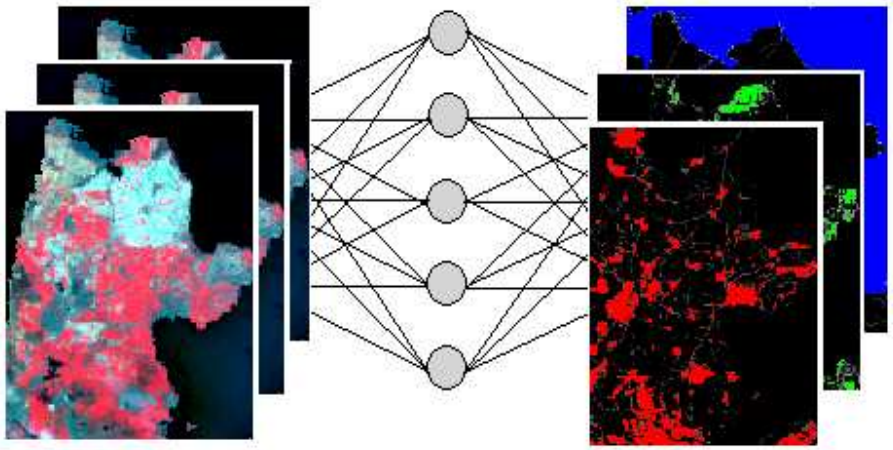
Centre for Geo-Information

Thesis Report GIRS-2006-26

CROP AREA ESTIMATIONS FROM TEMPORAL UNMIXING: A CASE STUDY USING MERIS WITH ARTIFICIAL NEURAL NETWORKS

Ernesto Bastidas Obando



April 2006

WAGENINGEN UNIVERSITY
WAGENINGEN UR

CROP AREA ESTIMATIONS FROM TEMPORAL UNMIXING: A CASE STUDY USING MERIS WITH ARTIFICIAL NEURAL NETWORKS

Ernesto Bastidas Obando

Supervisors:

Dr. Ir. J.G.P.W. (Jan) Clevers
R.(Raúl) Zurita Milla MSc.

A thesis submitted in partial fulfilment of the degree of Master of Science at Wageningen University and Research Centre, The Netherlands.

April 2006
Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2006-26
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

“No hay nostalgia peor que añorar lo
que nunca nunca jamas sucedio”.

Joaquín Sabina

ACKNOWLEDGEMENTS

Many people have collaborated in the development of this thesis and therefore I would like to acknowledge their support during its completion. First of all I would like to thank my supervisors Dr. Jan Clevers and Raúl Zurita Milla for their support and critical comments during the different stages of this research. Their constructive comments kept me focused.

I would like to extend my thanks to Dr. Tasking Kavzoglu for his readiness to provide me with feedback in the subject of artificial neural networks. I am grateful to Harm Bartholomeus for providing me with the LGN5 dataset. A word of gratitude goes to Willy ten Haaf for his advise during my enrolment in the MGI program.

The days in Gaia building, formerly Alterra, were very pleasant spent in the company of my MGI friends coming from different continents. Coffee breaks were always followed by good chats. I will always be indebted to my friends: Alejandra, Jaime, Mariana, Marcos, Roxina, Pablo, Eduardo, Edward, Marjolein. Thanks to them the day-to-day in the Netherlands was out of the routine. I wish them all the best in their future endeavors.

I thank my Colombian and Dutch families for their support during the development of this thesis. And last but not least, I would like to thank Manon for offering me her unconditional love and sharing my dreams and plans for the future.

ABSTRACT

The need for regular information on crop area distribution in European landscapes makes of MERIS a good candidate for temporal monitoring of agricultural systems. Nevertheless, MERIS 300 m spatial resolution might still be too coarse for a hard classification approach because crop fields can be smaller in size than a MERIS pixel. Therefore, the potential for retrieving sub-pixel information from temporal MERIS datasets over crop areas using feedforward artificial neural networks (FFNNs) was evaluated in this study. Two kinds of network structures, the first one using single pixel information and the second one using neighbor information from a 3 by 3 pixel window, were evaluated to estimate sub-pixel fractions based on five MERIS level 2 full resolution images. The five MERIS level 2 datasets were distributed over the growing season and covered the province of Noord-Holland in The Netherlands. A thematically and spatially aggregated version of the Dutch land use database (LGN5), with the main economically important agricultural land cover types for The Netherlands (grassland, potatoes, sugar beet and cereals), was used to reference the MERIS datasets. The LGN5 was also utilized for selection of training, validation and testing samples. Sub-pixel estimations coming from the trained 3 by 3 window FFNN were more precise than the estimations coming from a trained single pixel FFNN. The precision for potato area estimates with a 3 by 3 window FFNN for different spatial scales ranged with coefficient of correlation (R) values from 0.25 at 9 ha to 0.85 at ground areas over 81 ha. For grassland, area estimates ranged with R values from 0.75 at 9 ha to 0.95 at ground areas over 81 ha. Further studies should include a definition of the temporal profiles with more temporal datasets and adoption of look-up tables for training the neural networks.

Keywords: Crop distribution; MERIS; non linear; neural network; feedforward; unmixing

ABBREVIATIONS

ANN	Artificial Neural Network
AVHRR	Advanced Very High Resolution Radiometer
ESA	European Space Agency
FCM	Fuzzy C-Means
FFNN	Feedforward Artificial Neural Network
HRV	High Resolution Visible
JM	Jeffries-Matsushita
LGN	Land cover database for The Netherlands
LGN5	Land cover database for The Netherlands based on datasets of 2003 and 2004
M	Mahalanobis
MERIS	Medium Resolution Imaging Spectrometer
MGVI	MERIS Global Vegetation Index
MODIS	Moderate Resolution Imaging Spectrometer
MSE	Mean Square Error
MTCI	MERIS Terrestrial Chlorophyll Index
$\mathbf{N_i}$	Number of input nodes
$\mathbf{N_o}$	Number of output nodes
$\mathbf{N_p}$	Number of training samples
NIR	Near Infrared
NOAA	National Oceanic and Atmospheric Administration's
R	Coefficient of correlation
r	constant related to the noise level of the data
RD	Rijks Driehoek
RMSE	Root Mean Square Error
$\mathbf{RMSE_{normalized}}$	Normalized RMSE
SPOT	Satellite Pour l'Observation de la Terre
TOA	Top Of Atmosphere
TOAr	Top Of Aerosol
TOC	Top Of Canopy
TM	Thematic Mapper
VI	Vegetation Index
$\boldsymbol{\rho}$	Reflectance values

INDEX

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. BACKGROUND

Updated information on agricultural landscapes is necessary for land use planning, policymaking and monitoring food production. In The Netherlands, for example, the policy of crop rotation on grassland systems and high input – high output annual crops like potatoes and sugar beet led to a reduction of nutrients leaching to ground water and pesticides emissions to air (CCSE-SWCC 1997). In addition, monitoring crop distribution is important in regions where taxation and subsidies depend on the area and crop type. For example, starting July 1st 2006, the European Union will subsidize agricultural regions that convert sugar beet crop areas into other land use types in order to reduce sugar production (European-Commission 2006).

Traditionally, estimates of cropland distribution have been generated by supervised classification of multi-temporal satellite datasets of, for instance, National Oceanic and Atmospheric Administration's (NOAA) Advanced Very High Resolution Radiometer (AVHRR), Envisat Medium Resolution Imaging Spectrometer (MERIS) or Satellite Pour l'Observation de la Terre (SPOT) High Resolution Visible (HRV), among others. With multi-temporal AVHRR data, Quarmby *et al.* (1992) and Atkinson *et al.* (1997) developed land cover maps over large areas of Europe. However, this coarse scale might be still inappropriate for monitoring purposes due to the scale at which most land cover changes occur in Europe (Mücher *et al.* 2000). The use of high spatial resolution datasets like SPOT HRV or Landsat Thematic Mapper (TM) has generated well detailed information on crop cover at regional level (Jewell 1989, Murakami *et al.* 2001, De Wit and Clevers 2004). However, the large amount of data over large areas and its restricted availability due to cloud cover present limitations for using high spatial resolution for this approach (Brisco and Brown 1995, Sakamoto *et al.* 2005).

Medium resolution sensors such as Moderate Resolution Imaging Spectrometer (MODIS) and MERIS can bridge the gap between high and coarse spatial resolution. The area depicted by a MERIS pixel at 300 m is more than ten fold smaller than an AVHRR pixel and 100 fold larger than a Landsat TM pixel. With MERIS data, Verstraete *et al.* (1999) and Clevers *et al.* (2006) have demonstrated the potential of MERIS temporal and spatial characteristics for monitoring heterogeneous landscapes at regional and continental scales. Nevertheless, it is evident that for agricultural landscapes in Europe with field size lower than 300 m, MERIS pixels are coarse and great proportions of the image might be dominated by mixed pixels comprising different crop types.

In order to overcome the mixed pixel problem, a variety of physical, statistical and non-parametric techniques may be used to derive sub-pixel information from mixed pixels. For instance, physical linear mixing models (Foody and Cox 1994) assume a linear combination of class signatures weighted by the class proportions to estimate sub-pixel compositions. Statistical methods such as fuzzy C-means classification (Bezdek *et al.* 1984) assume a degree of membership associated to each class by the distance of a mixed pixel to the class mean. Although these methods have been applied for extracting sub-pixel information on land cover types, they are conditioned to data distribution such as Gaussian distribution (Paola and Schowengerdt 1995) and selection of endmembers (Atkinson *et al.* 1997).

These limitations have lead to more sophisticated methods like artificial neutral networks (Atkinson and Tatnall 1997) which do not assume a priori knowledge of the data. This non-parametric nature allows artificial neural networks to be more robust when datasets are not normally distributed and include mixed pixels in the training stage (Foody 2004). In addition, it is generally agreed that artificial neural networks produce classifications with higher accuracies than those generated through other methods (Atkinson and Tatnall 1997, Braswell *et al.* 2003, Foody and Mathur 2004). However, its complex design and slow training rate due to the trial and error process used to determine parameters such as number of training samples and number of layer nodes have limited their use in

classification approaches (Kavzoglu and Mather 2003). For this study an artificial neural network approach was used to estimate crop area from multi-temporal MERIS datasets.

1.2. PROBLEM DEFINITION

Preliminary studies have demonstrated the potential of using MERIS full resolution (300 m) for hard classification of land cover types (Verstraete *et al.* 1999, Clevers *et al.* 2006). However, hard classification of agricultural crops with MERIS data would be inappropriate for agricultural landscapes dominated by mixed pixels comprising various crop types. This study is therefore designed to derive sub-pixel area information from multi-temporal MERIS datasets on agricultural crops. A good methodology for deriving information on the area of agricultural crops from MERIS at sub-pixel level is not yet available. Further sub-pixel accuracy improvements of agricultural crops can be expected by using multi-temporal MERIS datasets over the growing season (Lobell and Asner 2004).

1.3. RESEARCH OBJECTIVES

The main objective of this thesis work was to evaluate the possibility of using multi-temporal MERIS datasets for extracting sub-pixel crop area information by using feedforward artificial neural networks. The inputs for the networks were multi-temporal MERIS information at pixel level and spatial information coming from neighboring pixels with a 3 by 3 window. The study area was the province of Noord-Holland, in The Netherlands.

For this purpose the following specific objectives were defined:

- To optimize data dimensionality from multi-spectral and multi-temporal MERIS datasets for input into a feedforward neural network.

- To evaluate the effect of neighbor information coming from MERIS datasets on sub-pixel estimates of crop area.

- To investigate the performance of sub-pixel crop area estimations of feedforward artificial neural networks with different scale of sub-pixel information and landscape heterogeneity.

These three objectives led to the following research questions:

- What kind of method is the most appropriate to assess the separability of crop types from multi-temporal and multi-spectral MERIS datasets at sub-pixel level?

- Can neighbor pixel information from MERIS datasets improve sub-pixel crop area estimates?

- Does the accuracy of sub-pixel crop area estimates change with the spatial scale of analysis and landscape heterogeneity?

1.4. STRUCTURE OF THIS REPORT

This report is structured in five chapters. Chapter one gives a short introduction to the subject, provides the problem definition and points out the objectives and research questions that will be investigated in this report. In chapter two brings a literature review regarding crop discrimination from remote sensing and sub-pixel extraction techniques with emphasis in FFNN. Chapter three provides all necessary information of the experimental setup and explains the used techniques. In chapter four, the results are presented and discussed. Finally conclusions and recommendations are presented in chapter five.

2. LITERATURE REVIEW

2.1. CROP DISCRIMINATION FROM REMOTE SENSING

Crop discrimination from remote sensing data has used mainly optical sensors with bands in the visible and near infrared (NIR) because most of the vegetation activity affects reflectance in those bandwidths. Reflectance in the visible part is low due to absorption by the chlorophyll. In the NIR region the absorption is very low, and reflectance is determined by the amount of transition between cell wall and air vacuoles in the leaf tissue. As result, NIR reflectance is high, and a steep slope occurs in the curve at about 700 nm, the so called red-edge region (Clevers and Jongschaap 2001). This spectral response changes with the phenological state of the crops. At early stages the visible and NIR would show a flat reflectance due to a bare soil signature. Once crops emerged and augment in biomass, reflectance in the NIR would steady increase showing a clear vegetation spectrum. Late in the season when crops have reached maturity, the NIR would decrease due to senescence and harvest. Therefore the use of multi-temporal datasets throughout the growing season are recommended for crop classification, since a single image contain a high level of spectral confusion between crops for their accurate classification (Jewell 1989). The optimal acquisition dates for satellite imagery are determined by the phenological characteristics and agricultural practices of crops and by environmental conditions. Nevertheless, the minimum required datasets for achieving good separability might include datasets taken at early, medium and late stages of the growing season (Murakami *et al.* 2001, De Wit and Clevers 2004, Sakamoto *et al.* 2005).

In addition to the temporal domain, it is often appropriate to have fine spatial resolution. Otherwise the spectral information at pixel level might come from mixed pixels comprising different crop types (Murakami *et al.* 2001). For some regions it is difficult to get cloud free fine spatial resolution data of, for instance, Landsat TM or SPOT HVR over the growing season and therefore coarse resolution sensors with higher temporal resolution might increase those chances. To overcome the mixed pixel problem in coarse

datasets various techniques may be used to derive sub-pixel information. The use of sub-pixel analyses and their implementation in vegetative environments are presented in sections 2.3 and 2.4.

2.2. MULTI-TEMPORAL PROFILE

The use redundant multi-spectral and multi-temporal information may lead to unexpected problems if artificial neural networks are used to produce thematic information. Redundancy of information may cause an increase of calculation time and inter class confusion caused by the use of unnecessary information (Kavzoglu and Mather 2002).

In order to reduce the redundant information coming from multi-spectral sensors with high correlated bands, as presented in MERIS (Clevers *et al.* 2006), it is possible to calculate the principal components or to define the most relevant bands, which are less correlated. In the case of vegetated environments a combination of bands in the red and NIR wavelength can also be used to reduce the spectral dimensionality of the dataset. An example of the latter is the use of vegetation indices.

Vegetation indices (VIs) have been formulated as indicators sensitive to vegetative properties like chlorophyll activity and canopy structure (De Jong *et al.* 2004). Thus, they aim at enhancing the vegetation signal and minimizing contributions from soil background, sun elevation angle and atmosphere. For example, the MERIS Global Vegetation Index (MGVI) is an index designed for canopy characterization (Gobron *et al.* 2004). On the other hand, the MERIS Terrestrial Chlorophyll Index (MTCI) is more sensitive to chlorophyll content (Dash and Curran 2004). Hence temporal profiles from VIs or bands in the red and NIR can be used for detection of crop changes throughout the growing season.

Classification accuracy of temporal profiles can differ from combinations of temporal datasets. However, the use of a large temporal dataset is not always necessary to achieve the best classification accuracy. For example, Murakami *et al.* (2001) used separability

analyses to evaluate the temporal combination for a optimal classification of cropping systems. They found that the accuracy provided with three temporal datasets would achieve nearly equal accuracy that the one obtained with nine datasets. This reduction in temporal information would considerable reduce the computation time of the artificial neural networks without affecting the classification accuracy. Therefore, separability analyses of the temporal profiles can help to reduce the data load into the neural networks.

2.3. SUB-PIXEL METHODS IN REMOTE SENSING

Several approaches may be used to extract information at sub-pixel level. These techniques can be grouped into two main approaches as defined by Foody (2004):

- Linear mixing models
- Soft classification models
    - Maximum likelihood
    - Fuzzy C-Means
    - Artificial neural networks

2.3.1. Linear mixing models

Linear mixing models simulate the reflectance of a pixel as a linear combination of the reflectances by endmembers (pure materials), weighted by the areal fraction of each endmember within the pixel:

$$p = \sum_{i=1}^{m} C_i p_i + \varepsilon \qquad \textbf{(1)}$$

where p is the observed pixel reflectance, C_i and p_i are the fractional cover and reflectance, respectively, of the ith endmember, and ε is a residual representing the model error. To solve for the fractions of each of m endmembers requires at least m equations, which are most commonly generated by repeating Equation 1 for different wavelengths.

However, the maximum amount of endmembers can never be more than the number of bands, because having more endmembers than bands results in more unknowns than equations (Lillesand and Kiefer 2000).

Although this simple approach has been widely used for deriving sub-pixel information of proportions of different vegetated environments (Foody and Cox 1994, Gong *et al.* 1994, Van Kootwijk *et al.* 1995, Lobell and Asner 2004), some authors are concerned about their use in vegetated environments (Borel and Gerstl 1994, Foody *et al.* 1997). Their major concern is that endmemeber selection in vegetated environments may be influenced by multiple scattering of leaves, soil background and by mixed endmembers. This effect makes the assumption of linear mixing models inappropriate in environments where non linear spectra mixing occurs. In addition, endmembers chosen from coarse datasets in vegetated environments may comprise areas with different phenological stages, thus hindering separability between classes (Braswell *et al.* 2003). Due to these limitations it has been suggested that a non-linear mixing approach may be more appropriate for vegetated environments than linear mixing models (Foody 2004, Lobell and Asner 2004).

In addition, the spatial resolution of the sensor may limit even further the selection of endmembers as the proportions of mixed pixels increases with a coarsening of the spatial resolution of the sensor (Atkinson and Aplin 2004). Assuming that for the region of Noord-Holland in the Netherlands, most of the agricultural fields have an area of less than 10 ha (Table 2), a MERIS scene with pixel size of 300 m may comprise a high percentage of mixed pixels reducing thus, the possibilities of a proper selection of endmember.

2.3.2. Soft classification models

Soft classification models offer a non linear approach to estimate land cover proportions at pixel level. They derive land cover proportions from the membership a pixel spectra information displays to each class. One concern in using this technique, likewise the linear mixing approach, is the requirement to have a well defined set of endmembers.

Nevertheless, some of these models like artificial neural networks have been refined and may not require an exhaustive selection of endmembers or even may allow the use of mixed pixels for training (Benediktsson *et al.* 1993, Benediktsson and Sveinsson 1997).

