3
A novel multitemporal cascade-classification method

The multitemporal classification method proposed in this work fits into the category of the semantic approaches mentioned in Section 2.2.2, more specifically it can be regarded as a f*uzzy multitemporal cascade-classification approach*, formally structured through the concept of *fuzzy Markov chain* (Avrachenkov and Sanchez, 2002).

Fuzzy reasoning aims at the modeling of reasoning schemes based on uncertain or imprecise information, and, as stated before, imprecision is an intrinsic component in the description of the mental process humans carry out when interpreting images. Fuzziness is, as explained by Bellman and Zadeh (1970) in a seminal paper, a major source of imprecision in many decision processes. They also pointed out that even when fuzziness can be simulated by a probabilistic model, it is generally beneficial to deal with it through fuzzy logic concepts.

Additionally, regarding the adopted technique for the implementation of the multitemporal classification method presented in this chapter, it has been showed (Kruse et al., 1987) that the fuzzy Markov chain is a robust system with respect to small perturbations of the transition matrix, which is not the case for the classical, probabilistic Markov chain.

As a cascade approach, classification results in terms of class membership values for the same geographical object at two dates are combined through the devised multitemporal method into a single consensus result. Moreover, the method can be set for the classification of both images simultaneously, or for the classification of only one of the two images. If simultaneous classification is intended, training samples for which the true classes are known at both epochs are required, if the goal is the classification of a single image, only training samples from the target date are needed to train the model.

It is important to observe that what has been said in the last paragraph on the demand for training samples does not take into consideration the particular

necessities of the classifiers employed at either epoch. Those classifiers are treated as black-boxes by the proposed method, which requires only that they provide a vector of class membership values – a *fuzzy label vector* – for each image object (that could in particular cases be an unitary vector). Additionally, the proposed method does not require that the same classifier be used for both images.

Before aggregating the monotemporal results, the fuzzy classifications associated to one of the epochs undergo a temporal transformation that projects them onto the other epoch. A fuzzy Markov chain is used to model the temporal transformation, which relies on a fuzzy transition matrix.

Additionally, an analytic technique to estimate class transition possibilities is introduced as an alternative for a stochastic approach, based on Genetic algorithms (Davis, 1990).

The proposed method can be applied to pixel-wise or segment-wise classification (Blaschke and Strobl, 2001), hence, the text that follows uses the term image objects to denote either the pixels or the segments being classified.

The next section presents the fundamentals of fuzzy Markov chains. The proposed multitemporal classification model is described in detail in the subsequent sections.

3.1. Fuzzy Markov chains

This section introduces the concept of fuzzy Markov chain (FMC). The description that follows is oriented to the use of FMC as a tool to model class dynamics, i.e. how an image object changes from one class to another through time, in the context of multitemporal classification of remotely sensed images. A more general presentation about the concept may be found in (Avrachenkov and Sanchez, 2002).

In this work, images are associated to their acquisition dates: $t_0+t\Delta t$, where t_0 is some stipulated initial time, Δt is a given time interval, and t is any integer number. For simplicity, the date $t_0+t\Delta t$ will be referred from this point on as time t, and $t_0+(t+1)\Delta t$ as time $t+1$, for $t \in Z$.

Let $\mathbf{\Omega} = \{\omega_1, \omega_2, \ldots, \omega_n\}$ be the set of n distinguishable land-use/land-cover (LU/LC) classes in the time span being considered. A binary fuzzy relation

can be defined on the Cartesian product $\mathbf{\Omega}\times\mathbf{\Omega}$ represented by a $n\times n$ transition matrix $\mathbf{T} = \{\tau_{ij}\}$. The symbol τ_{ij} stands for the possibility that an image object belongs to the class $\omega_i \in \mathbf{\Omega}$ at time t and to the class $\omega_j \in \mathbf{\Omega}$ at time $t+1$, with $0 \leq \tau_{ij} \leq 1$, for $i,j = 1,\ldots,n$.

This can be pictorially described by a class transition diagram (Figure 10), a weighted directed graph whose nodes correspond to classes and links to plausible class transitions between t and $t+1$. Each link is labeled with the class transition possibility τ_{ij}. For simplicity links with $\tau_{ij} = 0$ are not drawn.

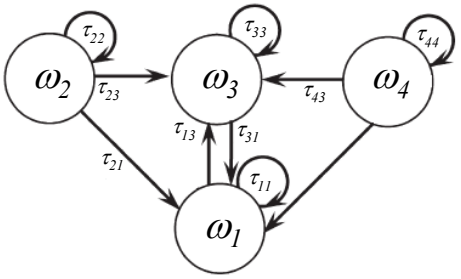

Figure 10. Class transition diagram.

