelling of population, employment and land-use change with a multiple activity-based variable grid cellular automaton. *Int. J. Geogr. Inf. Sci.,* 26(7), 1251-1280. https://doi.org/10.1080/13658816.2011.635146

Yang, Q., Li, X., and Shi, X. (2008). Cellular automata for simulating land use changes based on support vector machines. *Comput. Geosci.,* 34(6), 592-602. https://doi.org/10.1016/j.cageo.2007.08.003

Yang, Q., Shao, J., Scholz, M., Boehm, C., and Plant, C. (2012). Multilabel classification models for sustainable flood retention basins. *Environ. Model. Software,* 32, 27-36. https://doi.org/10.1016/j.envsoft.2012.01.001

Yang, Y. (1999). An evaluation of statistical approaches to text categorization. *Inf. Retrieval,* 1(1), 69-90. https://doi.org/10.1023/A:1009982220290

Younes, Z., Abdallah, F., Denoeux, T., and Snoussi, H. (2011). A dependent multilabel classification method derived from the k-nearest neighbor rule. *EURASIP J. Adv. Sig. Process.,* 1-14. https://doi.org/10.1155/2011/645964

Yu, J., Wong, W.K., Dietterich, T., Jones, J., Betts, M., Fery, S., Miller, J., and White, M. (2011). Multi-label classification for multi-species distribution modeling. *Proc. of the 28th International Conference on Machine Learning*, pp.1-4.

Zhang, M.L., and Zhou, Z.H. (2007). Ml-knn: A lazy learning approach to multi-label learning. *Pattern Recogn.,* 40(7), 2038-2048. https://doi.org/10.1016/j.patcog.2006.12.019