Maximum likelihood classification

This technique is based on the assumption that the position of the spectral signature of a mixed pixel relative to the means of pure endmembers gives an approximate maximum likelihood estimate of the proportion classes under normally distributed input datasets (Ichoku and Karnieli 1996). Thus, if two components are denoted by X and Y, the relationship for estimating the proportions of one of them in a pixel is given by Marsh *et al.* (1980):

$$P_y = 0.5 + 0.5\frac{d(m,x) - d(m,y)}{d(x,y)} \quad \textbf{(2)}$$

where P_y = proportion of class Y in the mixed pixel; $d(x,y)$ = Mahalanobis (M) distance between the mean endmember class X and Y; $d(m,x)$ = M-distance between the mixed pixel (m) and the endmember X; $P_y = 0$, if the estimate is negative; and $P_y = 1$, if the estimate is greater than 1. The M-distance may be calculated from:

$$d(x,y) = \sqrt{(x-y)^T \Sigma^{-1} + (x-y)} \quad \textbf{(3)}$$

where $d(x,y)$ is the M-distance of the mean endmembers X and Y with a covariance matrix Σ. T means transpose.

Although the maximum likelihood technique can be used to resolve proportional fractions, this technique is constrained to discriminate between two components. In addition, their use for sub-pixel information is being questioned as there is not a direct link between the proportional coverage of a class and its probability (De Bruin 2000).

Fuzzy C-Means (FCM)

The FCM is a model that has been used for the estimations of proportions of different plant communities (Wang 1990, Foody 1992, Foody and Cox 1994). The FCM is a non-hierarchical clustering algorithm that may be used to subdivide data into more than two components. Pixels are associated with several classes before being moved iteratively to different classes aiming at minimizing the least square error, J_m which is a weighted measure of the square distance between pixels and class centroids.

$$J_m(U,v) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m (d_{ik})^2 \qquad \textbf{(4)}$$

Where U is a fuzzy c-partition of the data containing n pixels $(x_1, x_2, ..., x_n)$, v_i is the center of the cluster i, d_{ik} is the distance between x_k and v_i measured using an appropriate weight matrix and *m* is a user defined weighting component that lies within the range $1 \le m \le \infty$, which determines the degree of fuzziness of an analysis. When $m = 1$ a conventional hard classification is obtained in which each pixel is associated with just one class. For the derivation of sub-pixel scale information, it is important that $m > 1$ assuring that multiple and partial class membership is allowed in the output. The grade of membership u_{ik} to a class is calculated from (Bezdek *et al.* 1984):

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}} \qquad \textbf{(5)}$$

Memberships close to unity indicate a high degree of similarity between x_k and a class *i* whereas memberships close to zero indicate little similarity.

Although the FCM has been commonly used to derive sub-pixel information, its accuracy relies on the specification of end-member spectra (Atkinson *et al.* 1997) and the careful selection of the value *m* used in the analysis (Foody 1996).

Artificial Neural Networks

Neural network models have been employed since the last decade in classification approaches. Their rapid development in remote sensing is due to their ability to perform more accurately than the aforementioned techniques (Benediktsson *et al.* 1993, Atkinson and Tatnall 1997, Braswell *et al.* 2003, Foody and Mathur 2004) and their ability to incorporate different types of data, including mixed pixels (Benediktsson *et al.* 1993, Benediktsson and Sveinsson 1997).

A description of their structure and functionality is presented in the following section.

2.4. ARTIFICIAL NEURAL NETWORKS (ANNs)

An Artificial Neural Network (ANN) can be presented as a data modelling tool that is able to capture and represent complex input/output relationships. They are the results of research to model the brain like computing (Haykin 1994). ANNs resemble the brain in two aspects:

- Knowledge is acquired by the network through a learning process;
- Inter node connection strengths known as synaptic weights are used to store knowledge.

The basic element of an ANN is the processing node or perceptron (Figure 1). Each processing node sums the value of its inputs. This is passed through an activation function to produce the node's output value.

There are many different types of organizing the processing nodes and so different types of neural networks (see for example, Demuth *et al.* (2005)). This section provides a description to the most commonly used ANN in remote sensing, the multi-layer perceptron, a feedforward artificial neural network model (FFNN).

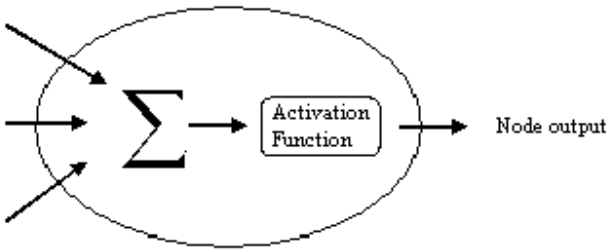


Figure 1. Internal structure of a perceptron.

The multi-layer perceptron model consists of several perceptrons organized into layers. A generic design of this type of network is shown in Figure 2.

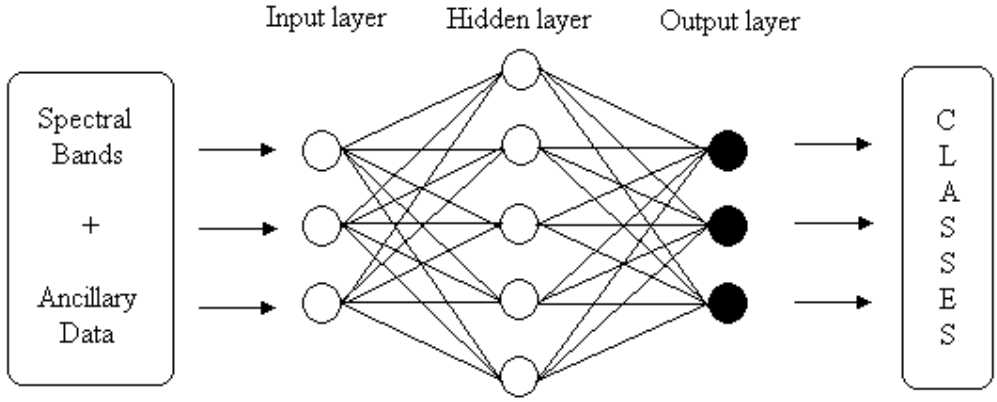


Figure 2. Generic three layer connected FFNN (Kavzoglu and Mather 2003).

The perceptrons are organized into layers. The first layer, which receives input information, is called the input layer. The last layer, which produces the output information, is called the output layer. Between the input and output layers can be one or more hidden layers. All layers are fully interconnected with the following layer but do not interconnect with other nodes in the same layer. Data is transmitted through connection between perceptrons in different layers. The input of each single perceptron is weighted according to:

$$net_j = \sum \varpi_{ij} o_i \qquad \textbf{(6)}$$

where net_j indicates the sum of the weighted inputs of the node j, ϖ_{ij} represents the weights between node i and node j, and o_i is the output from node i. The output from a given node j is then computed from:

$$o_j = f(net_j) \quad \textbf{(7)}$$

The function f is usually a non-linear log sigmoid function that is applied to the weighted sum of inputs before the signal passes to the next layer. The log sigmoid function maps the interval $(-\infty, \infty)$ onto (0,1) allowing the output of the signal to be interpreted as posterior probabilities summing to 1 (Bishop 1995).

$$f(net_j) = \frac{1}{1+e^{-net_j}} \quad \textbf{(8)}$$

Training a neural network consists of generating weights between each node that is accomplished by using known inputs and outputs and presenting them to the network. The weights are adjusted typically by the backpropagation algorithm, which modifies the weights until a target minimal error between the known and calculated outputs is achieved. The training of the backpropagation algorithm is carried out in two stages. Firstly, the network weights are randomly initialized, the input data is presented to the network and propagated forward to estimate the output value for each training. Secondly, the error is then back propagated through the network altering the weights of the connections according to the generalized delta rule (Rumelhart and MacClelland 1986):

$$\Delta\varpi_{ij}(n+1) = \eta(\delta_j o_i) + \alpha\Delta\varpi_{ji}(n) \quad \textbf{(9)}$$

where η is the learning rate parameter, δ_j is an index of the rate of change of the error, and α is the momentum parameter. This process of feeding-forward signals and

backpropagating the error is repeated iteratively until the error of the network is minimized or reaches a desirable magnitude.

The progress towards the minimum error has a relationship with the number of input parameters (Paola and Schowengerdt 1997). For example, if a network is provided with more information about the image like neighbor pixels, each iteration would have more information about the image than a single pixel iteration, making thus more progress per iteration to the desired error magnitude than just the information provided by a single pixel. In various land cover applications neighbor information has increased the accuracy while increasing the training time (Paola and Schowengerdt 1997, Murthy *et al.* 2003). However, the effect of the neighbor information, likewise a homogeneous filter, may smooth the image, losing thus some detail of information on single pixels (Paola and Schowengerdt 1997).

2.4.1. Drawback of Feed-Forward Neural Networks (FFNNs)

Training a FFNN involves setting several parameters that would influence the capability of the neural network to interpolate and extrapolate data that has not been presented before (generalization). These parameters include: the specification of the number of nodes and architecture, size of the training set and training time, among others.

Atkinson and Tatnall (1997), Paola and Schowengerdt (1997) and Kavzoglu and Mather (2002, 2003) have documented the effect of those parameters on the generalization of the neural network. However, there is still no scientific rule to determine the appropriate parameter values and therefore a trial and error strategy is often used.

Recently, some guidelines have been proposed to determine the range of those parameters. For example, the optimal amount of nodes and structure of a neural network should be not too small to identify the structure of the data (underfitting) or should not be too large to become overspecific for the data (overfitting) (Kavzoglu and Mather 2003). Overfitting occurs when the network memorizes a situation and it has not learned to generalize new situations. Thus, during the training the error achieved is very low but

when new data are presented the error is large. For example, Figure 3 (a) shows the response of a neural network that has been overfitted due to a large number of hidden nodes.

The response of an ANN does not lay merely in its structure but also in the representation of the training data for a particular problem and in the training process. Too few training samples are insufficient to derive the characteristics of the classes, while too many training sets may cause the network to overfit the data and require more learning time. Likewise, training time will affect the generalization of the neural network. The longer a neural network is trained for a specific data set, the more it will prone to memorize the training data and not be able to generalize (Kavzoglu and Mather 2003).

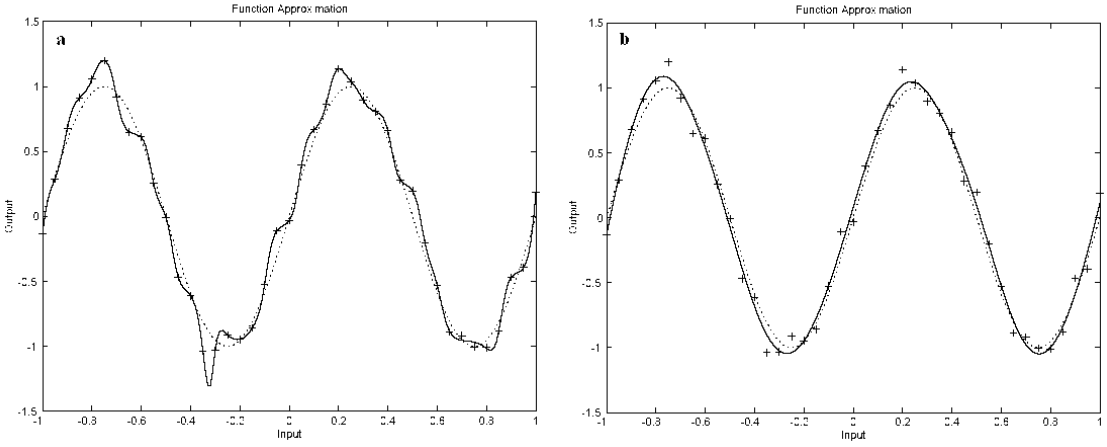


Figure 3. Response of neural networks that have been trained to approximate a sine function. Input function “doted line”, training data “+” symbol, approximation “solid line”. (a) overfitting, (b) good response (Demuth *et al.* 2005).

All those issues slow down the process of training a neural network and makes the use of the ANN more complicated than other methods. In addition, it is not exactly known how ANNs learn particular problems based on the parameter values and training sets and therefore they are often called black box methods.

2.4.2 FFNNs and crop classification

FFNNs have been used for crop classification at sub-pixel level with different parameters in different environments and with various sensors (Kavzoglu and Mather 1999, 2002, 2003, Murthy *et al.* 2003). This makes it difficult to compare or to determine the effects of different parameters on classification accuracies. Nevertheless, all authors have agreed on the superiority of FFNNs for crop classification approaches. For example, Murthy *et al.* (2003) compared a FFNN with a maximum likelihood method for wheat crop classification. They found that a FFNN was more precise in classifying wheat pixels than the maximum likelihood method. Despite the potential of FFNNs for classification approaches, there is a concern about the time employed to determine the appropriate parameter values towards operational application.

Recently, Kavzoglu and Mather (2003) conducted a study to gain some insights into the behavior of FFNNs for crop classification purposes. In this study a guideline that facilitates the selection of appropriate parameter values for designing and training of FFNNs is proposed (this guideline is presented in section 3.3.5).

Based on the literature review, MERIS high temporal and multi-spectral resolution data offer an excellent opportunity to enhance the separability between crops. The aforementioned studies confirm the potential of FFNNs for discriminating sub-pixel composites from coarse datasets. Therefore, for this research we designed a methodology to evaluate the potential of extracting sub-pixel crop information from temporal MERIS datasets by using ANNs of the type FFNN for a heterogeneous landscape in Europe.

3. MATERIALS AND METHODS

3.1 STUDY AREA

The study area includes a large part of the province of Noord-Holland in The Netherlands (Figure 4). The selection for this area was based on the availability of a training dataset with good accuracy (overall accuracy 78%) and accessibility of MERIS 2004 images.

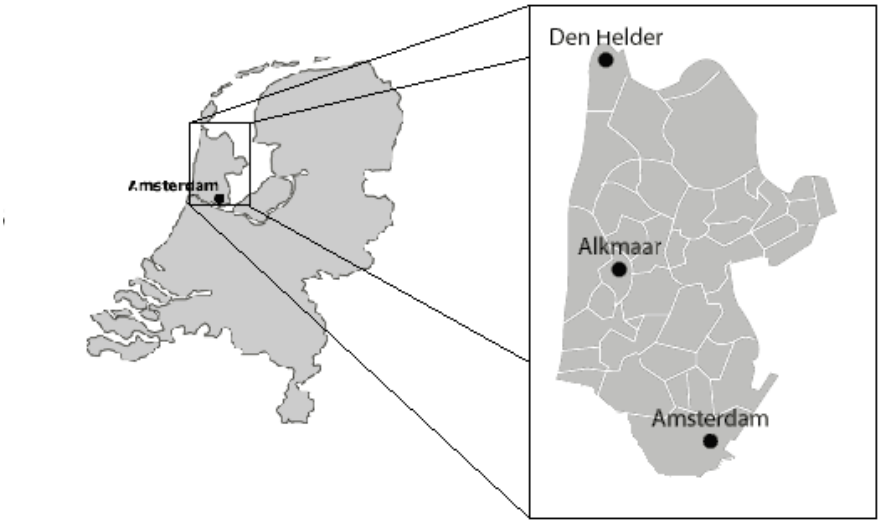


Figure 4. Study area province of Noord-Holland

3.2 DATASETS

3.2.1. Ground cover information

For ground cover information, the latest version of the Dutch land use database (LGN5) was used. It is a geographical database that describes the land use in The Netherlands with a grid structure of 25 meters, with an application scale of 1:50.000. It uses the stereographic projection of the Dutch national coordinate system Rijks Driehoek (RD). The nomenclature of the LGN5 database consists of 39 classes covering urban areas,

water, forest, various agricultural and natural land cover types. LGN is produced from multi-temporal classification of satellite imagery with ancillary data. Currently, the version 5 is based on satellite data of the year 2003 for the provinces of the east of the Netherlands and satellite data of 2004 for the provinces of the western part of the country. The overall classification accuracy for all provinces is 78% with values ranging from 46% till 93% (Hazeu 2005). The 39 classes of the LGN5 were recoded into the nine classes: potatoes, sugar beet, grassland, cereals, other crops, built up, forest, natural vegetation and water. Potatoes, sugar beet and grassland are economically important agricultural land cover types in the Netherlands. Cereals are used in crop rotation programs of potatoes and sugar beet. Built up relates to urban areas and green houses. Other crops comprise areas with maize, spring and autumn flower bulbs and tree nurseries. Forest contains deciduous and coniferous forests. Natural vegetation includes sand dunes, shrubs, swamps and heathland. Water covers salt and fresh water bodies. Information on the frequency distribution of the nine classes is given in Table 1. Information on the distribution of agricultural fields based on crop area is given in Table 2.

Table 1. Frequency distribution of cover types in the province of Noord-Holland

Land cover	% of total area
Potatoes	2.5
Sugar beet	1.4
Cereals	2.0
Grassland	22.3
Other crops	6.8
Built up	12.5
Forest	3.6
Natural vegetation	4.4
Water	44.5

Table 2. Frequency distribution of agricultural fields based on crop area for the year 2002 in the province of Noord-Holland (from CBS (2004)).

Crop type	% of fields with < 10 ha	% of fields between 10 and 20 ha	% of fields between 20 and 30 ha	% of fields with > 30 ha
Potatoes	9.60	5.18	2.17	2.07
Sugar beet	12.29	4.44	0.72	0.39
Cereals	12.66	6.98	1.88	1.28
Grassland	17.00	7.87	1.26	14.19
Total	51.55	24.48	6.02	17.94

For this study, the LGN5 was aggregated to 300 meters cell size displaying the largest cover type fraction per pixel and keeping information on cover composition in fractional images per class. For example, Figure 5 shows that a pixel from the aggregated LGN5 is displayed as built up because this class covers the largest fraction of this pixel. Nevertheless, information on fractional compositions per class (i.e. grassland) is stored in its respective fractional image. This aggregated database was used as a reference for the co-registration of MERIS images and for the selection of training, testing and validation datasets.

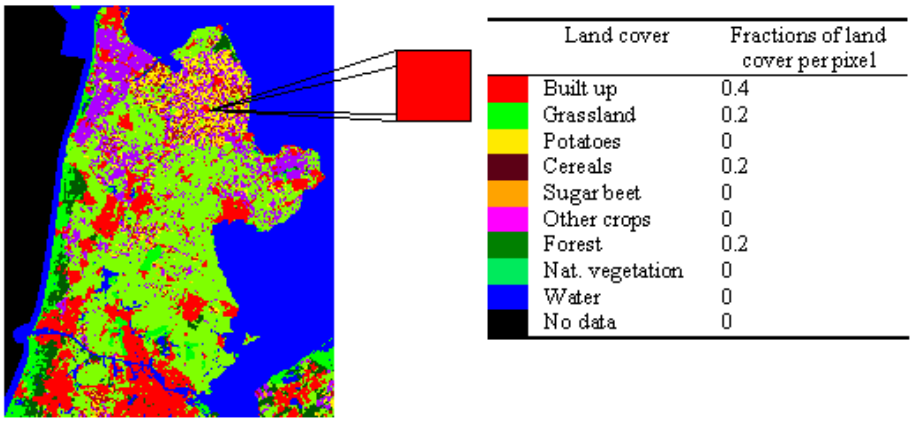


Figure 5. The Dutch land use database aggregated to 9 classes and 300 m pixel size with fractional distribution for Noord-Holland.

3.2.2 MERIS Data

MERIS is one of the payload components of the European Space Agency's (ESA) environment research satellite Envisat, launched in March 2002. MERIS is a 68.5^{o} field of view push-broom imaging spectrometer that measures the solar radiation reflected by the Earth, at ground spatial resolution of 300 m (Full resolution) and 1200 m (Reduced resolution). MERIS is a 15 band programmable imaging spectrometer, in the visible and near infrared. MERIS allows a global coverage of the Earth in 3 days.