The vector ${}^t\boldsymbol{\alpha} = [{}^t\alpha_1,\ldots,{}^t\alpha_n]$ with $0 \leq {}^t\alpha_i \leq 1$ represents, for a particular image object, the fuzzy classification defined over $\mathbf{\Omega}$ at time t, where ${}^t\alpha_i$ denotes the membership value of the object to class ω_i (for all $\omega_i \in \mathbf{\Omega}$) at time t. It is further assumed that ${}^t\alpha_i$ is a function of attribute values of the image object at time t.

Based on the fuzzy label vector ${}^t\boldsymbol{\alpha}$ and on the transition matrix $\mathbf{T}$, the fuzzy Markov chain model estimates the class membership values, represented by the vector ${}^{t+1}\boldsymbol{\beta} = [{}^{t+1}\beta_1,\ldots,{}^{t+1}\beta_n]$, for the same object one time unit later by applying the following formula:

$$ {}^{t+1}\beta_j = \underset{i=1,\ldots,n}{\perp}\left\{\top\left({}^t\alpha_i, \tau_{ij}\right)\right\} \text{ for } i,j = 1,\ldots,n \qquad (2) $$

The symbols $\top$ and $\perp$ represent respectively a *t-norm* and a *s-norm*. These are basic operations that correspond, respectively, to the intersection and to the union of fuzzy sets. A function qualifies as a *t-norm* or as an *s-norm* by satisfying a group of conditions well established in fuzzy set theory. A thorough

presentation on this topic, including a precise definition of these concepts, may be found in (Klir and Yuan, 1995).

The transition law introduced in Equation (2) can be expressed in a more compact form by the equation below.

$$^{t+1}\boldsymbol{\beta} = {}^{t}\boldsymbol{\alpha} \circ \mathbf{T} \tag{3}$$

In the expression "$\mathbf{A}\circ\mathbf{B}$", the symbol "$\circ$" denotes a special type of matrix multiplication, analogous to conventional matrix multiplication, where the product is replaced by a t-norm and the summation is replaced by a s-norm operator.

At this point, it is convenient to clarify the adopted notation. The symbols $^{t+1}\boldsymbol{\beta}$ and $^{t+1}\boldsymbol{\alpha}$ denote different distributions, although both refer to the same object at the date *t+1*. While $^{t+1}\boldsymbol{\alpha}$ has been computed based on feature values at time *t+1* without any temporal transition transformation, $^{t+1}\boldsymbol{\beta}$ is the result of applying the FMC transition law, Equation (3), on the membership grades in $^{t}\boldsymbol{\alpha}$ computed on the feature values of the image object corresponding to the same geographical object at time *t*.

3.2. Problem statement

Let $^{t}\mathbf{I}$ and $^{t+1}\mathbf{I}$ denote two co-registered images of the same geographical area acquired respectively at dates *t* and *t+1*. Accordingly, $^{t}\mathbf{x}$ and $^{t+1}\mathbf{x}$ stand for the feature vectors composed of spectral and spatial feature values describing the same geographical object respectively in $^{t}\mathbf{I}$ and $^{t+1}\mathbf{I}$. Let us further denote by $^{t}\mathbf{w}$ and $^{t+1}\mathbf{w}$ the *crisp label vectors* for the object being analyzed at times *t* and *t+1*. Both $^{t}\mathbf{w}$ and $^{t+1}\mathbf{w}$ are *n*-dimensional unitary vectors of the form [0 … 1 … 0] having "1" in its *i*-th component and "0" otherwise, indicating that the object belongs to the class ω_i at a particular time. Formally, $^{t}\mathbf{w}$ and $^{t+1}\mathbf{w}$ belong to a *n*-dimensional space $\boldsymbol{\Omega}^n$, where:

$$\boldsymbol{\Omega}^n = \{\, \mathbf{w} = [\, w_1, \ldots, w_n] \mid w_i \in [0,1] \text{ for all } i=1,\ldots,n \text{ and } \|\mathbf{w}\|=1\}^5. \tag{4}$$

[5] $\|\cdot\|$=denotes the vector module

If one is concerned only with the classification of the image objects in the later date, the multitemporal classification problem can be defined as the problem of generating the vector ${}^{t+1}\mathbf{w}$ for each image object, based on the feature vectors ${}^{t}\mathbf{x}$ and ${}^{t+1}\mathbf{x}$, in other words, it is about finding a function **M** of the form:

$$ {}^{t+1}\mathbf{w} = \mathbf{M}\,({}^{t}\mathbf{x}, {}^{t+1}\mathbf{x}) \tag{5} $$