MERIS data is provided in three levels of processing: level 0, level 1 and level 2. Level 0 consists on the core information recorded in packets by the instrument. This information is not generally available to users and it serves as basis for level 1. Level 1 comprises geo-coded top of atmosphere (TOA) data radiances [$Wsr^{-1}m^{-2}\mu m^{-1}$] and it is the base for level 2. Level 2 provides reflectance values for the different kinds of data products. Level 2 reflectances are different in nature depending on the surface (MERIS-Envisat 2005):

- over clouds, they are TOA reflectances,
- over land, they are Top Of Aerosol (TOAr) reflectances corrected only for Rayleigh diffusion but not corrected for the diffusion by aerosols,
- over water, they are surface reflectances.

For our study we used the products over land of MERIS level 2. MERIS level 2 land surface products provide TOAr reflectance in 13 bands as band 11 (760 nm) and band 15 (900 nm) were excluded because they are strongly influenced by O_2 and water vapor in the atmosphere, respectively. The Rayleigh correction was calculated taking into consideration pixel optical thickness and geometry and it was derived from a look up table (Santer *et al.* 2000).

MERIS level 2 has addressed the correction by aerosols and angular perturbations in order to provide users with Top of Canopy (TOC) reflectance in two bands one in the red, band 8 (681 nm), and near infrared (NIR), band 13 (865 nm). Aerosol scattering information was obtained in the blue region namely band 2 (442 nm) where values are

much sensitive to atmospheric scattering than longer wavelengths. A parametric anisotropic function was implemented to simulate the variations in the signal of the aforementioned bands due to angular observation and illumination at TOAr. Then a combination of the blue band with the red and NIR is done to calculate the rectified red and rectified NIR bands that would have measured in the red and NIR at the TOC. The algorithm proposed to generate the rectified red and NIR is defined as:

$$g(B_1, B_2) = \frac{l_{0,1}B_1^2 + l_{0,2}B_2^2 + l_{0,3}B_1B_2 + l_{0,4}B_1 + l_{0,5}B_2 + l_{0,6}}{l_{0,7}B_1^2 + l_{0,8}B_2^2 + l_{0,9}B_1B_2 + l_{0,10}B_1 + l_{0,11}B_2 + l_{0,12}} \quad \textbf{(10)}$$

$$\rho_{rectified\_red} = g(\rho_{band2}, \rho_{band8}) \quad \textbf{(11)}$$

$$\rho_{rectified\_NIR} = g(\rho_{band2}, \rho_{band13}) \quad \textbf{(12)}$$

where ρ are reflectance values and $l_{0,n}$ are the coefficients for a polynomial provided in Gobron *et al.* (2004).

In addition, MERIS level 2 land surface product includes the MERIS Global Vegetation Index (MGVI) (Gobron *et al.* 2004) and the MERIS Terrestrial Chlorophyll Index (MTCI) (Dash and Curran 2004). The MGVI algorithm is generated from the rectified red and rectified NIR bands and it uses like the rectified bands Equation 10 for its calculation. The MGVI is given by:

$$MGVI = g(\rho_{rectified\_red}, \rho_{rectified\_NIR}) \quad \textbf{(13)}$$

where ρ are reflectance values for the rectified red and rectified NIR bands described above. The coefficients $l_{0,n}$ are given in Gobron *et al.* (2004).

The MTCI is defined as:

$$MTCI = \frac{\rho_{band10} - \rho_{band9}}{\rho_{band9} - \rho_{band8}} \quad \textbf{(14)}$$

where ρ are reflectance values for different bands.

For this study the full resolution MERIS level 2 reflectance values, MGVI and MTCI for the region of Noord-Holland from 2004 was used. Specifications of the spectral bands of MERIS Level 2 are given in Table 3. Cloud free images from April 1st, August 8th and September 10th were available and visually assessment of images from May 15th and July 14th showed less than 15% cloud cover. The images are shown in Figure 6.

Table 3. The bands of the MERIS level 2.

Band	Band center [nm]	Bandwidth [nm]
1	412.5	9.9
2	442.4	10.0
3	489.7	10.0
4	509.7	10.0
5	559.6	10.0
6	619.6	10.0
7	664.6	10.0
8	680.9	7.5
9	708.4	10.0
10	753.5	7.5
12	778.5	15.0
13	864.8	20.0
14	884.8	10
Rectified red	681	
Rectified NIR	865	

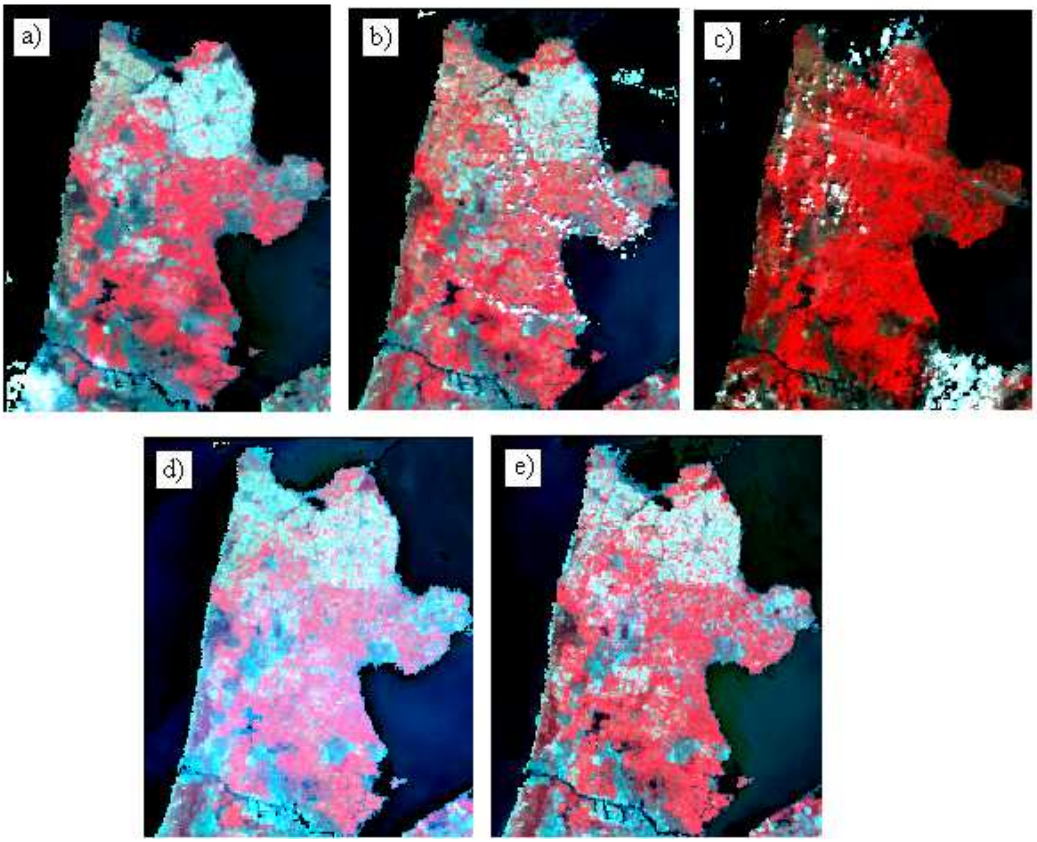


Figure 6. MERIS full resolution images of a) April 1st, b) May 15th , c) July 14th , d) August 8th and e) September 10th 2004. Bands 14, 8 and 3 are depicted in RGB.

3.2.3 Crop calendar

As crops show relative phenological differences throughout the growing season, a crop calendar depicting the growing season of the most important crops for the Netherlands was used for filtering the multi-temporal datasets (Figure 7). In this calendar the length of the growing season is determined by the phenological characteristics of the crops and agricultural practices such as: planting, fertilizing and harvesting

The available satellite scenes cover the planting and harvest dates for most of the crops. For example, sugar beet and potatoes would present a soil spectrum in the scenes of April 1st and May 15th because they are sowed at the end of May. Then they might present a vegetative spectrum for the scenes of July 14th and August 8th. In September, however, potatoes are harvested and therefore sugar beet would keep a vegetative spectrum for this

date. Cereals would present a vegetative spectrum for the scenes of April 1st and May 15th. Their vegetative spectrum would increase between those dates due to an augment in biomass. Nevertheless, for the scene of July 14th and August 8th they might show depletion in the vegetative spectrum because of senescence and further harvesting.

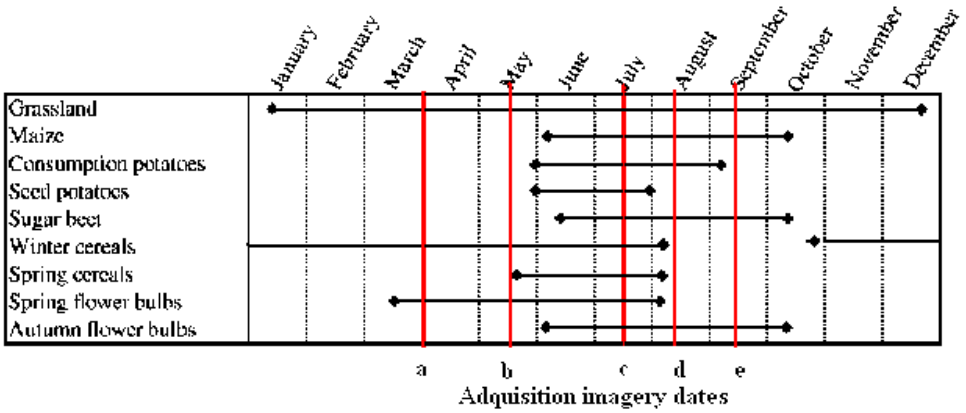


Figure 7. Crop calendar of The Netherlands. a) April 1st, b) May 15th , c) July 14th , d) August 8th and e) September 10th 2004 (modified from De Wit and Clever (2004).

3.3. METHODOLOGY

The methodology used in this research involves three main parts: preparation, processing and analysis (Figure 8). Preparation includes the selection of the study area from MERIS images and their co-registration with the resampled LGN5 dataset. Processing deals with the selection of the most suitable spectral bands from MERIS and the definition of the temporal profiles based on separability analysis. Also this section covers the preparation of the training, validation and testing datasets and the selection of the parameter values of the FFNNs. In the analysis part, the FFNNs are trained with the training and validation dataset. The trained ANN with the best performance is singled out for its simulation with the testing datasets. Statistical analyses were used to draw conclusions for the different estimated fractions at different scales of analysis.

3.3.1. Preparation

In the preparation stage the LGN5 database was resampled to 300 meters cell size assigning the majority class as label but keeping information on cover composition in fractional images per class as showed in Figure 5. This resampled database was used as reference for geometric corrections of MERIS images and selection of ground data.

In addition, the original LGN5 database 25 m cell size was used to analyze the landscape properties for the province of Noord-Holland. For this purpose we used the Patch Analysis 3.0 software (Rempel 2003) to estimate the class area, number of patches per class, mean patch size and standard deviation of patches of each cover type.

Before any analysis, the satellite images need some correction procedures due to the geometric and radiometric distortions during the acquisition process. These corrections can be divided in two categories: geometric and radiometric corrections.

Geometric corrections

Geometric corrections are necessary to reduce the effect of geometric distortions and they enable us to match the temporal datasets with the resampled LGN5 database. All MERIS level 2 images were geo-referenced with the BEAM 3.4 software (Brockmann 2005) using the latitude and longitude coordinates that are provided by MERIS metadada for nine tie points in the image. Then, the geographic projection was reprojected into the stereographic projection of the Dutch national coordinate system Rijks Driehoek (RD) using ENVI. Visual differences existed between the resampled LGN5 and satellite images, therefore we performed an image to image co-registration between each temporal dataset and the resampled LGN5. For each image to image co-registration 10 ground control points were recorded between the two images. A nearest neighbor resampling function was used because it preserves the information of the image pixels most closely.

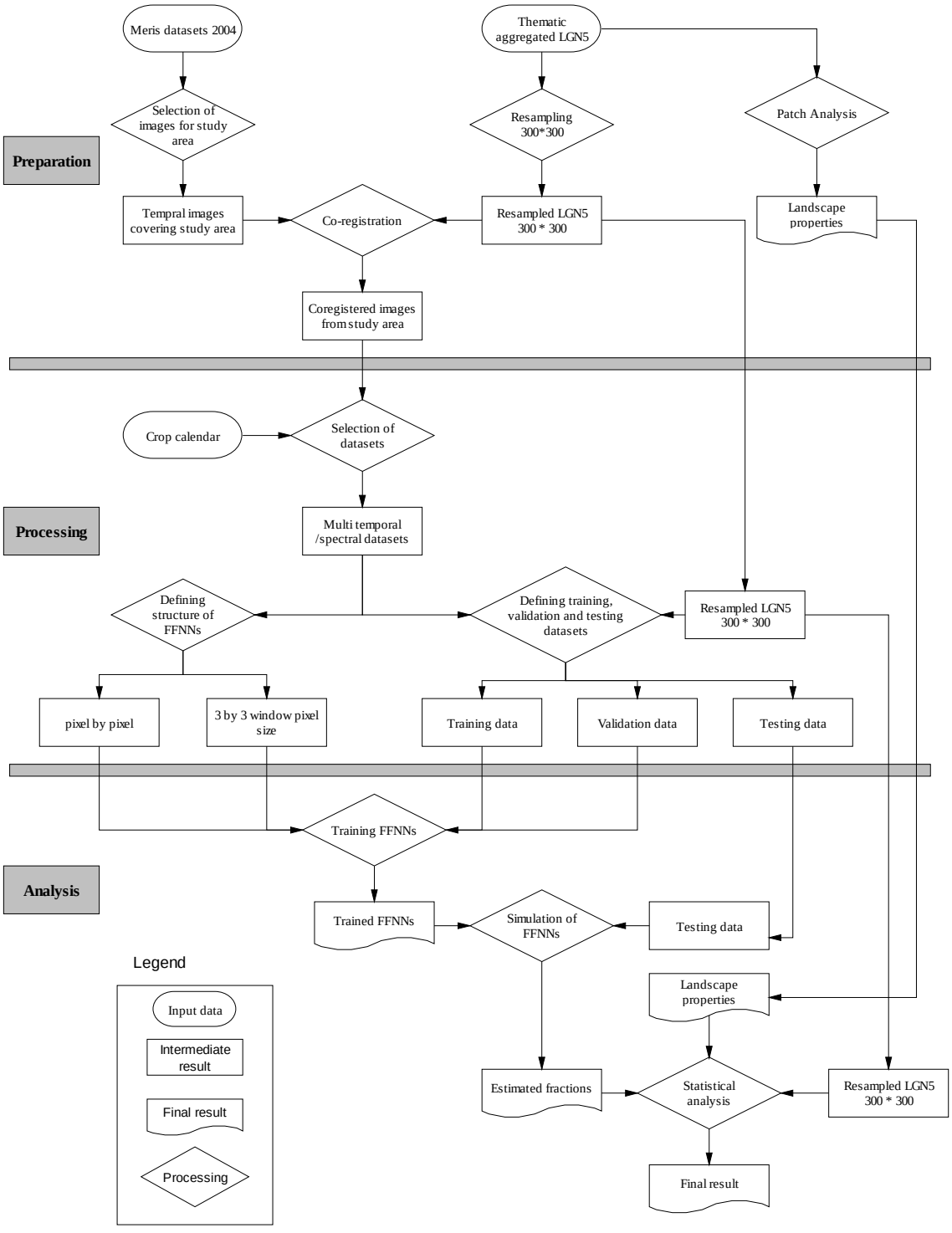


Figure 8. Flowchart of the methodology

Radiometric corrections

As we are using reflectance values from the MERIS level 2 products, radiometric corrections were already performed for incoming radiation and sun elevation.

3.3.2. Processing

This stage comprises the analysis of the spectral and temporal profiles, definition of the structure and parameters of the FFNNs, which are described separately in further sections, and the selection of the training, validation and testing data.

3.3.3. Analysis spectral and temporal profile

Individual spectral bands from MERIS level 2 were correlated to assess the degree of redundancy of the different bands. The MGVI that combines information from the rectified red and rectified NIR were used to generate various temporal profiles. Those temporal profiles allowed us to measure the degree of interclass separability based on the spectral response derived from the phenological state of the crops. The temporal profiles ranged from using one image to using all five images over the crop cycle (Table 4). For example, temporal profile 1 was chosen to evaluate the effect of class separability of sugarbeet and potatoes when only one image is available in the middle of the growing season. Temporal profile 2 evaluated the interclass separibility of sugar beet and potatoes when images were available at the onset, middle and end of their growing season. Temporal profile 3 allowed us to inspect the separability of cereals from other classes. Temporal profile 4 excluded the first scene, which might be considered out of the growing season and not redundant for crop separability between potatoes and sugar beet. Temporal profile 5 took into account all satellite datasets available for this study.

Table 4. Temporal profile set.

Temporal profile number	Image combination	Exclude images
1	c	a b d e
2	b c d	a e
3	a b d e	c
4	b c d e	a
5	a b c d e	

a) April 1st, b) May 15th, c) July 14th, d) August 8th and e) September 10th 2004.

The separability was measured with the Jeffries-Matsushita (JM) distance (Swain and King 1973). In similar studies, Murakami *et al.* (2001) and Kavzoglu and Mather (2002) agreed on the superiority of this distance for defining temporal scene combinations with high separability. The JM distance has an upper limit of 1.414 indicating high separability between classes. In this study a value lower than 1.3 for two classes was considered as having low separability and we decided therefore to merge those classes. The JM distance is represented by:

$$JM_{ij} = \sqrt{2(1-\exp(-\alpha))} \quad \textbf{(15)}$$

and

$$\alpha = \frac{1}{8}(\mu_i - \mu_j)^T \left(\frac{C_i + C_j}{2} \right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \left(\frac{\left| C_i + C_j \right| / 2}{\sqrt{\left| C_i \right| * \left| C_j \right|}} \right) \quad \textbf{(16)}$$

where *i* and *j* are the classes being compared, μ_i and μ_j are the mean vectors of classes *i* and *j*, C_i and C_j are the sample variance-covariance matrices for class *i* and *j*, $|C_i|$ and $|C_j|$ are the determinants of C_i and C_j, respectively, and T indicates the transposition,

3.3.4. FFNN structure

The simplest structure to be used in a neural network is reading one multi-spectral pixel into the network. Thus, one input node is used to represent each band (Paola and Schowengerdt 1995). However, a possible extension is to use spatial texture information, for example by using a 3 by 3 window for each band as input. Although this structure increases the input nodes by nine fold per node, it might enhance the classification performance by smoothing single pixel errors. On the other hand, this structure might cause a loss of information in heterogeneous landscapes. In this study we decided to use both structures to evaluate the effect of spatial information from MERIS on the sub-pixel accuracy (Paola and Schowengerdt 1997).

3.3.5. FFNN parameters

As mentioned before setting the parameters of neural networks is a difficult task and most of the times these parameters are based on trial and error approaches. In order to ease this process, we used the following guideline proposed by Kavzoglu and Mather (2003):

- *Estimate the number of hidden layer nodes required using the expression* $N_p/[r*(N_i+N_o)]$♣ *formulated by Garson (1998).*
- *Define the number of output layer nodes by considering the nature of the problem and the availability of ground reference data.*
- *Randomly select a number of training samples. Sample sizes should range between* $30*N_i*(N_i+1)$ *and* $60*N_i*(N_i+1)$♣ *as proposed by Hush (1989).*
- *Set initial weights to a small range with a mean value of 0. (e.g. [-0.15 , 0.15] or [-0.25, 0.25]).*
- *Employ a validation dataset to terminate the training process. The validation dataset may include around 50-100 samples for each class.*
- *Use the output encoding scheme of [0 1 0] to represent output classes.*
- *Use a shuffling mechanism for the learning process to present the inputs to the network in a randomly defined order.*

♣ N_i and N_o are the number of input and output nodes, respectively.

N_p is the number of training samples.

"r" is a constant that is related to the noise level of the data range from 5 for clean data to 10 for noisy data depending of the noise level of the data.

Although these guidelines pose a base for designing and using ANN in remote sensing image classification, they can be adjusted with other approaches found in the literature that might suit better this study. Hence the considerations taken in this study for the aforementioned recommendations and other parameters are further described in this section.

Number of input layer nodes

The number of input nodes corresponded to the temporal image combination and spectral bands or VIs per image for the pixel by pixel FFNN. For the 3 by 3 window FFNN the number of input nodes was nine fold the number of nodes used in the pixel by pixel structure.

Number of output layer nodes

The number of output layer nodes was determined after knowing the number of classes having good separability, as defined by the JM distance with values larger than 1.3 (section 3.3.3). Nevertheless, the level of complexity for the ANN to distinguish the desired classes depends on the number of input and output nodes. The complexity for the ANN would increase if the number output nodes is larger than the number of input nodes (Kavzoglu and Mather 2003). Therefore, it was appropriate to define ANN parameters such as number of hidden layer nodes and training samples from methods including the number of input nodes and output nodes for their estimation.

Number of hidden layers and hidden layer nodes

Estimating the number of hidden layers and hidden layer nodes is an important task in designing a neural network (Kavzoglu and Mather 2003). A single hidden layer is sufficient for classification of multispectral imagery and it is the most common used in literature (Paola and Schowengerdt 1995). However, when the number of output nodes gets near 20 more flexibility can be gained by a second hidden layer in the network (Kanellopoulos *et al.* 1992). Hence based on the number of classes a single hidden layer would be sufficient for this study.

The method recommended in the guideline to estimate the number of hidden layer nodes, formulated by Garson (1998) (Table 5), depends on the r constant that for this study we assigned a value of 10 due to the large amount of noise expected from classes with similar spectral profiles like potatoes and sugar beet. Thus, a large amount of training samples is needed to obtain a considerably amount hidden nodes to train the network. Nevertheless, as the amount of training data is a common problem to all supervised

algorithms (Borel and Gerstl 1994, Atkinson *et al.* 1997, Foody *et al.* 1997), we evaluated also other methods which do not include the number of training samples as parameter but use the number of input and output nodes for estimating the number of hidden layer nodes.

Table 5. Approaches proposed to compute the optimum number of hidden layer nodes.

Heuristic	Source
$2N_i$ or $3N_i$	(Kanellopoulos and Wilkinson 1997)
$(N_i + N_o)/2$	(Ripley 1993)
$N_p/[r(N_i + N_o)]$	(Garson 1998)
$\frac{2 + N_o * N_i + \frac{1}{2} N_o (N_i^2 + N_i) - 3}{N_i + N_o}$	(Paola 1994)

N_i and N_o are the number of input and output nodes, respectively.

N_p is the number of training samples.

r = 10. Due to the amount of noise expected from classes with similar spectral profiles like potatoes and sugar beet.

Number of training samples

The number of training samples has been acquainted by various authors as the parameter of major influence at the training stage. This is because the neural network learns the characteristics of the data from the sample values. Too few or too large training samples may lead, respectively, to insufficient training or overfitting of the network. In order to determine the appropriate number of training samples we used two approaches, the first one using the method presented in the guideline that is suggested by Hush (1989) as:

$$N_p = 30 * N_i * (N_i + 1) \quad \textbf{(17)}$$

and the second one using the method proposed by Mather (1999), which takes into consideration input and output data, is estimated as:

$$N_p = 30 * N_i * N_o \quad \textbf{(18)}$$

where N_p is the number of training samples, N_i is the number of input nodes and N_o is the number of output nodes in the network.

Training, validation and testing data

For training and simulation of the FFNNs the dataset was divided into training, validation and testing data. The networks were trained with the training and validation dataset using multi-temporal profiles coming from ρ and VIs in the input layer and the resampled LGN5 data with the known land cover fractions as target dataset in the output layer. For simulating the trained networks, the multi-temporal profiles of the testing data were used in the input layer and the resampled LGN5 data were used to validate the estimated fractions coming from the trained FFNNs.

The selection of the training, validation and testing dataset was done with a function developed in MATLAB 7.0. This function randomly selects a desired number of pixels from each class from the multi-temporal profiles and the resampled LGN5, while excluding the pixels contaminated with clouds. The number of training samples per class was equally distributed from the total training samples estimated with Equations 17 and 18. The number of testing and validation samples included 50 samples per class as suggested in the guideline proposed by Kavzoglu and Mather (2003).

Initial weight ranges

Initialization of random weight values with a mean value of zero in the range of [-0.15, 0.15] or [-0.25, 0.25] has produced similar error rates and achieved higher values of classification accuracy than larger ranges [-0.6, 0.6] (Kavzoglu and Mather 2003). Therefore a lower range of [-0.2, 0.2] was used for this study.

Learning rate

The learning rate was not implemented in this study because we used as training algorithm the resilient back propagation. This algorithm is commonly used and recommended along with log sigmoid functions (Demuth *et al.* 2005). The resilient back

propagation algorithm updates the weight values based on the derivative of the performance function with respect to the weight value and not to a learning rate and momentum as used by other approaches like the standard steepest decent algorithm. A complete description of the resilient back propagation algorithm is given by Demuth *et al.* (2005).

Output fractions

The output fractions in this study used one output node per class with values ranging between 0 and 1. The value 1 was assigned when the node fully corresponded to the selected class, whereas 0 to the nodes that corresponded to other classes. Values in between indicated the proportions for the different land cover types. For example, if there are five classes then a code of 0 1 0 0 0 was used to represent a pure node for the class 2.

Stopping criteria

The stopping criteria method called cross validation was used in this study to prevent overfitting while training the neural network. This method uses two sets: a training set and a validation set. The training set is utilized for computing the error gradient and updating the network weights and biases. The validation subset is used to monitor the error on the validation set during the training process. The validation set normally follows the decreasing trend of the training error; however, when there is an overfitting of the training set, the validation error increases. When an increase of the validation error iteratively occurs, the training is stopping and the weights and biases at the minimum of the validation error are returned. The cross validation method is implemented in the neural network toolbox of MATLAB 7.0.

3.4. ANALYSIS

The selection of the most appropriate parameter values for the number of input nodes and the number of hidden layers was based on an iterative approach where the response to each parameter was analyzed independently. Since one parameter value was evaluated each time, the rest of the parameters had to be fixed when training the network. The FFNNs were trained using the MATLAB Neural Network toolbox. The performance for each parameter was analyzed from the Mean Square Error (MSE) on the training. The parameter values having the lowest MSE were chosen to train the FFNNs. The trained FFNNs were then simulated with the testing dataset and their estimations were thoroughly evaluated with statistical analysis.

3.4.1 Simulations

For the simulations we used the simulation function from the Neural Network toolbox along with the trained FFNNs and the testing dataset. The relationship between network response and known fractions per class was evaluated using linear regressions (Equation 19) and calculating the coefficient of correlation (R) (Equation 21) between the network response and the corresponding targets.

$$y = bx + c \qquad \textbf{(19)}$$

The slope, b and constant c are given by:

$$b = R\left(\frac{\sigma_{\hat{a}}}{\sigma_a}\right) \qquad c = m_{\hat{a}} - b * m_a \qquad \textbf{(20)}$$

where the slope b, is R multiplied by the ratio of the standard deviations, with $\sigma_{\hat{a}}$ being the standard deviation of the estimated fractions per pixel and σ_a the standard deviation of the known fractions per pixel for class *a*. The constant c can be calculated using the means of the estimated $m_{\hat{a}}$ and known m_a fractions.

The coefficient of correlation R is represented by:

$$R = \frac{\mathrm{cov}(a, \hat{a})}{\sigma_a \sigma_{\hat{a}}} \quad \textbf{(21)}$$

Where $\mathrm{cov}(a, \hat{a})$ is the covariance between known and estimated values, σ_a and $\sigma_{\hat{a}}$ are the standard deviations of known and estimated values for class *a*.

3.4.2 Analysis of estimations at different scale of sub-pixel information

To evaluate the impact of the estimated fractions at different scales, we selected two regions of 900 ha (10 by 10 MERIS pixels) and considered spatial scales ranging from a single MERIS pixel (9 ha) to the whole region. Figure 9 depicts the aggregation of the sub-pixel information at different scales. the spatial scales used for the analysis and the process of aggregation. At each scale, the total area of class *a* was calculated for the estimated and known values. Thus, at a scale of 9 ha the region was divided into 10 by 10 pixel square sub-regions and the total area of class *a* was computed for each sub-region. Then, the estimated and known class *a* areas were analyzed at each scale using the Root Mean Square Error ($RMSE$) (Equation 22), the normalized RMSE ($RMSE_{normalized}$) (Equation 23), Bias (Equation 24) and R (Equation 21).

$$RMSE_s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{a}_i - a_i)^2} \quad \textbf{(22)}$$

$$RMSE_{normalized} = \frac{RMSE_s}{\sum_{i=1}^{n} a_i \Big/ n} \quad \textbf{(23)}$$

$$Bias = \frac{\sum_{i=1}^{n}(\hat{a}_i - a_i)}{n} \quad \textbf{(24)}$$

where, $RMSE_s$ is the RMSE proportional to the spatial scale s (ground area) over which class a area is computed, $RMSE_{normalized}$ is the $RMSE_s$ normalized for the known class a area, n is the number of sub-regions and a_i and $\hat{a}_i$ are the known and estimated areas of class a in the i th sub-region, respectively. The RMSE informs about the inaccuracy of the prediction, the R informs about the prediction variance and the Bias defines the systematic error of the estimations related to the known areas.

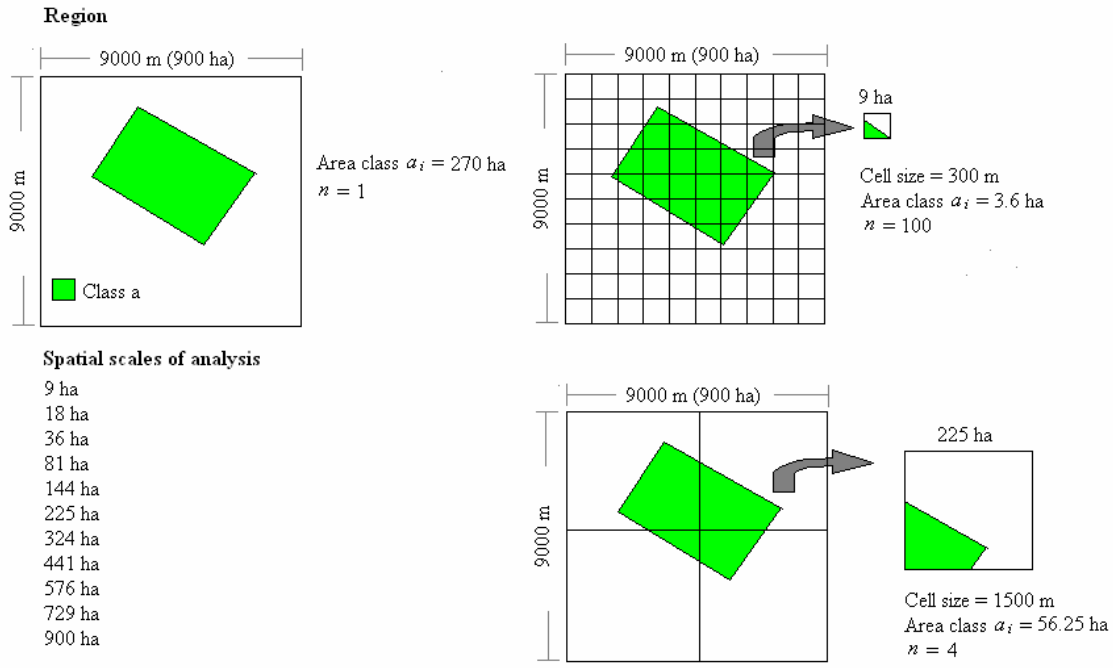


Figure 9. Spatial scales of analysis and aggregation of sub-pixel information.

4. RESULTS AND DISCUSSION

4.1. LANDSCAPE PROPERTIES

Sub-pixel estimates of land cover proportions would depend on landscape properties such as patch size, patch variability and land cover proportions (Moody and Woodcock 1994, Huang *et al.* 2002, Lobell and Asner 2004). This is because the percentage of correctly classified pixels decreases with the degree of heterogeneity within a pixel (Moody *et al.* 1996).

Table 6 presents the patch size, patch variability and land cover proportions of each class for the study area. All crop classes except grassland comprised very little proportions of the study area. Nevertheless their mean patch size (~ 10 ha) was as large as the spatial resolution covered by a MERIS pixel (9 ha). Sugar beet, potatoes, cereals and forest had smaller patches sizes than the mean area depicted by the other classes. Nevertheless, all land cover types had a greatly varying patch size, which describes the high heterogeneity of the landscape. This great variation in the patch size of crops could have been expected from the information given in Table 2, where crop fields vary from small plots with less than 10 ha to crop fields of more than 30 ha (i.e. grassland or potatoes). Nevertheless, visual assessment of the spatial distributions of grassland, built up, natural vegetation and water patches indicated that these patches are more tied to its same class than patches of sugar beet, potatoes and cereals. Therefore, the former classes are more likely to be surrounded by its same class than by any other class when the landscape is aggregated to 300 m.

Table 6. Patch metrics for the study area based on LGN 5 dataset.

Class	Class area [ha]	Number of patches	Mean patch size [ha]	Standard deviation of patches [ha]
Built up	78771	14863	5.30	544.44
Grassland	135263	13967	9.68	85.77
Potatoes	14811	1409	10.51	14.00
Cereals	12612	1071	11.78	14.02
Other crops	41291	3313	12.46	42.29
Forest	21788	4976	4.38	40.29
Nat. veg.	26372	2754	9.58	137.59
Sugar beet	8468	906	9.35	8.44
Water	266806	2471	107.97	2914.21

4.2. IMAGE GEOMETRIC CORRECTIONS

Assuming that the resampled LGN5 with a 300 m pixel size had a geometric inaccuracy in the order of half a MERIS pixel (Clevers *et al.* 2005). The shift of the image to image co-registration using 10 ground control points recorded between the resampled LGN5 and reprojected MERIS images appeared to be systematic for all MERIS datasets. It was in the order of half a pixel as shown in Table 7.

Table 7. RMSE [in pixels] of the georeferencing to the Dutch RD coordinate system.

Recording date	RMSE X coord.	RMSE Y coord.	RMSE total
01-04-2004	0.413	0.214	0.466
15-05-2004	0.365	0.270	0.454
17-07-2004	0.276	0.380	0.470
08-08-2004	0.308	0.352	0.468
10-09-2004	0.360	0.281	0.458

4.3. SPECTRAL AND TEMPORAL PROFILES

In order to guarantee the collection of spectral signature over pure pixels, 10 pure pixels per class that were surrounded or adjacent by pixels of the same class were selected from the dataset. Figure 10 and Annex I present the mean spectral signatures for each land cover type. Although these signatures showed severe overlap, the land cover spectra were as expected for all classes. Annual crop types showed a flat reflectance signature at the onset and end of the growing season. Instead, grassland depicted a clear vegetation spectrum with a high reflectance in the near infrared (NIR) for all the dates. Forest and natural vegetation had similar spectral signature over time but the signature of forest for April 01st (Figure 10) was lower in the NIR because leaf unfolding of deciduous forests and understory forest species for 2004 occurred at the end of April. Instead, natural vegetation areas comprised evergreen shrubs and grassland species that presented visible leaf surface at the end of March (Van Vliet 2005). Built up areas showed a similar low NIR spectrum over dates, which can be explained by the effect of green areas such as gardens and parks allocated within the built up areas. The spectra of water in the visible and near infrared showed the lowest value and it can clearly be distinguished from other spectra. Hence, this class can be expected to have good classification accuracy with only one image.

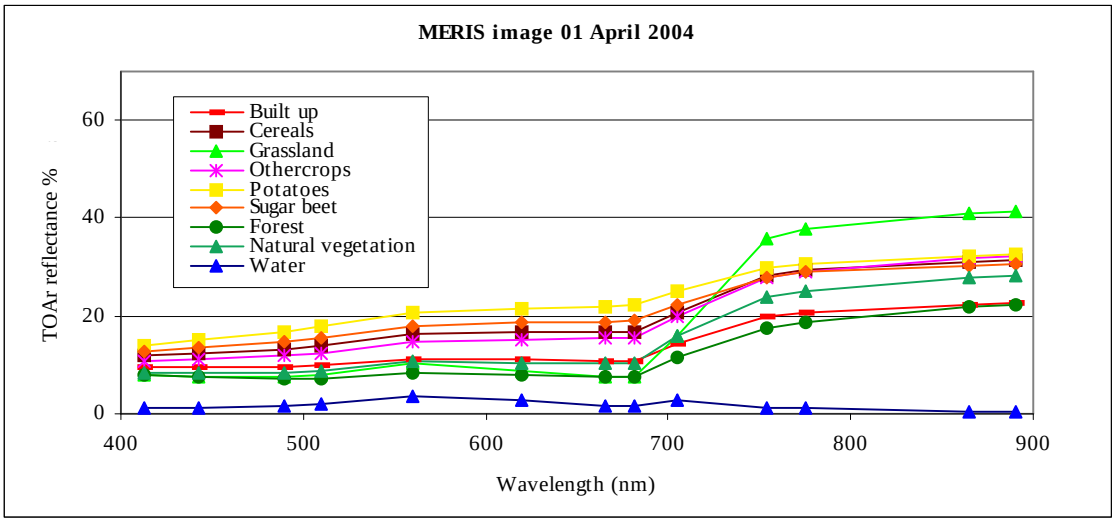


Figure 10. Spectral signatures for land cover types derived from MERIS level 2 image.

The temporal visualization of the nine land cover types was facilitated by plotting the MGVI profiles over time (Figure 11). Cereals, potatoes and other crops depicted a complete crop cycle. Their profile steady increased due to an aunment in biomass over time in May – June, reaching peak greenness in June-July, and then a negative trend in July – September due to senescence and harvest. Sugar beet showed a similar pattern but it was harvested in November. Grassland did not have major changes over time; only a smooth depletion in August indicated a reduction in biomass possibly due to the mowing cycle and grazing. Forest and natural vegetation had similar trends over time. Their profile increased due to aument in biomass between April and May reaching greenees peak in May. Then their profile kept constant over time. Built up and water showed a flat profiles over time but water had the lowest profile due to the low spectral values in the red and NIR.