It is interesting to consider that a method that classifies a later image exploring the information of an earlier image should work equally well if the image being classified is the earlier one using data from a later date. In principle, by just inverting the order of the model's arguments (the feature vectors from the two points in time), the objects being classified would relate to the earlier date, whereby their features at a later date are used as additional information. With this in mind, the multitemporal classification problem can be formulated in more general terms as consisting of the determination of the crisp label vectors ${}^{t}\mathbf{w}$ and ${}^{t+1}\mathbf{w}$ based on ${}^{t}\mathbf{x}$ and ${}^{t+1}\mathbf{x}$. Formally it is about designing a function M of the form:

$$ ({}^{t}\mathbf{w}, {}^{t+1}\mathbf{w}) = \mathbf{M}\,({}^{t}\mathbf{x}, {}^{t+1}\mathbf{x}) \tag{6} $$

Clearly the above formulation subsumes that of Equation (5).

3.3. General classification model

In this section a particular design for the classification model stated above is presented. The terms *monotemporal* and *multitemporal* will be used hereafter to designate classifiers whose inputs refer respectively to a single date or to multiple dates.

The multitemporal classifier proposed here can be viewed as the combination of two monotemporal classifiers. Let the first monotemporal classifier be represented by a function ${}^{L}\mathbf{C}$ that computes membership values for the object being classified at the later time $t+1$, based exclusively on the feature values at time $t+1$, extracted from image ${}^{t+1}\mathbf{I}$. The monotemporal classifier ${}^{L}\mathbf{C}$ produces a n-dimensional fuzzy label vector represented by ${}^{t+1}\boldsymbol{\alpha} = [{}^{t+1}\alpha_1, {}^{t+1}\alpha_2, \ldots, {}^{t+1}\alpha_n]$, where ${}^{t+1}\alpha_i$ stands for the membership of the image

object to class ω_i, for all $\omega_i \in \mathbf{\Omega}$ and for at least one i, ${}^{t+1}\alpha_i \neq 0$. So, ${}^{L}\mathbf{C}$ can be viewed as a function of the form:

$$ {}^{t+1}\boldsymbol{\alpha} = {}^{L}\mathbf{C}\ ({}^{t+1}\mathbf{x}) \tag{7} $$

A second monotemporal classifier ${}^{E}\mathbf{C}$ is applied to the object feature vector ${}^{t}\mathbf{x}$ at time t. Analogously to the first monotemporal classifier, it generates a fuzzy label vector ${}^{t}\boldsymbol{\alpha}$, formally:

$$ {}^{t}\boldsymbol{\alpha} = {}^{E}\mathbf{C}\ ({}^{t}\mathbf{x}) \tag{8} $$

If the objective is the classification at the later time *t+1,* the FCM transition law can be applied to infer the membership values at time *t+1* based on the membership distribution ${}^{t}\boldsymbol{\alpha}$ for the earlier time *t.* Thus, if **T** is the class transition matrix representing the class transitions in two consecutive instants, the classification at time *t+1* can be estimated through combining equations (3) and (8), yielding:

$$ {}^{t+1}\boldsymbol{\beta} = {}^{E}\mathbf{C}\ ({}^{t}\mathbf{x}) \circ \mathbf{T} \tag{9} $$

The two fuzzy label vectors ${}^{t+1}\boldsymbol{\alpha}$ and ${}^{t+1}\boldsymbol{\beta}$ are then combined in the next step by an *aggregation* function **F** to form a *multitemporal fuzzy label vector* ${}^{t+1}\boldsymbol{\mu}= [{}^{t+1}\mu_1, ..., {}^{t+1}\mu_n]$ given by:

$$ {}^{t+1}\boldsymbol{\mu} = \mathbf{F}({}^{t+1}\boldsymbol{\alpha}, {}^{t+1}\boldsymbol{\beta}) = \mathbf{F}[{}^{L}\mathbf{C}\ ({}^{t+1}\mathbf{x}) , {}^{E}\mathbf{C}\ ({}^{t}\mathbf{x}) \circ \mathbf{T}] \tag{10} $$

The final step is the defuzzification, performed by a function of the form:

$$ \mathbf{H}: [0\ 1]^n \rightarrow \Omega^n \tag{11} $$

which transforms the fuzzy label vector ${}^{t+1}\boldsymbol{\mu}$ into a crisp one. Putting it all together, the multitemporal classifier ${}^{F}\mathbf{M}$ is given by Equation (12), depicted graphically in Figure 11.

$$ {}^{t+1}\mathbf{w} = {}^{F}\mathbf{M}({}^{t+1}\mathbf{x}, {}^{t}\mathbf{x}) = \mathbf{H}\{\ \mathbf{F}[{}^{L}\mathbf{C}\ ({}^{t+1}\mathbf{x}) , {}^{E}\mathbf{C}\ ({}^{t}\mathbf{x}) \circ \mathbf{T}]\ \} \tag{12} $$