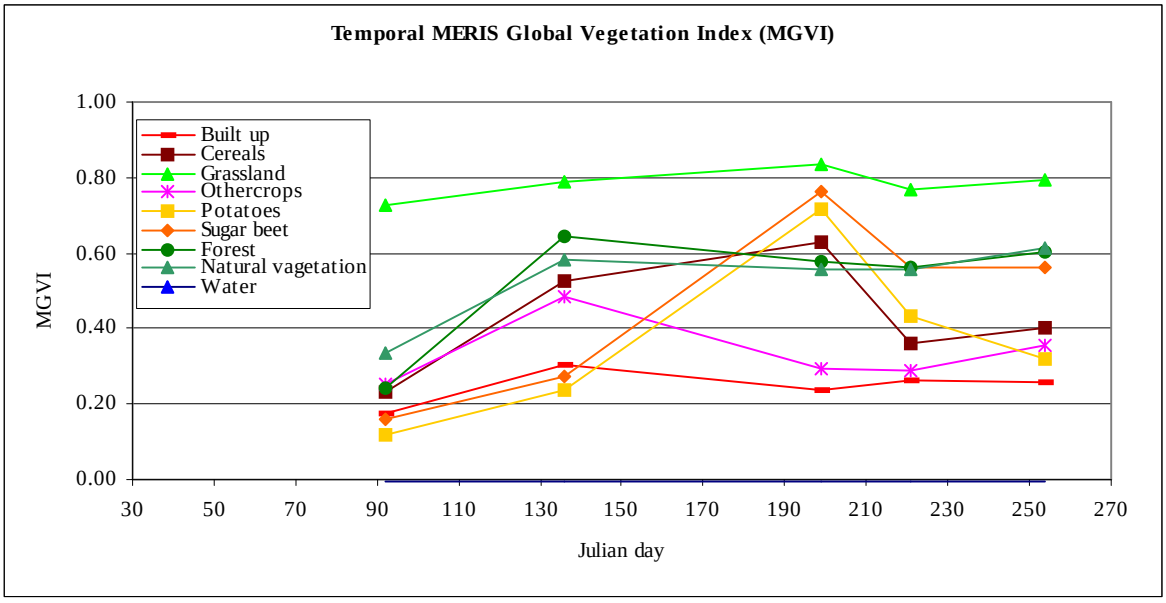


Figure 11. Temporal MERIS Global Vegetation Index (MGVI) spectra for image dates.

The spectral profiles for the different land cover types also illustrated the high degree of correlation for bands in the visible (400 – 700 nm) and NIR (750 – 900 nm) wavelength. Therefore, a correlation matrix between the spectral bands was calculated and its results are presented in Table 8. The visible spectra (bands 1 till 8) and the NIR (bands 10 till 14) were very correlated. The band 9, designed for the red edge region, showed a

moderate high correlation with the visible bands. It was interesting that the correlation of the rectified red with band 8 and rectified NIR with band 13 was not 1. This can be associated with the correction carried out to produce spectral reflectances at TOC. Nevertheless, the correlation for the rectified red and band 8 was lower than the correlation for the rectified NIR and band 13. This difference between the two correlations can be expected because lower wavelengths are more sensitive to atmospheric scattering than larger wavelengths (Gobron *et al.* 2004).

Table 8. Correlation matrix for the MERIS image of April 1st, 2004.

Band	1	2	3	4	5	6	7	8	9	10	12	13	14	Rec. Red	Rec. NIR	MGVI	MTCI
[nm]	412	442	490	510	560	620	665	681	705	754	775	865	890	681	865		
2	0.99	1															
3	0.98	0.99	1														
4	0.98	0.99	1	1													
5	0.93	0.95	0.96	0.98	1												
6	0.96	0.96	0.98	0.99	0.98	1											
7	0.96	0.95	0.98	0.98	0.95	1	1										
8	0.96	0.95	0.98	0.98	0.95	0.99	1	1									
9	0.70	0.73	0.76	0.79	0.89	0.83	0.79	0.79	1								
10	0.30	0.30	0.31	0.27	0.10	0.29	0.35	0.35	0.20	1							
12	0.32	0.32	0.32	0.29	0.12	0.31	0.36	0.37	0.18	1	1						
13	0.35	0.35	0.35	0.31	0.14	0.32	0.38	0.38	0.18	0.99	1	1					
14	0.35	0.35	0.35	0.32	0.14	0.32	0.38	0.38	0.18	0.99	0.99	1	1				
Rec. Red	0.84	0.88	0.91	0.92	0.90	0.95	0.96	0.96	0.79	0.30	0.32	0.33	0.33	1			
Rec. NIR	0.28	0.29	0.28	0.25	0.08	0.26	0.32	0.32	0.24	0.98	0.98	0.99	0.99	0.24	1		
MGVI	0.70	0.68	0.68	0.67	0.55	0.69	0.72	0.73	0.10	0.95	0.95	0.96	0.96	0.88	0.95	1	
MTCI	0.59	0.59	0.60	0.60	0.44	0.66	0.68	0.67	0.21	0.83	0.84	0.83	0.83	0.68	0.83	0.80	1

(0.30) = negative correlation

The MGVI and MTCI had a strong, positive correlation between them and with bands located in the NIR while displayed a negative correlation with the bands in the visible region. Nevertheless, the MGVI showed a stronger correlation with NIR reflectances than the MTCI.

Despite the fact that this correlation matrix was calculated over grassland areas for the image of April 1st, the MTCI showed a weaker negative correlation to red reflectances than the MGVI. This unexpected response of the MTCI, which was designed to be chlorophyll sensitive (Dash and Curran 2004), can be attributable to the fact that the bands 8, 9 and 10 used to calculate the MTCI did not consider aerosol corrections and biodirectional reflectance factor issues. Another reason could have been the there were not much photosynthetic activity over grassland for this image.

The principal component analysis for the MERIS image of April 1st showed that most of the information (99.78%) was captured in the first three components (Annex II). The first component had high positive loadings from the MGVI, MTCI, and NIR bands. The second component had high positive loadings from the NIR bands and the third component included loadings from all bands in the visible region. Hence, most of the information of a MERIS level 2 image could be comprised in two bands: one in the red like the rectified red and other in the NIR like the rectified NIR. In this study we found it appropriate not to limit the information from temporal MERIS images to the combination of two bands but also to explore the combination of MGVI and MTCI MERIS level 2 products.

4.4. CLASS SEPARABILITY FROM TEMPORAL PROFILES

The class separability was measured with the JM distance (Equation 15) using endmembers defined from 10 pixels per class known to consist of a single cover type. The temporal profiles were designed with the MGVI, which combined information from the rectified red and rectified NIR bands. The class water was excluded from the separability analysis calcualted with the JM distance because visual comparisons of its spectra in the red and NIR regions showed very good separability over time. Table 9 shows that the separability between classes increased as temporal datasets were added up. A temporal profile coming from five dates was optimal to achieve good separability between crop types. Only the class forest and natural vegetation presented a distance lower than 1.3, which indicated low separability in this study. Therefore these two classes

were merged into a natural vegetation class. Finally eight classes comprising built up, grassland, potatoes, cereals, other crops, natural vegetation, sugar beet and water were used as target in the output nodes. The JM distance values between classes for all temporal profiles are given in Annex III.

Although Murakami *et al.* (2001) stated that good separabilities between crops could be achieved with at least three images distributed throughout the growing season (i.e. May, July and August), we found that for our study it was not the case. This can be expected because the temporal and spectral patterns showed similar behavior between class types except for water. Another explanation for this could be that the aforementioned authors used a large set of temporal fine resolution datasets from SPOT HVR allowing them to extract a well defined set of endmembers.

Table 9. Classes with low separability (JM distance < 1.3) per scene combination

Temporal profile 1	Temporal profile 2	Temporal profile 3	Temporal profile 4	Temporal profile 5
Cereals & Potatoes	Potatoes & Sugar beet	Forest & Nat. veg.	Forest & Nat. veg.	Forest & Nat. veg.
Cereals & Sugar beet	Built up & Other crops	Potatoes & Sugar beet	Potatoes & Sugar beet	
Potatoes & Sugar beet	Forest & Nat. veg.	Cereals & Other crops	Cereals & Potatoes	
Built up & Other crops	Cereals & Potatoes	Cereals & Potatoes	Built up & Other crops	
Nat. veg. & Potatoes	Cereals & Forest	Built up & Other crops		
Cereals & Forest	Forest & Other crops	Other crops & Potatoes		
Cereal & Nat. veg.	Forest & Grassland	Nat. veg. & Potatoes		
Forest & Sugar beet	Nat. veg. & Other crops			
Nat. veg. & Sugar beet				
Grassland & Potatoes				
Cereals & Grasland				
Forest & Potatoes				
Grassland & Sugar beet				
Grassland & Nat. veg.				
Forest & Grassland				
Nat. veg. Other crops				
Built up & Nat. veg.				
Forest & Nat. veg.				

4.5. SELECTION OF FFNN PARAMETERS

The majority of the parameters for the FFNNs were based on results of previous sections and literature review. These parameters are shown in Table 10. The number of input layer nodes corresponded to the number of chosen bands per image determined in section 4.3 and the number of temporal images giving the largest separability of crops found in section 4.4. Hence, 10 input layer nodes were assigned to the pixel by pixel FFNN and 90 input layer nodes were allocated to the window 3 by 3 FFNN. The number of output layer nodes makes reference to the eight land cover types defined with the JM distance in section 4.4. We used a single hidden layer because it gave enough flexibility for the eight land cover classes to be distinguished in this assignment (Kanellopoulos *et al.* 1992). The amount of validation and testing samples were arbitrarily based on the guideline proposed by Kavzoglu and Mather (2003) with 50 samples per class. A log sigmoid transfer function was motivated for the hidden and output layers because it allows the outputs to be interpreted as probabilities ranging from 0 to 1. Since these outputs approximate to probabilities they sum to 1, representing each output node the class composition within a pixel (Bishop 1995). When using this function, it is recommended to adopt a resilient back propagation algorithm because this algorithm updates the weight values based on the derivate of the sigmoid function and not with a learning rate or momentum. Besides, this algorithm has proven to be fast and efficient in memory when used with a log sigmoid function (Demuth *et al.* 2005). The weight initialization values ranged from -0.2 to 0.2 as recommended by Kavzoglu and Mather (2003). The cross validation stopping criteria would stop the training when one of the following situations occurs:

- iterations exceed the limit of 5000
- Mean Square Error (MSE) achieved is lower than 0.01
- MSE error of the validation dataset increases after 20 consecutive iterations.

However, Table 10 does not include all necessary parameters to run the FFNNs. We still need to evaluate the type of input data for training the networks, number of hidden layer nodes and number of training samples. These parameters, though well documented in

literature, are case specific and therefore they were discriminated after evaluating MSE coming from trained networks. The networks were run ten times in order to avoid inconsistency for the MSE values and their average was used to define the performance of the network. Each parameter is analyzed independently in the following subsections.

Table 10. Constant parameter values for training the neural network.

Parameter	Value
Number of input layer nodes	10 for the pixel by pixel FFNN 90 for the 3 by 3 window FFNN
Number of output nodes	8 nodes
Number of hidden layers	1 layer
Number of validation sample	50 per class
Number of testing sample	50 per class
Transfer function	Log sigmoid
Training algorithm	Resilient backpropagation
Initial weight ranges	-0.2, 0.2
Stop criteria	Cross validation or early stopping
Iterations	5000
Goal	0.01 MSE
Number of consecutive failure iterations	20

4.5.1. Input data

In order to compare the performance of the FFNN with vegetation indices (VIs), using the MTCI and MGVI, and reflectance values (ρ), with the rectified red and rectified NIR, we iteratively trained the FFNNs 10 times with the parameters described in Table 10 along with the arbitrary parameters given in Table 11. The number of hidden layer nodes were estimated according to Ripley (1993) for both networks because it includes as parameters the number of input nodes and output nodes and it does not rely on the selection of constant values for its calculation (Table 5). The training sample used for both networks was estimated according to Mather (1999) (Equation 18) because it also includes as parameters the number of input nodes and output nodes. Nevertheless the calculated amount of training samples for a 3 by 3 window network was too large for this study (refer to this issue in section 4.5.3). Therefore, we used the same amount of training samples calculated for the pixel by pixel network structure.

Table 11. Parameters used for evaluating different input data

Pixel by pixel network	3 by 3 window network
9 hidden layer nodes	49 hidden layer nodes
2400 training samples	2400 training samples

Figure 12 depicts the average MSEs and standard deviations for trained networks with VIs and ρ values. The pixel by pixel network and the 3 by 3 window network had lower average MSEs when trained with ρ values than VIs. Also their standard deviation was slight smaller for the trained networks with ρ values than VIs. A similar pattern was obtained for the training iterations of the networks trained with ρ. Annex IV presents the iteration values obtained for each FFNN. It seemed that the sigmoid transfer function behaved better with input datasets ranging between 0 and 1 in the case of ρ than larger than 1 from the MTCI. In general, the trained 3 by 3 window FFNN reached lower MSE than the pixel by pixel FFNN.

An alternative to use the MTCI as input dataset could have been its normalization to a range between 0 and 1. In this study, this preprocessing step was not done and we continued training the networks with ρ due to its better performance.

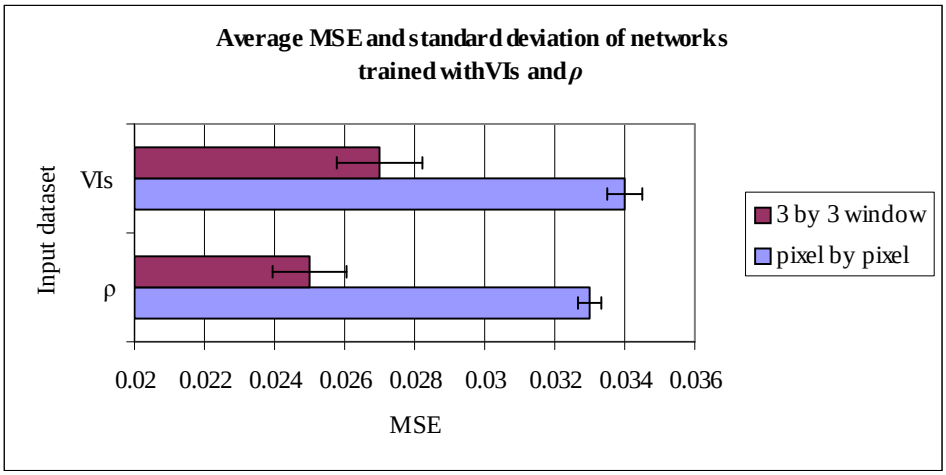


Figure 12. Average MSE and standard deviation response for trained FFNN with VIs and ρ.

4.5.2. Number of hidden layer nodes

The number of hidden layer nodes was estimated using the heuristic proposed by Ripley (1993), Paola (1994), Kanellopoulos and Wilkinson (1997) and Garson (1998) described in Table 5. The number of hidden layer nodes for pixel by pixel FFNN were 9, 30, 30, and 13, respectively. Estimates for the 3 by 3 window FFNN were 49, 341, 180 and 2, respectively.

Similar to the approach followed to determine the appropriate input data, to determine the number of hidden layer nodes we trained the networks with the parameters described in Table 10 along with the parameters presented in Table 12.

Table 12. Parameters used for evaluating different hidden layer nodes

Pixel by pixel network	3 by 3 window network
2400 training samples ρ input data	2400 training samples ρ input data

Figure 13 illustrates the average MSEs and standard deviations for trained networks using the hidden layer nodes estimated with the heuristics presented in Table 5. For the pixel by pixel FFNN there were no great differences of the average MSEs among the number of hidden nodes nor the standard deviations. However, the method suggested by Garson (1998) with 13 hidden layer nodes and estimated with an r value of 10 performed slightly better.

For the 3 by 3 window FFNN, the method suggested by Garson (1998) with 2 hidden layer nodes gave a larger average MSE than other methods because the number of hidden layer nodes was too small to discriminate between classes. This small number resulted from the increase of input nodes in the denominator and the limited number of training samples in the numerator. Instead, the methods proposed by Ripley (1993), Paola (1994) and Kanellopoulos and Wilkinson (1997) with 49, 341 and 180 hidden layer nodes, respectively, had equal average MSEs and similar standard deviations.

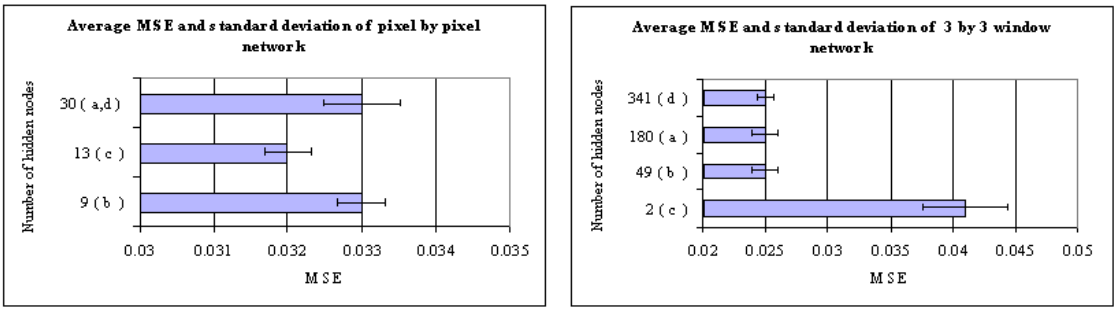


Figure 13. Average MSE and standard deviation response for trained FFNNs with different numbers of hidden nodes. Number of hidden layer nodes calculated from: a) Kanellopoulos and Wilkinson (1997), b) Ripley (1993), c) Garson (1998) and d) Paola (1994).

This indicated that estimating the exact number of hidden layer nodes might require the use of different methods to define the optimum bounds for a desired network response. Nevertheless, the final selection of the number should be towards the lower bound as it would be more efficient in saving memory, training would be iteratively faster and the network would have less probabilities of memorizing the training data (Demuth *et al.* 2005).

For this study we decided to select 13 hidden layer nodes proposed by Garson (1998) for the pixel by pixel network and 49 hidden layer nodes proposed by Ripley (1993) for the 3 by 3 window network.

In general, the 3 by 3 window FFNN had lower average MSE and standard deviation than the pixel by pixel FFNN, except for the 3 by 3 window network trained with 2 hidden layer nodes. The training iterations decreased with an increase of the hidden layer nodes. Annex V presents the iteration values for each FFNN.

4.5.3. Number of training samples

The number of training samples was calculated using the methods proposed by Hush (1989) in Equation 17, and Mather (1999) in Equation 18. Estimates for the pixel by pixel FFNN were 3300 for Equation 17 and 2400 for Equation 18.

Estimations for the 3 by 3 window FFNN were 245700 for Equation 17 and 21600 for Equation 18. These amounts of training samples were inappropriate because they require a larger dataset than the one used in this study. Instead, we used the number of training samples calculated for the pixel by pixel network. It is important to stress that the performance of a network with a large training sample may not differ much from that of a much smaller set when the selection of the training data is randomly and equally distributed per class (Paola and Schowengerdt 1995, Kavzoglu and Mather 2003).

Similar to previous parameters, we trained the FFNNs with the parameters described in Table 10 along with the most suitable parameters found in section 4.5.1 and section 4.5.2, both summarized in Table 13.

Table 13. Parameters used for evaluating different training samples.

Pixel by pixel network	Window 3 by 3 network
13 hidden layer nodes	49 hidden layer nodes
ρ input data profile	ρ input data profile

There were minor differences between the average MSEs of trained networks with different training datasets (Figure 14). The 3 by 3 window FFNN trained with a dataset of 3300 pixels had a lower average MSE than the 3 by 3 window FFNN trained with 2400 pixels. Nevertheless, standard deviations of the MSEs were larger for the dataset of 3300 pixels.

The pixel by pixel FFNNs had similar MSEs and standard deviations for both datasets. Hence the difference in the number of training dataset in this study did not have major influence for training the FFNNs. Also similar as in previous sections, a 3 by 3 window FFNN had lower average MSE than a pixel by pixel FFNN.

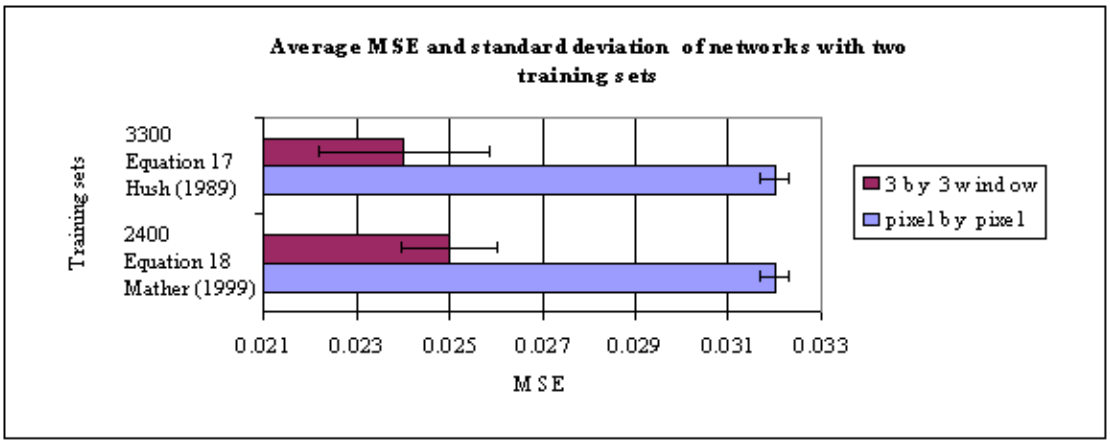


Figure 14. Average MSE and standard deviation response for trained FFNNs with two training sets.

4.5.4. Optimum parameter values

After evaluating several parameters in previous sections the optimum parameter values for training the pixel by pixel and window 3 by 3 FFNNs for this study are displayed in Table 14.

Table 14. Optimum setting of network structures

Parameters	Pixel by pixel FFNN	3 by 3 window FFNN
Input data profile	ρ	ρ
Number of input layer nodes	10	90
Number of hidden layer nodes	13	49
Number of hidden layers	1	1
Number of output layer nodes	8	8
Number of training samples	2400	2400
Number of validation samples	400	400
Number of testing samples	400	400
Initial weight range	[-0.2, 0.2]	[-0.2, 0.2]
Stopping criteria	Cross validation	Cross validation
Iterations	5000	5000
Goal	0.001 MSE	0.001 MSE
Number of consecutive failure iterations	20	20
Transfer function	Log sigmoid	Log sigmoid
Training algorithm	Resilient backpropagation	Resilient backpropagation

4.6. SIMULATIONS

Although the performance of the trained networks can be evaluated to some extent by the MSE of the training set, we decided to evaluate the network response with data that has not been presented before (testing data). For this purpose, we used a trained network with parameters provided in Table 14 and simulated its response with the prepared testing data.

Figures 15 shows that the estimated fractions for the pixel by pixel network did not have a perfect fit with the known fractions for all cover types (slope equal to one). The estimated fractions for built up, natural vegetation, water and grassland classes had a strong positive correlation with correlation coefficient values (R) of 0.71, 0.70, 0.87 and 0.77, respectively. Potatoes and sugar beet estimated fractions had a positive but moderate correlation with R values of 0.57 and 0.55, respectively. The estimated fractions for cereals and other crops classes were estimated, with positive but low correlation, with R values of 0.45 and 0.48, respectively. Nevertheless, the outliers had a great impact for the moderate and low correlation values. The scatter plots showed that the outliers had a great impact for the moderate and low correlation values

For most classes the data points were clustered towards the lower left corner of the plot (indicating the large amount of low fractions per class). There were overestimations for very low fractions and underestimations for high fractions. This response can be explained by the behavior of the log sigmoid transfer function, as it tends to saturate with high values and overemphasize with low fractions.

It was interesting that saturation for high values was different for the land cover types. For example, the saturation for the simulated fractions for potatoes was 0.4, whereas for grassland it reached 0.9. This can be explained by the fact that most of the fractions of datasets used for training potatoes ranged from 0 to 0.6, whereas for grassland they were well distributed between 0 and 1. Information on the fraction distribution of the training dataset is presented in Annex VII.

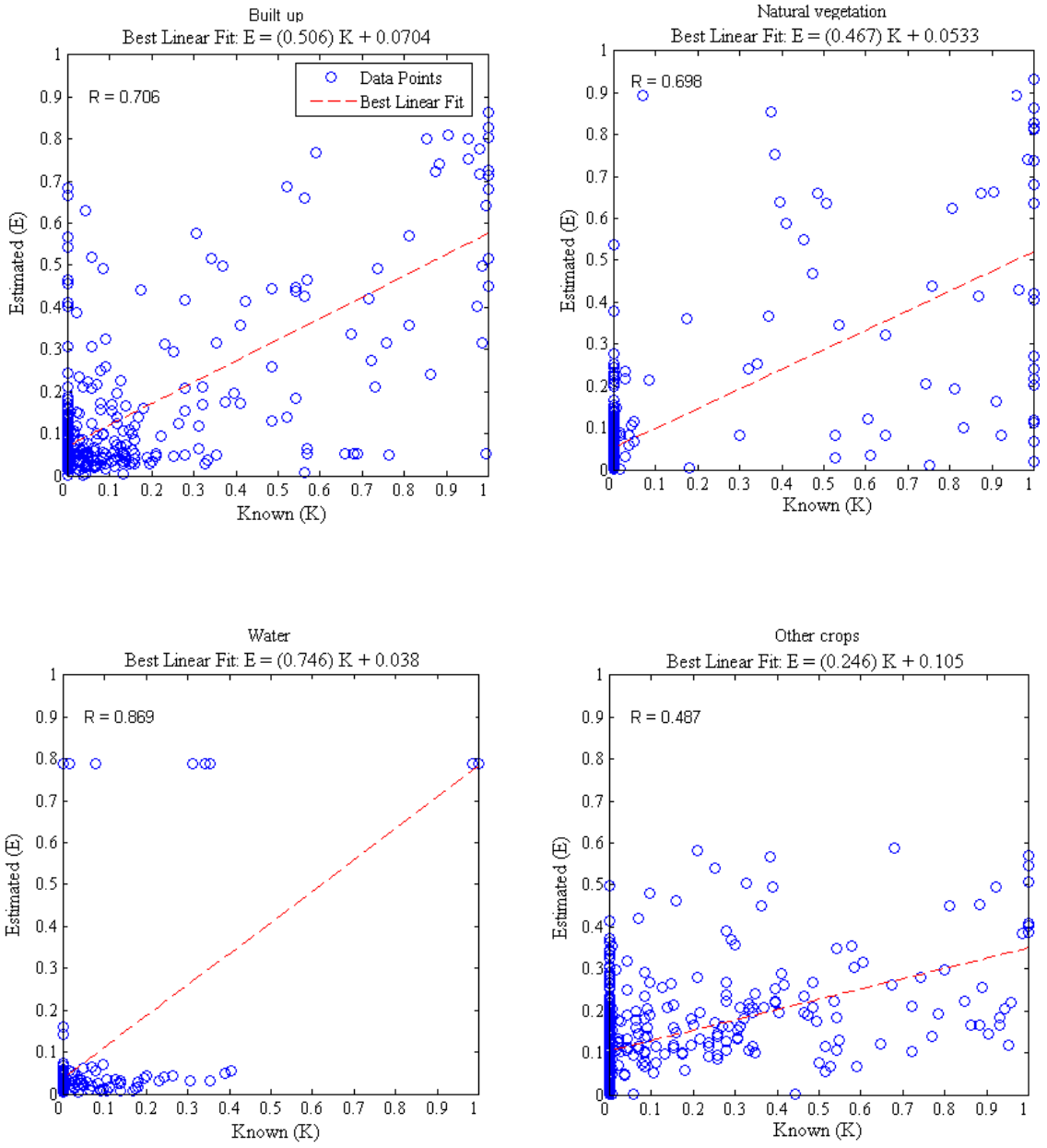


Figure 15 a. Linear regressions between estimated fractions from a pixel by pixel FFNN and known fractions.

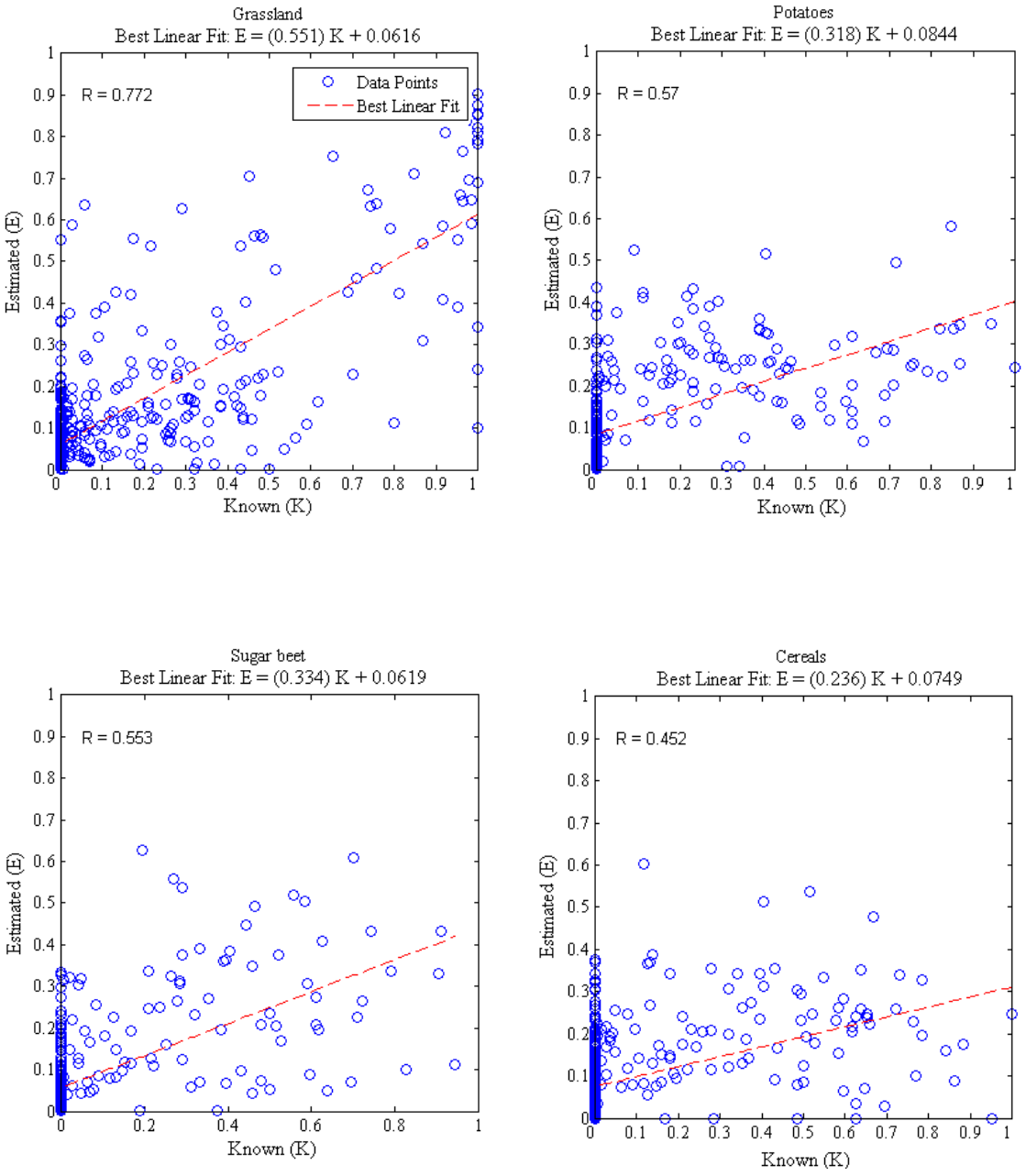


Figure 15 b. Linear regressions between estimated fractions form a 3 by 3 window FFNN and known fractions.

Figure 16 displays the linear regressions between known and estimated fractions for the 3 by 3 window FFNN. For all classes the points followed the best linear fit. Also there was a slight increase in their slope values. For example, grassland had a slope value of 0.653, compared to just 0.552 using a pixel by pixel FFNN. The built up, natural vegetation, water and grassland estimated fractions had a strong positive correlation with R values of 0.70, 0.80, 0.95 and 0.81, respectively. The estimated fractions for other crops, sugar beet, potatoes and cereal classes had a positive but moderate correlation with R values of 0.51, 0.61, 0.60 and 0.60, respectively. In general there was an improvement of the correlation values for natural vegetation, water, other crops, grassland, sugar beet and cereals when compared to the R values using a pixel by pixel FFNN. This improvement in the correlation values was due to the reduction of error estimations associated with outliers. This reduction could be expected as the 3 by 3 window FFNN was trained with more information on the image.

In addition to the progress of estimated fractions with neighbor information, there was a reduction of the saturation of high estimated fractions. For example, maximum estimated values for water was 0.9 with a 3 by 3 window FFNN, compared to 0.8 with a pixel by pixel FFNN.

The better performance for the simulated fractions coming from a 3 by 3 window FFNNs could have been foreseen during the selection of the network parameters (section 4.5). In all those evaluations the trained 3 by 3 window FFNNs had lower average MSEs than the networks trained with a pixel by pixel FFNN.

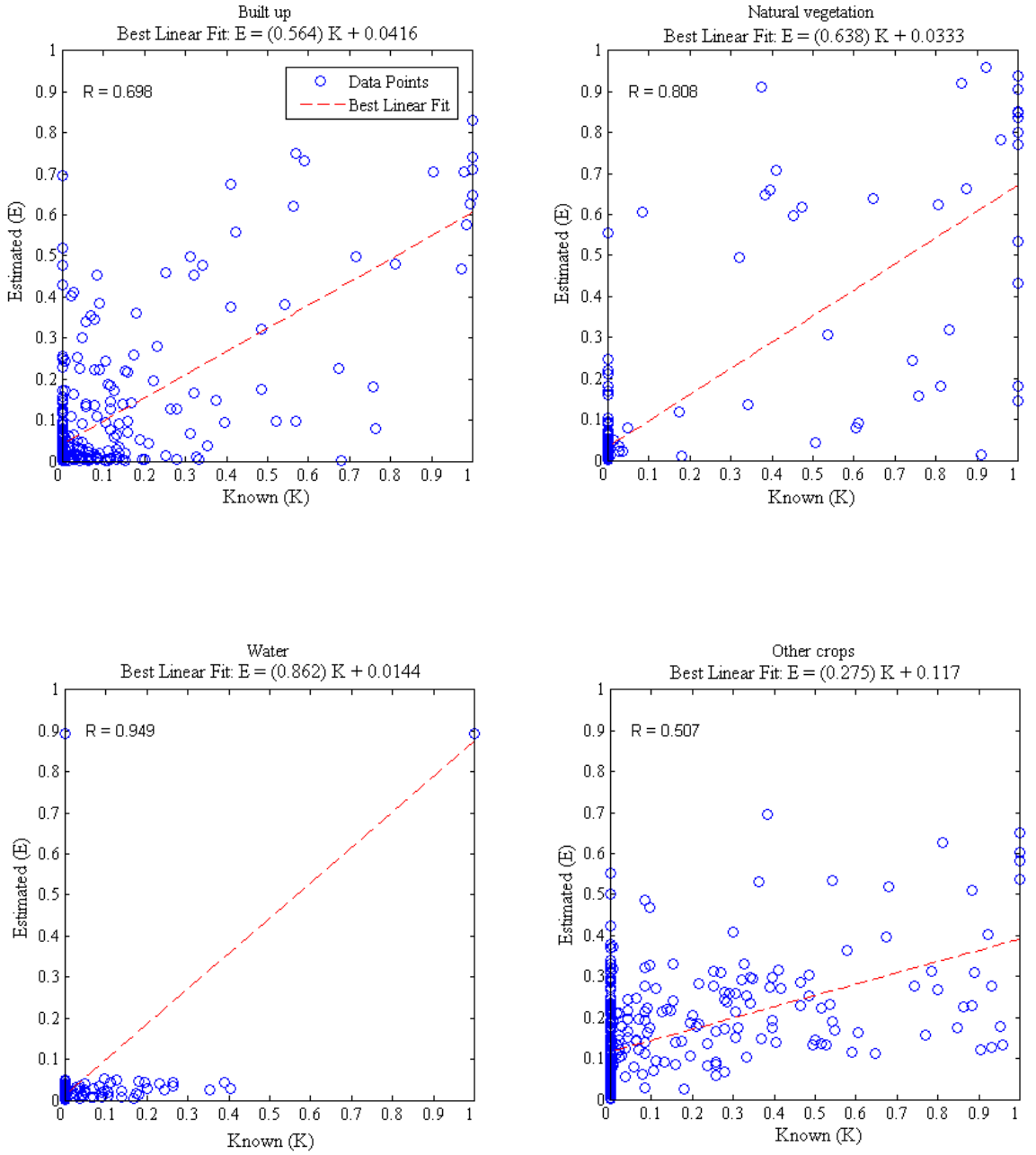


Figure 16 a. Linear regressions between estimated fractions from a pixel by pixel FFNN and known fractions.

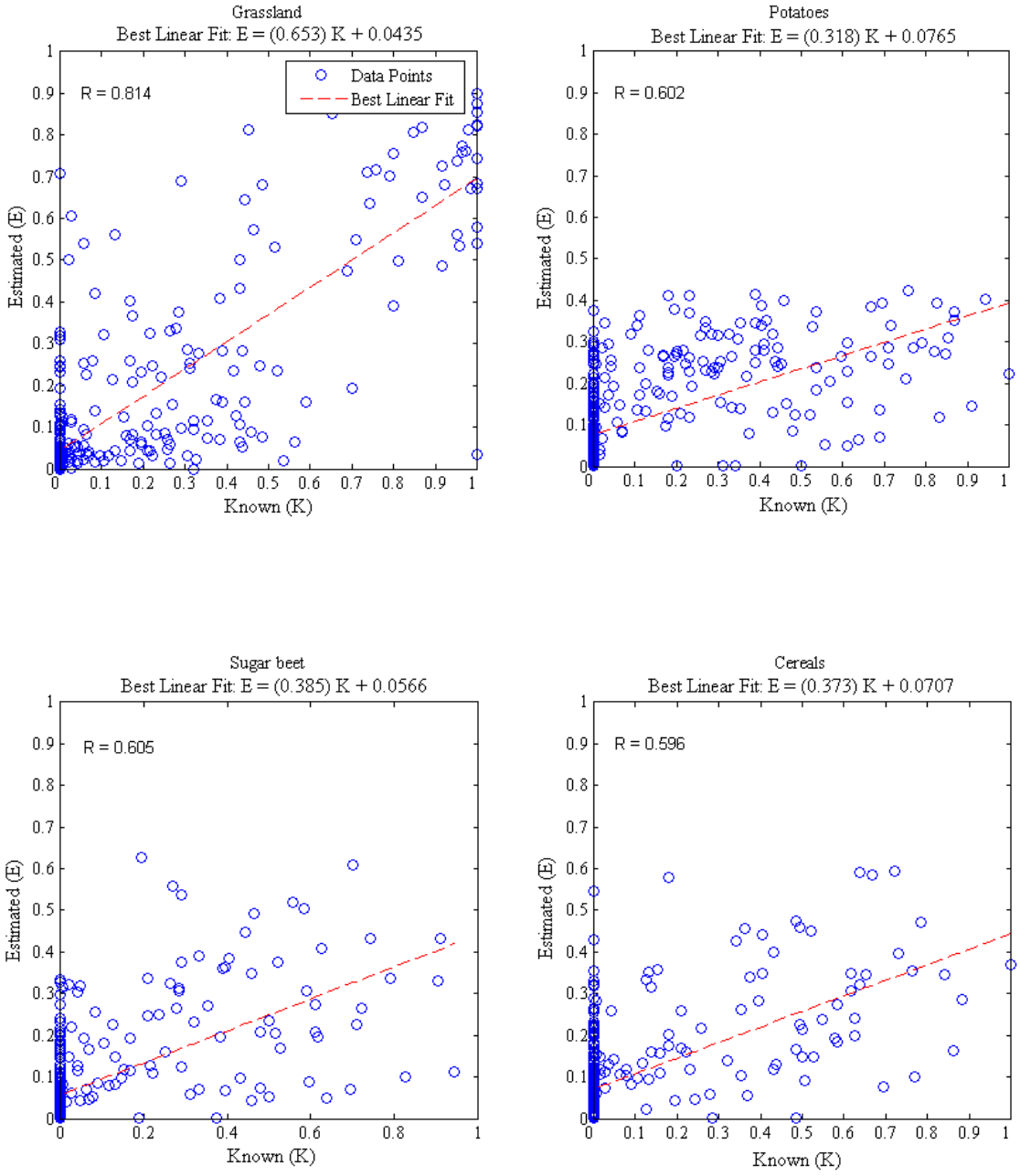


Figure 16 b. Linear regressions between estimated fractions from a 3 by 3 window FFNN and known fractions.