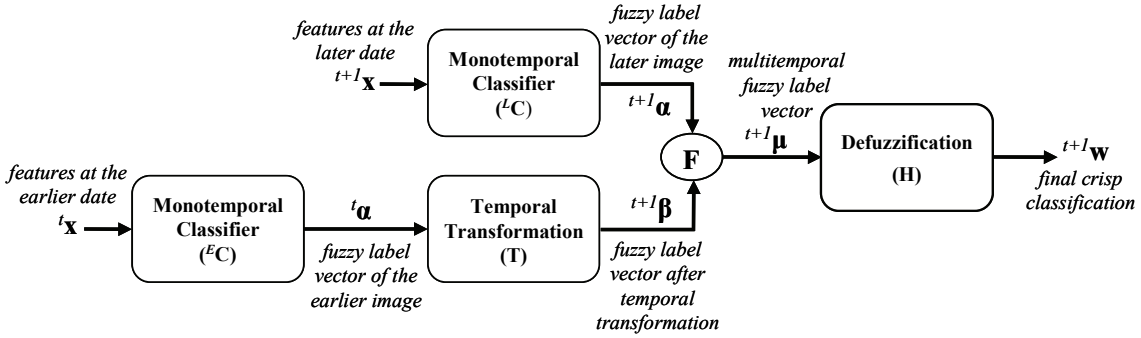

Figure 11. The *forward* multitemporal classification model.

The schema depicted in Figure 11 and formally described by Equation (12) uses features of two dates but aims at classifying the objects at the later date. This will be called hereafter as the *forward* formulation of the model, denoted by the prefix *F*.

The same reasoning can be used for a model that aims at the classification of the earlier image. By rearranging the terms of Equation (12) and assuming that the inverse of a fuzzy relation represented by $\mathbf{T}$ is given by $\mathbf{T}^{-1}$, where $\mathbf{T}^{-1}$ is equal to the transpose of $\mathbf{T}$. Thus, the *backward* formulation, illustrated in Figure 12, is given by:

$$^{t}\mathbf{w} = {}^{B}\mathbf{M}(^{t}\mathbf{x}, {}^{t+1}\mathbf{x}) = \mathbf{H}\{\, \mathbf{F}[{}^{E}\mathbf{C}\,(^{t}\mathbf{x}) \,,\, {}^{L}\mathbf{C}\,(^{t+1}\mathbf{x}) \circ \mathbf{T}^{-1}] \,\} \qquad (13)$$

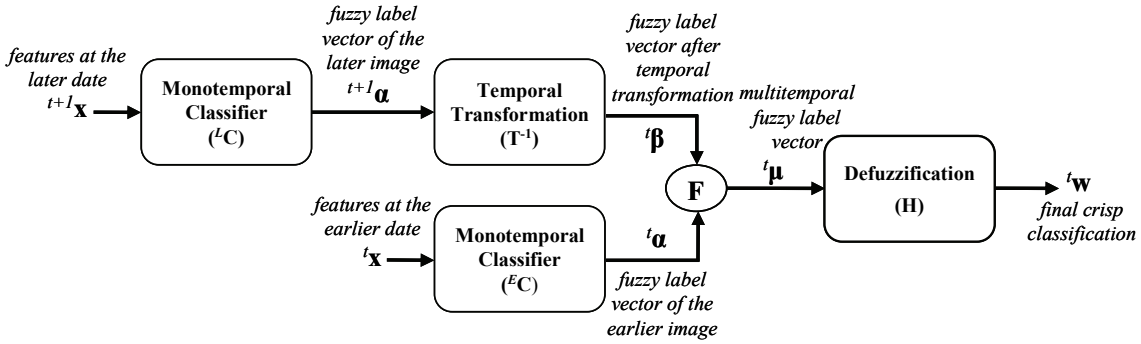

Figure 12. The *backward* multitemporal classification model.

3.4. Particularization of the classification model

FMC models may be built using any t-norm and s-norm composition. However, the *max-product* composition is favored hereafter, since it leads to a simple model with an intuitive interpretation, as will be later shown. Thus Equation (2) takes the form:

$$ {}^{t+1}\beta_j = \max_{i=1,\ldots,n} \left\{ {}^{t}\alpha_i \, \tau_{ij} \right\} \text{ for } i,j=1,\ldots,n \tag{14} $$

Equations (12) and (13) describe the proposed solution to the multitemporal classification problem formulated in Section 3.2 in quite general terms. No restrictions are imposed on the input features ${}^{t}\mathbf{x}$ and ${}^{t+1}\mathbf{x}$ at both dates, nor on the design of classifiers ${}^{E}\mathbf{C}$ and ${}^{L}\mathbf{C}$.