4.7. ANALYSIS OF ESTIMATES AT DIFFERENT SCALES

The information provided during simulations gave us some insight about the potential of the FFNN to estimate sub-pixel fractions. However, to evaluate in detail the sub-pixel estimations and their spatial distribution at different scales we need to proceed with another evaluation approach. For this purpose we selected two regions of 900 ha (10 by 10 MERIS pixels) for potatoes, region I, and grassland, region II, displayed in Figure 17. Potatoes class was chosen because it had similar patch size and patch variability with sugar beet and cereals, whereas grassland had a greater varying patch size and its pixels in the LGN5 were more tied to its same class showing more homogeneity than areas with potatoes, sugar beet or cereals. For these two regions a set of statistics comprising the root mean square error $(RMSE)$, the normalized RMSE $(RMSE_{normalized})$, Bias and coefficient of correlation (R) was calculated. As this approach contains different scales of analysis, it is important to mention that we refer the estimated and known sub-pixel values to the area covered by potatoes or grassland over the scale (ground area) in consideration as depicted in Figure 9.

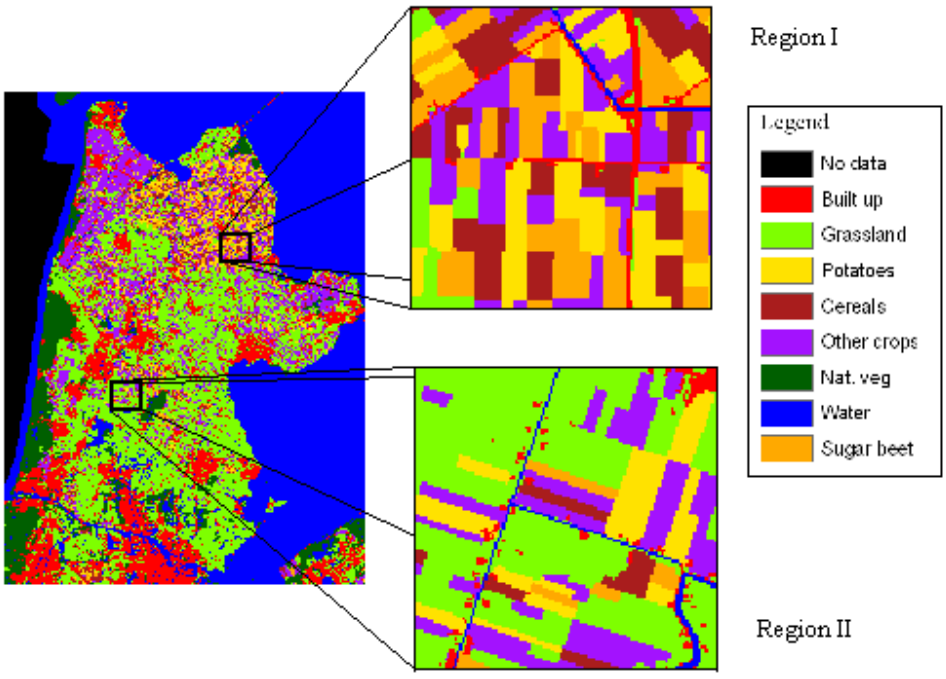


Figure 17. Regions I and II selected from the LGN 5.

4.7.1. Analysis for potato estimates

Figure 18 shows in detail the region I with its respective sub-pixel fractions, known and estimated values from a pixel by pixel FFNN and a 3 by 3 window FFNN.

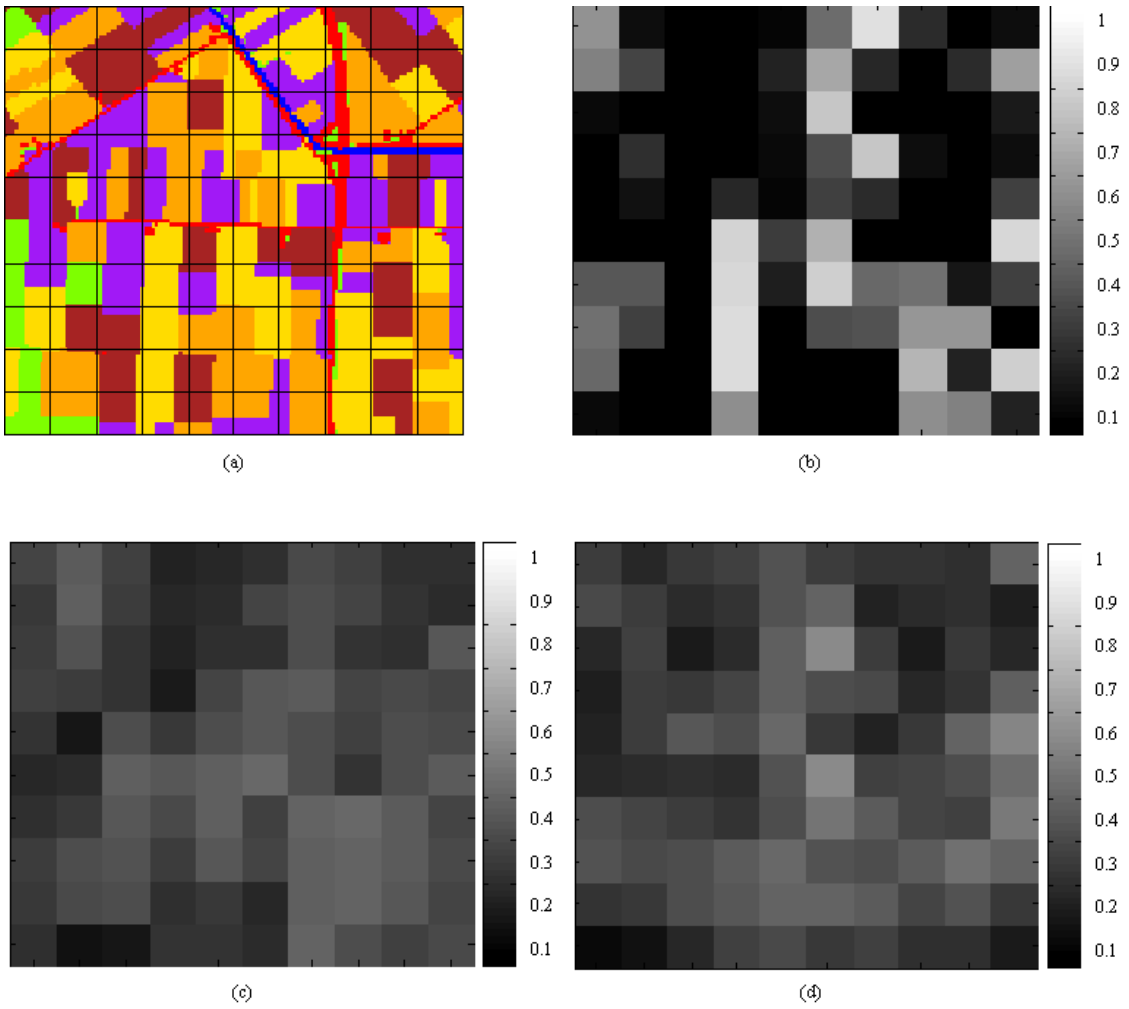


Figure 18. Region I for potato estimates. (a) shows the land use distribution (LGN 5) overlaid by a grid cell resembling MERIS pixels. (b) the fractions of sub-pixels covered by potatoes represented in grey scale (known fractions). (c) estimated fractions with a pixel by pixel FFNN structure. (d) estimated fractions with a 3 by 3 window FFNN structure.

Figure 18a shows that a large part of the MERIS pixels for this region are dominated by mixed pixels. Visual comparisons of the fractional image using the pixel by pixel FFNN (Figure 18c) and 3 by 3 window FFNN (Figure 18d) with known potato fractions (Figure 18b) did not show any correspondence at pixel level.

Statistical comparisons for both FFNN structures, showed in Figure 19, confirmed the above visual assessment. The R values for both networks were around 0.25 at a ground area of 9 ha (MERIS pixel size of 300m). Nevertheless, an increase in area of analysis at 81 ha produced an increase in R values of 0.41 in a pixel by pixel FFNN and of 0.85 in a window 3 by 3 FFNN. This increase in precision was also showed in the calculated RMSE and the $RMSE_{normalized.}$ Their values decreased from 26% error area of estimates to around 7 % for the RMSE and from 120% error area of estimates to around 5% for the $RMSE_{normalized}$. Similar results were obtained by Townshend *et al.* (2000), Huang *et al.* (2002) and Lobell and Asner (2004). They associated the low accuracies at sensor pixel level to uncertainties coming from atmospheric effects, sensor response and landscape heterogeneity.

As we aggregated the spatial area of analysis to a ground area of 18 ha (2 MERIS pixels) over the horizontal axis (E-W) and vertical axis (N-S), the calculated R values from the pixel by pixel FFNN showed some big differences. The aggregated area over the horizontal axis (E-W) had a R value of 0.53, while the aggregated area over the vertical axis (N-S) had a R value of 0.3. This difference of 0.2 could have been originated from erroneous estimations during the simulations because this effect is smoothed out when using a 3 by 3 window FFNN.

The information provided by the $RMSE_{normalized}$ had great differences with the information provided by the RMSE. For example at a scale of 9 ha the error of estimates with the RMSE was 2.52 ha (28%) for the estimated potato area while the correlation of the estimated values was very low with R values of 0.25 for both networks. Instead the error of estimates calculated with the $RMSE_{normalized}$ was 10.8 ha (120%). This difference in values was because the average known potato area per sub-region was 2.06 ha which is

below the estimated error calculated with the RMSE. Hence the normalization of the RMSE for information on area distribution can provide measurements that are better related to other accuracy methods used for sub-pixel estimations like coefficients of correlation.

The bias error of the potato estimates was around 7% and positive that meant a low systematic overprediction of the estimates. This can be expected from the behavior of the log sigmoid transfer function, as it tends to overemphasize low estimates.

As region I included small areas covered with grassland, estimations over grassland at 9 ha had a positive moderate correlation with a R value of 0.50 for the pixel by pixel FFNN and positive strong correlation with a R value of 0.75 for the 3 by 3 window FFNN.

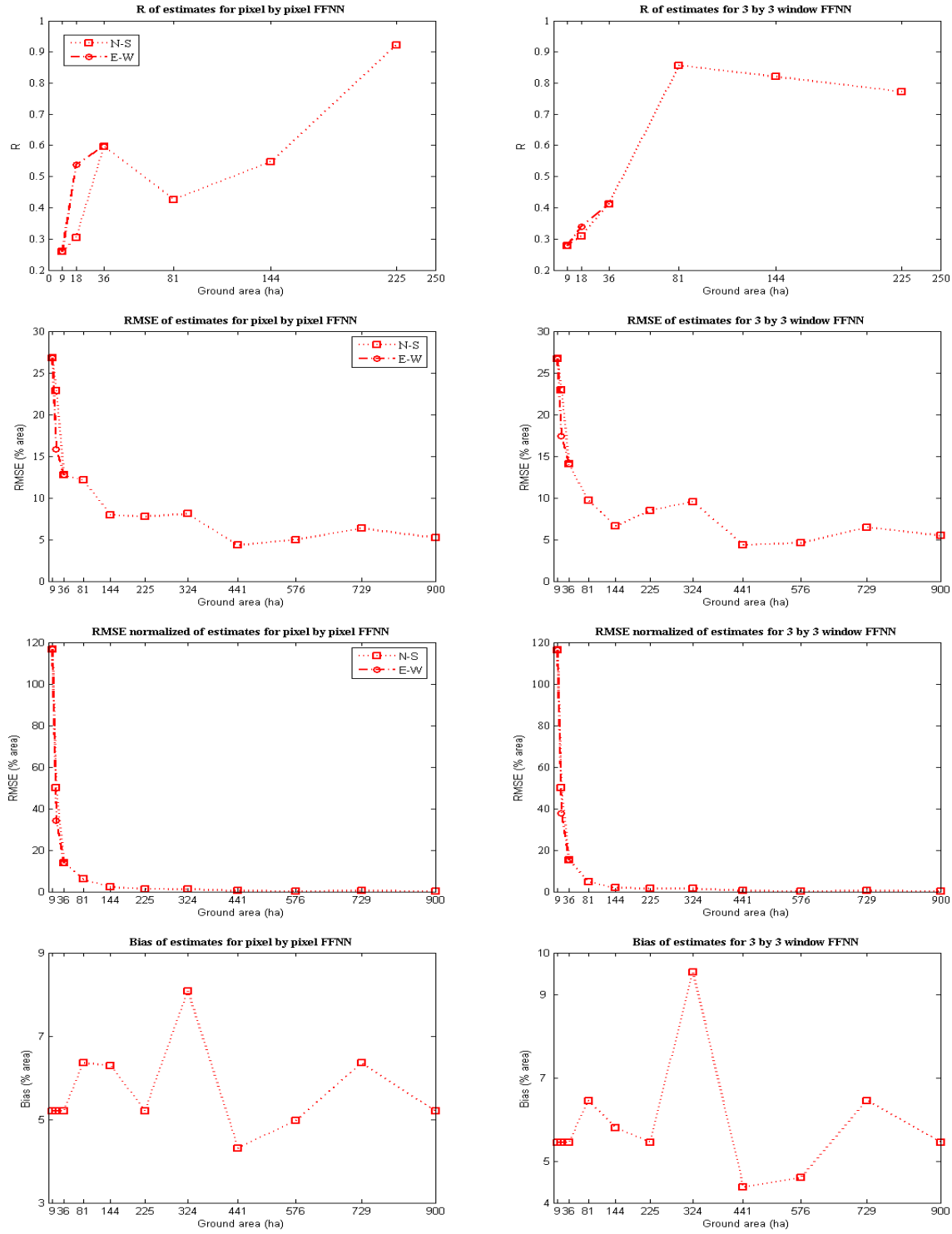


Figure 19. RMSE, $RMSE_{normalized,}$ R and Bias for MERIS estimates of potato area at different spatial scales. The legend N-S and E-W corresponds to the aggregation along the vertical axis and horizontal axis, respectively, at analysis scale of 18 ha.

4.7.2. Analysis for grassland estimates

As low estimates of potato fractions could be related to the very little composition of this class in the landscape and lack of training samples with large fractions, we decided to carry out a similar analysis with grassland that comprised 22.3% of the whole study area and was trained with equally distributed fractions. Figure 20 presents the region II with its respective sub-pixel fractions, known and estimated values for grassland, from a pixel by pixel FFNN and a 3 by 3 window FFNN.

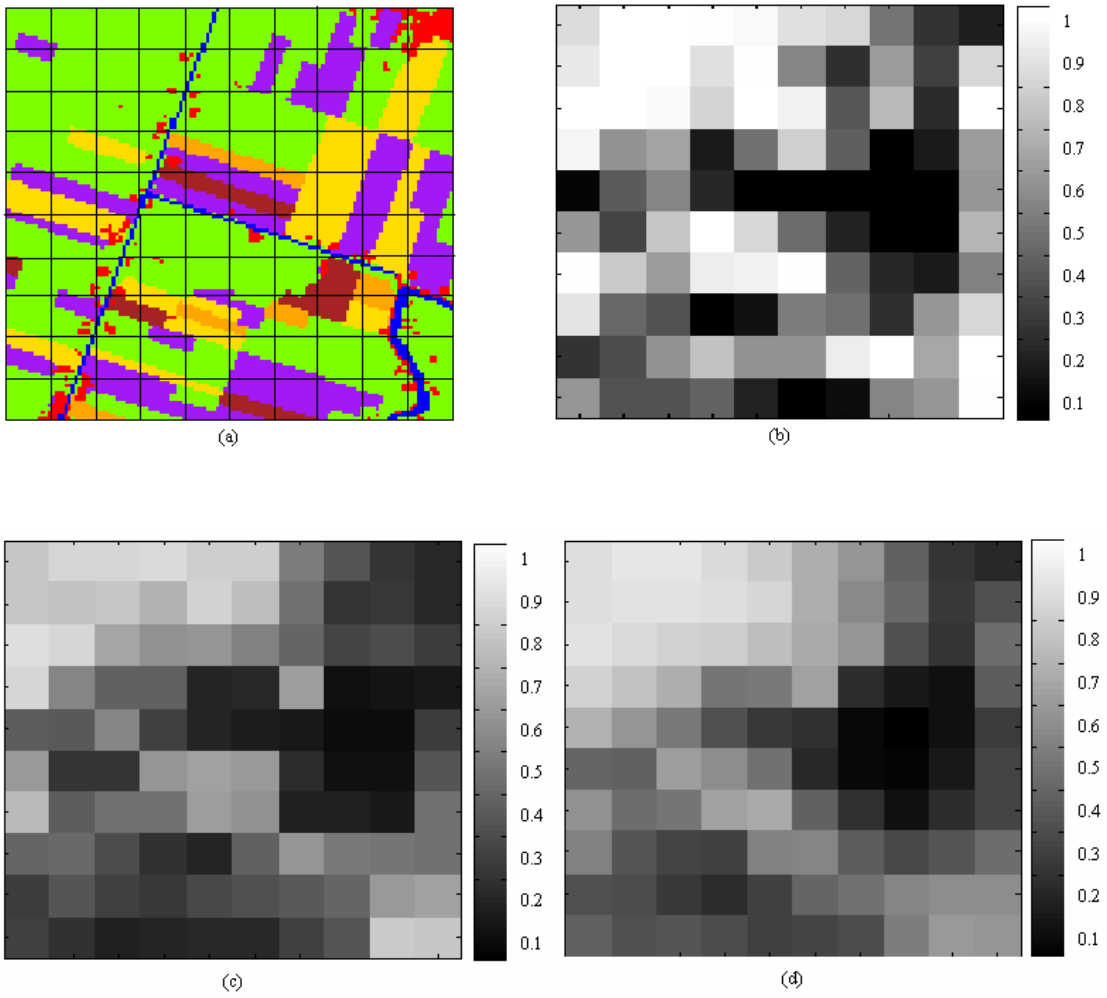


Figure 20. Region II for grassland estimates. (a) shows the land use distribution (LGN 5) overlaid by a grid cell resembling MERIS pixels. (b) the fractions of sub-pixels covered by grassland represented in grey scale (known fractions). (c) estimated fractions with a pixel by pixel FFNN structure. (d) estimated fractions with a window 3 by 3 FFNN structure.