For the hardening function **H**, the aggregation function **F** and the t-norm and s-norm compositions building up the operator "$\circ$", there are various conceivable alternatives. This work does not investigate all these alternatives, but concentrates instead on one implementation of the model proposed in the previous section, which is defined by a particular choice for, **F**, "$\circ$" and **H**, as follows.

The defuzzyfication step is carried out by a hardening function **H** that selects the fuzzy set with the highest membership grade, formally:

$$ [w_1,\ldots,w_n] = \mathbf{w} = \mathbf{H}(\boldsymbol{\mu}) = \mathbf{H}([\mu_1,\ldots,\mu_n]) \text{ where } w_j = \begin{cases} 1 \text{ for } \mu_j = \max\{\mu_1,\ldots,\mu_n\} \\ 0 \text{ otherwise} \end{cases} \tag{15} $$

For aggregation, a function **F** whose outcome is the product of corresponding elements of the input fuzzy vectors was selected. Thus:

$$ {}^{t+1}\boldsymbol{\mu} = \mathbf{F}({}^{t+1}\boldsymbol{\alpha}, {}^{t+1}\boldsymbol{\beta}) = [{}^{t+1}\alpha_1 \, {}^{t+1}\beta_1, \, {}^{t+1}\alpha_2 \, {}^{t+1}\beta_2, \ldots, {}^{t+1}\alpha_n \, {}^{t+1}\beta_n] \tag{16} $$

In (Mota et al., 2007) the geometric mean is used instead of the product for aggregating fuzzy label vectors. Since the selected hardening function just takes the class with the highest membership grade, the geometric mean and the product turn out to be equivalent in terms of the final classification outcome.

Finally, the max-product composition was chosen for the vector and matrix operator "∘" introduced in Section 3.1. This particular choice of **H**, **F** and "∘" results in a classification model with interesting characteristics. Let's first consider the forward model depicted in Figure 11. According to equations (14) and (16):

$$ {}^{t+1}\mu_m = {}^{t+1}\alpha_m \; {}^{t+1}\beta_m = {}^{t+1}\alpha_m \max_l \left({}^{t}\alpha_l \tau_{lm} \right) = \max_l \left({}^{t}\alpha_l \tau_{lm} \; {}^{t+1}\alpha_m \right) \tag{17} $$

The hardening function **H** defined in Equation (15) will finally assign the object at time $t+1$ to the class ω_j for which:

$$ {}^{t+1}\mu_j = \max_m \left({}^{t+1}\mu_m \right) = \max_m \left(\max_l \left({}^{t}\alpha_l \tau_{lm} \; {}^{t+1}\alpha_m \right) \right) = \max_{l,m} \left({}^{t}\alpha_l \tau_{lm} \; {}^{t+1}\alpha_m \right) \tag{18} $$

Applying the same reasoning for the backward model illustrated in Figure 12, and assuming that the inverse of **T** is given by its transpose, $\mathbf{T}^{-1}$, results that the object at time t will be assigned to the class ω_i for which:

$$ {}^{t}\mu_i = {}^{t}\alpha_l \max_m \left({}^{t+1}\alpha_m \tau_{ml}^{-1} \right) = \max_{m,l} \left({}^{t+1}\alpha_m \tau_{ml}^{-1} \; {}^{t}\alpha_l \right) = \max_{l,m} \left({}^{t}\alpha_l \tau_{lm} \; {}^{t+1}\alpha_m \right) \tag{19} $$

Now one can explore the fact that the right side of equations (18) and (19) are equal. This means that the values $i=l$ and $j=m$ for which ${}^{t}\alpha_l \tau_{lm} \; {}^{t+1}\alpha_m$ is maximum over $l,m=1,\ldots,n$ define the classes ω_i and ω_j to which the object is to be assigned respectively at times t and $t+1$.

In other words, to find the classes to which the object belongs to at times t and $t+1$, we simply need to find the indices i and j that maximizes the product ${}^{t}\alpha_i \tau_{ij} \; {}^{t+1}\alpha_j$ for all possible class transitions, what can be done by the algorithm presented in Figure 13.

It is not difficult to demonstrate that the proposed multitemporal classification model subsumes the model introduced in (Mota et al. 2007). As the later aims at the classification of image objects at a later time, it will be confronted with the forward formulation of the proposed model.

```
procedure Multitemporal_Classification (ᵗx, ᵗ⁺ᵏx, ᴱC, ᴸC, T, n)
begin
  for each image object
    begin
      compute ᵗα=ᴱC(ᵗx) and ᵗ⁺ˡα=ᴸC(ᵗ⁺ˡx)
      for l:=1 to n
        for m:=1 to n
          compute p_lm = ᵗα_l τ_lm ᵗ⁺¹α_m
      return classes ω_i for t and ω_j for t+l, where p_ij = max_{l,m}{p_lm}
    end
end
```

Figure 13. Algorithm for determining the classes of an image object at times *t* and *t+1* (argument *n* represents the total number of classes).

The model from (Mota et al. 2007) requires that the class of the object being classified is known for the earlier date (*t*). This can be represented by a crisp label vector ${}^{t}\mathbf{W} \in \mathbf{\Omega}^{n}$, with the true classification at time *t*.

Such information can be understood as the outcome of an ideal classifier that always produces the true classification (${}^{t}\mathbf{W}$). How this classifier is designed, whether it is automatic or not, does not matter here, as long as its output, namely, the true classification at time *t,* is available. Let's then use such ideal classifier as the earlier classifier in the scheme of Figure 11. In this case:

$$ {}^{t}\boldsymbol{\alpha} = {}^{t}\mathbf{W} \tag{20} $$

Let's now assume that the object being classified belongs to class ω_l at time *t*. Hence, ${}^{t}\mathrm{W}_l$ is the only non zero element of ${}^{t}\mathbf{W}$. Introducing this information in Equation (17) results that the object being classified will be assigned to the class $\omega_i \in \mathbf{\Omega}$ at time *t+1,* for which:

$$ {}^{t+1}\mu_i = \max_k \left\{ {}^{t+1}\alpha_k \tau_{lk} \right\} \tag{21} $$

which matches the model proposed in (Mota et al., 2007).

Finally, an analogy with the Bayes classifier induces an intuitive interpretation for the factors $\max_l\left({}^{t}\alpha_l \tau_{lm}\right)$ and $\max_m\left({}^{t+1}\alpha_m \tau_{ml}^{-1}\right)$ that appear respectively in equations (17) and (19). Those factors can be understood as a sort of "a priori possibility" of classes being considered for the object at a particular date. Taking the forward formulation, the factor $\max_l\left({}^{t}\alpha_l \tau_{lm}\right)$ in Equation (17) can be regarded as the a priori possibility for class ω_m in time $t+1$, based on the description of the object at time t. The same reasoning applies for the backward formulation.

3.5. Estimating transition possibilities

The estimation of transition possibilities is a crucial issue in this proposal. This section presents estimation procedures that try to maximize classification accuracy computed over a given training set.

Classification accuracy is assumed to be given by a function G, which computes somehow the agreement between the crisp label vectors generated by the multitemporal classifier ($^{t}\mathbf{w}$, $^{t+1}\mathbf{w}$) and the corresponding true crisp label vectors ($^{t}\mathbf{W}$, $^{t+1}\mathbf{W}$) for a given set S of image objects (the training set).

Function G may be associated to classification accuracy at the earlier date, at the later date, or at both dates simultaneously. In any case, according to equations (12) and (13), the outcome of the multitemporal classifier depends on the selected monotemporal classifiers $^{E}\mathbf{C}$ and $^{L}\mathbf{C}$, on the transition possibilities $\mathbf{T}$, and on the feature values of the objects in set S, described by $^{t}\mathbf{x}$ and $^{t+1}\mathbf{x}$.

The estimation procedure consists of finding the set of transition possibility values $\mathbf{T} = \{\tau_{ij}\}$ that maximizes the selected accuracy function, for the selected image objects, and for the selected monotemporal classifiers. This is formally expressed by:

$$\mathbf{T} = \arg\max_{\tilde{\mathbf{T}}}\left\{G\left(S, {}^{E}\mathbf{C}, {}^{L}\mathbf{C}, \tilde{\mathbf{T}}\right)\right\} \tag{22}$$

To be rigorous, the classification accuracy depends also on the implementation of the t-norm and s-norm, on the selected aggregation method and

hardening function. Since these options of the multitemporal classification model are kept fixed throughout this work, and on behalf of notation simplicity, the dependence on them does not appear in Equation (22).

At this point it is worth clarifying what is actually meant in this work by *transition possibility* and how it relates to the concept of *transition probability* as defined in the context of the probabilistic Markov model formulation. In probabilistic Markov models the probability of a transition from class ω_i to class ω_j is given by the probability that an image object belongs to class ω_j in time $t+1$, given that it belongs to the class ω_i in time t. Note that transition probabilities can be estimated (for a sufficiently large training set) by the proportion of objects belonging to class ω_i at time t and to class ω_j one time unit later. Clearly, transition probabilities do not depend on the classifiers being used to discriminate the classes, neither on how classification accuracy is measured in a given application.

Transition possibilities are different in this respect. Considering Equation (22) as a definition of transition possibilities, it becomes clear that they not only depend on the temporal class dynamics in the target area, but also on the monotemporal classifiers and on the selected accuracy function. Therefore, as the transition possibilities depend on the characteristics of all these elements[6], any change in those characteristics may imply in significant differences in the transition possibility values. This can be observed in the experiments reported in Chapter 4, in which two different monotemporal classifiers for the earlier date and two different accuracy functions are analyzed.

3.5.1. Accuracy functions

Any accuracy function consistent with the formulation of Equation (22) can be used to guide the transition possibilities estimation for the proposed multitemporal classification model. In this work two very general accuracy functions were selected: average per class accuracy and overall accuracy.

[6] In fact, the transition possibility values reflect the entire design of the multitemporal classifier.

Given a set S of image objects and a transition possibility matrix **T**, and considering the classification results for one of the two images, classification accuracy for class ω_i can be defined as:

$$f_i(S,\mathbf{T}) = \frac{c_i(S,\mathbf{T})}{n_i(S)} \tag{23}$$

where $c_i(S,\mathbf{T})$ is the number of image objects in S belonging to class ω_i, correctly classified, and $n_i(S)$ is the number of image objects of class ω_i in set S. Clearly $f_i(S,\mathbf{T})$ provides a measure of the capacity of the classifier to correctly recognize objects of class ω_i.

Now the *average class accuracy* (G_a) in terms of class accuracies can be expressed as:

$$G_a(S,\mathbf{T}) = \frac{1}{n}\sum_{i=1}^{n} f_i(S,\mathbf{T}) \tag{24}$$

where n is the number of classes.

Similarly, the *overall accuracy* may be expressed as:

$$G_o(S,\mathbf{T}) = \sum_{i=1}^{n} f_i(S,\mathbf{T})\pi_i(S) \tag{25}$$

where $\pi_i(S) = n_i(S)/n$ is the proportion of image objects in S that belong to ω_i at the later date. For simplicity, the dependence of the functions on the monotemporal classifiers are omitted in equations (23) to (25).

Equation (25) shows explicitly the dependence of the overall accuracy on the distribution $\{\pi_i(S)\}_{i=1,\ldots,n}$ of objects in S among classes. Hence, if the overall accuracy is the accuracy metric of concern, the distribution of objects among the classes in the training set must be as similar as possible to the distribution of all objects in order to avoid erroneously rating the accuracy measure for some classes.

It was stated before that transition possibility estimates expresses not only temporal class dynamics since they depends on particular factors, such as the monotemporal classifiers and the choice of sample segments (see Equation (22)).

From the discussion above it becomes clear that if overall accuracy is the accuracy metric used in the estimation procedure, the transition possibility values will incorporate information about class transition frequencies. In this case, one can even say that if no temporal correlation exists among the classes in the two epochs, transition possibilities will be strongly influenced by class occurrence frequency information.

3.5.2. Estimation techniques

The computation of the transition possibilities defined in Equation (22) requires an optimization procedure; nevertheless the classification model is not bound to any particular parameter optimization technique. The theoretical maximum number of class transition possibilities to estimate is equal to the square of the number of classes (n^2). In most practical applications this number can be reduced by setting to zero the possibility of class transitions that can never occur. This information is a form of prior knowledge that can be obtained from an expert on the target site.

In the experiments described in Chapter 4 both the average class accuracy and the overall accuracy are used as accuracy functions. As for the optimization procedure, a stochastic and an analytical method were used.

In (Mota et al., 2007), a Genetic algorithm (GA) was used to estimate transition possibilities, using as objective function the average class accuracy (Equation 25) calculated over a training set *S* of image objects for which the true classes at the later date were known. Inspired in the evolution theory, Genetic algorithms represent a stochastic technique applied in many optimization problems (Davis, 1990).

Although the technique is computational intensive, and does not guarantee that the global optimal will be reached, they can work well with virtually any objective function. In this work, GA is one of the techniques used to estimate transition possibilities. As it will be described in Section 4.4, the technique was used in two sets of experiments, one using average class accuracy and the other using overall accuracy as objective functions.

In addition to the above mentioned stochastic approach, an analytical procedure was devised for the estimation of the transition possibilities, and in the experiments the classification results using transition matrixes estimated through the two methods were later compared.

The analytic procedure is based on the decision rule presented in Section 3.4, namely, finding the indices *i* and *j* that maximizes the product ${}^{t}\alpha_i \tau_{ij} {}^{t+1}\alpha_j$ for all possible class transitions, being ω_i the classification result for the object at the earlier time, and ω_j the class to which the object is assigned at the later time.

Let's assume that there is a set of geographical objects known to belong to classes ω_i and ω_j respectively at times *t* and *t+1*. These objects can be eventually used as training patterns for the monotemporal classifiers ${}^{E}\mathbf{C}$ and ${}^{L}\mathbf{C}$ according to a conventional supervised approach. Once ${}^{E}\mathbf{C}$ and ${}^{L}\mathbf{C}$ have been trained, the membership values ${}^{t+1}\boldsymbol{\alpha} = [{}^{t+1}\alpha_1, {}^{t+1}\alpha_2, \ldots, {}^{t+1}\alpha_n]^T$ and ${}^{t}\boldsymbol{\alpha} = [{}^{t}\alpha_1, {}^{t}\alpha_2, \ldots, {}^{t}\alpha_n]^T$ can be computed for each training object.

Assuming that the proposed classification model works properly for the training set, the following requirements must be true:

$$ {}^{t}\alpha_i \tau_{lm} {}^{t+1}\alpha_m - {}^{t}\alpha_i \tau_{ij} {}^{t+1}\alpha_j < 0 \quad \text{for all } (l,m) \neq (i,j) \tag{26} $$

Let's now take the sigmoid function given by:

$$ sig(x) = \frac{1}{1+\exp(-ax)} \tag{27} $$

whose plot is shown in Figure 14.

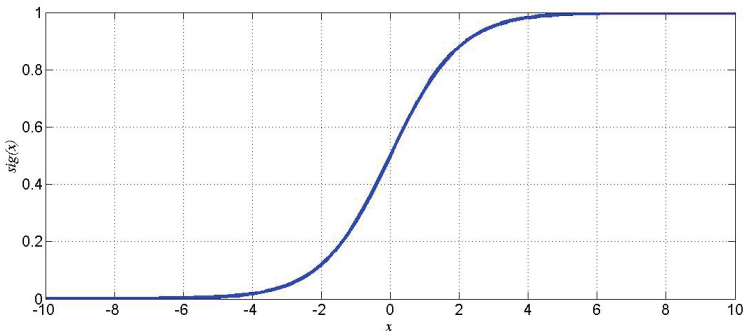

Figure 14. Plot of the sig(*ax*) function for *a*=10.

For a stable classification it is desirable that the left side of the inequality in Equation (26) is as close to -1 as possible, hence:

$$sig\left({}^{t}\alpha_{l}\tau_{lm}{}^{t+1}\alpha_{m} - {}^{t}\alpha_{i}\tau_{ij}{}^{t+1}\alpha_{j}\right) = 0 \quad \text{for all } (l,m) \neq (i,j) \tag{28}$$

and this should be valid for all training samples.

The equation above generates a homogeneous nonlinear equation system consisting of up to n^2-1 equations for each sample in up n^2 unknowns (the state transition possibilities τ_{ij}). Recall that the number of unknowns can be reduced by making $\tau_{ij} = 0$ for the impossible transitions.

This equation system can be solved by standard optimization techniques such as the Newton-Gauss based methods (Nocedal and Wright, 1999). The search space must be restricted to the interval [0 1], and a good starting solution can be built by setting all τ_{ij} (the ones that must be estimated) equal to 0.

As it was said in the previous section, if overall accuracy is the metric of concern, the training set must have a similar object-per-class distribution with respect to the actual distribution, considering all objects in the two images. To be precise, the training set must have a frequency of class transitions similar to that of the whole set of objects. As that information is not available beforehand, in the experiments described in Chapter 4 transition possibilities were estimated using classified objects from a pair of two earlier images.

On the other hand, if average class accuracy is the target accuracy metric, the frequency of class transitions in the training set must be similar with respect to each other. To achieve that, one could chose the same number of samples for each class transition, However, considering that some class transitions are more common than others, it makes more sense to introduce a weighting term in the optimization procedure related to the class transition frequencies. Furthermore, as the selected analytic optimization technique is based on least-square approximation, when the target metric is average class the square root of the total number of occurrences of each class transition in the training set is introduced as a denominator in Equation (28), as shown in Equation (29).

$$\frac{sig\left({}^{t}\alpha_{l}\tau_{lm}{}^{t+1}\alpha_{m} - {}^{t}\alpha_{i}\tau_{ij}{}^{t+1}\alpha_{j}\right)}{\sqrt{F_{ij}}} = 0 \quad \text{for all } (l,m) \neq (i,j) \tag{29}$$

where F_{ij} stands for the number of occurrences of transitions from class ω_i to class ω_j in the training set S.

As mentioned before, although very versatile, GAs are computational intensive, and this represents a relative advantage of calculus-based methods, like the one just described. In the experiments reported in Chapter 4, for instance, the processing time for the estimation of a single transition matrix – both estimation procedures coded within the MATLAB software (Mathworks, 2009), and executed on a dual-core CPU @ 2.40 GHz – was of 2.0 seconds for the analytical method, and 11.6 seconds for the GA-based technique.