Region II shows that large proportion of the MERIS pixels are dominated by the class grassland. Visual comparisons of the fractional images derived using the pixel by pixel FFNN, Figure 20c, and 3 by 3 window FFNN, Figure 20d with known grassland fractions Figure 20b presented a pretty good spatial correspondence between the known proportions and its estimated values at pixel level. Also grassland patches were spatially well distributed throughout the region.

The statistics showed in Figure 21 for both FFNN structures confirmed the above visual assessment. Estimated grassland fractions were positive and strong correlated with R values around 0.72 at 9 ha for both networks.

An increase in area of analysis of the estimates, for the pixel by pixel FFNN and 3 by 3 window FFNN, produced also an increase in the correlations, with R values of 0.95 at 81 ha, compared to 0.72 at 9 ha. This increase in precision was also apparent by a decrease in RMSE and $RMSE_{normalized}$. The RMSE were reduced from 28% at 9 ha to 14 % at 81 ha for the pixel by pixel FFNN and from 25% at 9 ha to 11% at 81 ha for the 3 by 3 window FFNN. The $RMSE_{normalized}$ percentages were reduced from 50% at 9 ha to 3% at 81 ha for the pixel by pixel FFNN and from 45% at 9 ha to 3% at 81 ha for the window 3 by 3 FFNN. In addition, uncertainties associated with the direction of aggregation at a scale of 18 ha were not significant.

The bias for grassland estimated values was around -7%, which meant an underprediction over high fractions. It was negative due to the tendency of the log sigmoid transfer function to underestimate high fractions.

In general the window 3 by 3 FFNN structure was more precise and accurate in estimating sub-pixel area information for grassland than the pixel by pixel FFNN at a scale of 9 ha.

As region II includes small areas covered with potatoes, we calculated its correlation between estimates and known areas at a spatial scale of 9 ha. The pixel by pixel FFNN

had a positive low correlation with a R value of 0.48, while the 3 by 3 window FFNN had a positive low correlation with a R value of 0.50. Despite that R values were low correlated at 9 ha, they are larger to the correlations measured in region I, for potatoes, with R values of 0.25 at 9 ha. This general improvement in correlation values over more homogeneous areas like in region II might be explained by the fact that accuracy of estimations significantly decreases over heterogeneous areas (Moody *et al.* 1996). Nevertheless, the distribution of training samples used to train the networks might have also significantly influenced the estimations over heterogeneous and more homogenous regions. For example, grassland areas were positive strong correlated with R values around 0.75 for both regions.

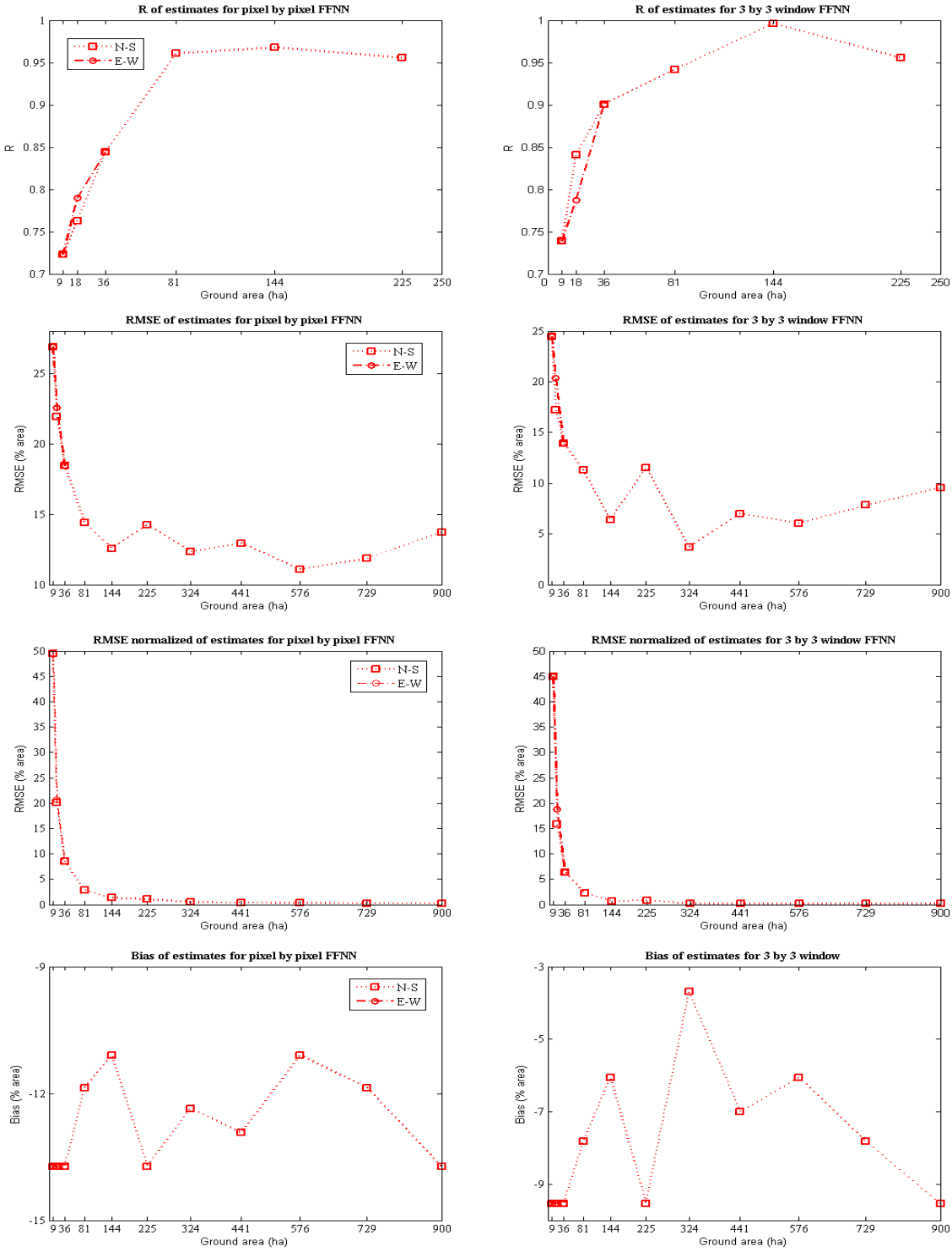


Figure 21. RMSE, $RMSE_{normalized}$, R and Bias for MERIS estimates of grasland area at different spatial scales. The legend N-S and E-W corresponds to the aggregation along the vertical axis and horizontal axis, respectively, at an analysis scale of 18 ha.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. CONCLUSIONS

Results of this study show the potential of extracting sub-pixel crop area estimates from time series of MERIS level 2 datasets by using a pixel by pixel Feedforward Artificial Neural Network (FFNN) and a 3 by 3 window FFNN.

The approach used in this study to assess the separability of crop types from multi-temporal and multi-spectral datasets followed an evaluation of the spectral signatures of crops per image and over time. Calculation of correlation coefficients and principal components for the MERIS level 2 bands and products showed that MERIS have a high degree of redundant information in its spectral bands for the visible and near infrared. Vegetation indices (VIs) showed high correlation with the NIR bands. Most of the information could thus be comprised in MERIS level 2 products rectified red and NIR or a combination of the two using the MGVI. Separability of temporal spectral signatures measured with the JM distance revealed that for good separability, distance values between 1.3 and 1.4, the quantity of MERIS images was more important than their position at the onset, middle and end of the growing season. The resulting spectral signatures using all five MERIS datasets showed good separability between crops, only the classes forest and natural vegetation, which had a JM-value of 1.25, were merged into a single class.

Based on the temporal profiles and output classes the best parameter values for training of FFNNs were obtained by using ρ as input datasets, 13 hidden layer nodes for a pixel by pixel FFNN as proposed by Garson (1998) and 49 hidden layer nodes for a 3 by 3 window FFNN as proposed by Ripley (1993) and 2400 training samples as proposed by Hush (1989). Although the guideline proposed by Kavzoglu and Mather (2003) facilitates the process of design and use of neural networks towards an operational application,

parameters such as number of hidden layer nodes and number of training samples need to be adjusted to case-by-case.

Neighbor pixel information from MERIS datasets improved the sub-pixel crop estimates. The sub-pixel estimations with a 3 by 3 window FFNN, for instance, had coefficient of correlation (R) values of 0.81 for grassland, 0.61 for sugar beet, 0.60 for potatoes and 0.60 for cereals; compared to sub-pixel estimations with R values of 0.77 for grassland, 0.55 for sugar beet, 0.57 for potatoes and 0.45 for cereals using a pixel by pixel FFNN. The improvement in the correlations can be explained by a reduction of errors associated with the outliers. Another interesting effect of using neighbor information was the reduction of the saturation effect for high estimates due to the behavior of the log sigmoid function. For example, maximum estimated value for water was 0.9 with a 3 by 3 window FFNN, compared to 0.8 with a pixel by pixel FFNN.

Accuracy of sub-pixel crop area estimates changed with the spatial scale of analysis, which ranged from 9 ha to 900 ha. Analysis for potato estimates at different spatial scales yielded R values from 0.25 at 9 ha to 0.85 at spatial scales over 81 ha using a 3 by 3 window FFNN. For grassland, there was an increased from R values of 0.72 at 9 ha to 0.95 at spatial scales over 81 ha using a 3 by 3 window FFNN. Such increase in precision as function of the spatial scale produced a reduction for RMSE values. For example, the $RMSE_{normalized}$ over grassland were reduced from 45% error area of estimates at 9 ha to 3% at 81 ha for the 3 by 3 window FFNN. Accuracy of crop area estimates over heterogeneous regions as depeicted in region I, for potatoes, was lower than in regions with a more homogenous area as presented in region II, for grassland. Nevertheless, this low estimation can not only be related to heterogeneity of the landscape but also to the training samples used for describing potato and grassland fractions. Besides, it was important to quantify the systematic error coming from the log sigmoid function that was around 7% area of estimates for low estimated values and –7% for high estimated values.

5.2. RECOMMENDATIONS

Future research drawing on and furthering the present study may include:

Definition of temporal profiles with higher temporal resolutions would provide a better definition of the endmembers. In addition, definition of endmembers in terms of VI's should be further explored with VI's ranging between 0 and 1 or otherwise normalization into this range.

Although most of the ANN studies in remote sensing take the transfer sigmoid function as given, implementation of other transfer functions with output values ranging between 0 and 1 could reduce the saturation effect for high fractions and overestimation of low fractions.

The training sample can be optimized by the use of look up tables combining the different ranges of proportion per class of interest. If needed training samples for pure pixels might also be derived from a combination of crop growth models and radiative transfer models (Lobell and Asner 2004).

The effect on accuracy of sub-pixel area estimates using other sources of ancillary data like soil or topographic maps needs further exploration. In addition, the evaluation of the accuracy at sub-pixel level needs a standardization of statistical methods. In this study, the implemented $RMSE_{normalized}$ proves to be useful when evaluating accuracy estimates at pixel level. It normalizes the RMSE with information on class distribution, making the measurements better related to other accuracy methods used for sub-pixel estimations like the coefficient of correlation.

6. REFERENCES

Atkinson, P. M., M. E. J. Cutler, and H. Lewis 1997, Mapping sub-pixel proportional land cover with AVHRR imagery, *International Journal of Remote Sensing*, **18**, 917-935.

Atkinson, P. M., and A. R. L. Tatnall 1997, Neural networks in remote sensing, *International Journal of Remote Sensing*, **18**, 699-709.

Atkinson, P. M., and P. Aplin 2004, Spatial variation in land cover and choice of spatial resolution for remote sensing, *International Journal of Remote Sensing*, **25**, 3687-3702.

Benediktsson, J. A., P. H. Swain, and O. K. Ersoy 1993, Conjugate-gradient neural networks in classification of multisource and very-high-dimensional remote sensing data, *International Journal of Remote Sensing*, **14**, 2883-2903.

Benediktsson, J. A., and J. R. Sveinsson 1997, Feature extraction for multisource data classification with artificial neural networks, *International Journal of Remote Sensing*, **18**, 727-740.

Bezdek, J. C., R. Ehrlich, and W. Full 1984, FCM: the fuzzy c-means clustering algorithm, *Computers and Geosciences*, **10**, 191-203.

Bishop, C. M. 1995, *Neural Networks for Pattern Recognition*, Oxford University Press Inc, Oxford, 482 pp.

Borel, C. C., and S. A. W. Gerstl 1994, Nonlinear spectral mixing models for vegetative and soil surfaces, *Remote Sensing of Environment*, **47**, 403-416.

Braswell, B. H., S. C. Hagen, S. E. Frolking, and W. A. Salas 2003, A multivariable approach for mapping sub-pixel land cover distributions using MISR and MODIS: Application in the Brazilian Amazon region, *Remote Sensing of Environment*, **87**, 243-256.

Brisco, B., and R. J. Brown 1995, Multidate SAR/TM synergism for crop classification in western Canada, *Photogrammetric Engineering and Remote Sensing*, **61**, 1009-1014.

Brockmann, C. 2005, Tutorial BEAM software, Access date: 22-09-2006, http://www.brockmann-consult.de/beam/tutorials.html.

CBS 2004, Frequency distribution of agricultural fields based on crop area for the year 2002, Centraal Bureau voor de Statistiek, Access date: 11-04-2006, http://statline.cbs.nl/StatWeb/Start.asp?lp=Search/Search&LA=EN&DM=SLEN.

CCSE-SWCC 1997, Policy instruments for environmental protection in agriculture: analytical review of the literature, Eastern Canada Soil and Water Conservation Centre, Grand falls, Access date: 28-02-2006, http://www.ccse-swcc.nb.ca/publications/english/policy_iep.pdf.

Clevers, J. G. P. W., and R. Jongschaap 2001, Imaging Spectrometry for Agricultural Aplications, in *Imaging Spectrometry*, edited by F. D. Van der Meer and S. M. De Jong, Kluwer Academic Publishers, Dordrecht, pp. 157-199

Clevers, J. G. P. W., R. Zurita Milla, M. E. Schaepman, and H. M. Bartholomeus 2005, Using Meris on Enivisat for land cover mapping. In Proceedings of the 2004 ENVISAT & ERS Symposium, Salzbug, Austria, pp. 10.

Clevers, J. G. P. W., M. E. Schaepman, C. A. Mücher, A. J. W. De Wit, R. Zurita Milla, and H. M. Bartholomeus 2006, Using MERIS on ENVISAT for land cover mapping in the Netherlands, *International Journal of Remote Sensing*, **In press**.

Dash, J., and P. J. Curran 2004, The MERIS terrestrial chlorophyll index, *International Journal of Remote Sensing*, **25**, 5403-5413.

De Bruin, S. 2000, Querying probabilistic land cover data using fuzzy set theory, *International Journal of Geographical Information Science*, **14**, 359-372.

De Jong, S. M., E. J. Pebesma, and F. D. Van der Meer 2004, Spatial Variability, Mapping Methods, Image Analysis and Pixels, in *Remote Sensing ImageAnalysis: Including the Spatial Domain*, edited by S. M. De Jong and F. D. Van der Meer, Kluwer Academic Publishers, Dordrecht, pp. 17 - 35

De Wit, A. J. W., and J. G. P. W. Clevers 2004, Efficiency and accuracy of per-field classification for operational crop mapping, *International Journal of Remote Sensing*, **25**, 4091-4112.

Demuth, H., M. Beale, and M. Hagan 2005, Neural Network Toolbox For Use with MATLAB, edited, Math Works, Natick.

European-Commission 2006, CAP Reform: EU agriculture ministers adopt groundbreaking sugar reform, European Commission press release, Brussels, Access date: 10-03-2006, http://europa.eu.int/rapid/pressReleasesAction.do?reference=IP/06/194&format=HTML&aged=0&language=EN&guiLanguage=en.

Foody, G. M. 1992, A fuzzy sets approach to the representation of vegetation continua from remotely sensed data: an example from lowland heath, *Photogrammetric Engineering & Remote Sensing*, **58**, 221-225.

Foody, G. M., and D. P. Cox 1994, Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions, *International Journal of Remote Sensing*, **15**, 619 - 631.

Foody, G. M. 1996, Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data, *International Journal of Remote Sensing*, **17**, 1317-1340.

Foody, G. M., R. M. Lucas, P. J. Curran, and M. Honzak 1997, Non-linear mixture modelling without end-members using an artificial neural network, *International Journal of Remote Sensing*, **18**, 937-953.

Foody, G. M. 2004, Sub-Pixel Methods in Remote Sensing, in *Remote Sensing Image Analysis: Including The Spatial Domain*, edited by S. M. d. Jong and F. D. v. d. Meer, Kluwer Academic Publishers, Dordrecht, pp. 37 - 49

Foody, G. M., and A. Mathur 2004, A relative evaluation of multiclass image classification by support vector machines, *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 1335-1343.

Garson, G. D. 1998, *Neural Networks: an introductory guide for social scientists*, Sage Publications, London, 130 pp.

Gobron, N., O. Aussedat, B. Pinty, M. Taberner, and M. M. Verstraete 2004, MERIS: level 2 land surface products algorithm theoretical basis document, European Commision JRC, Report: 2.10, Ispra, 20 pp.

Gong, P., J. R. Miller, and M. Spanner 1994, Forest canopy closure from classification and spectral unmixing of scene components - multisensor evaluation of an open canopy, *IEEE Transactions on Geoscience and Remote Sensing*, **32**, 1067-1073.

Haykin, S. 1994, *Neural Networks: a comprehensive fundation*, Macmillan, New York, 696 pp.

Hazeu, G. W. 2005, LGN5 crop accuracy estimation, Alterra, Wageningen UR, Wageningen, Access date: 19-08-2005, http://www.lgn.nl.

Huang, C., J. R. G. Townshend, S. Liang, S. N. V. Kalluri, and R. S. DeFries 2002, Impact of sensor's point spread function on land cover characterization: Assessment and deconvolution, *Remote Sensing of Environment*, **80**, 203 - 212.

Hush, D. R. 1989, Classification with neural networks: a performance analysis. In Proceedings of the Proceedings of the IEEE International Conference on Systems Engineering, Dayton, Ohio, pp. 277-280.

Ichoku, C., and A. Karnieli 1996, A Review of Mixture Modeling Techniques for Sub-Pixel Land Cover Estimation, *Remote Sensing Reviews*, **13**, 161-186.

Jewell, N. 1989, An evaluation of multi-date SPOT data for agriculture and land use mapping in the United Kingdom, *International Journal of Remote Sensing*, **10**, 939-951.

Kanellopoulos, I., A. Varfis, G. G. Wilkinson, and J. Megier 1992, Land-cover discrimination in SPOT HRV imagery using an artificial neural network - a 20-class experiment, *International Journal of Remote Sensing*, **13**, 917 - 924.

Kanellopoulos, I., and G. G. Wilkinson 1997, Strategies and best practice for neural network image classification, *International Journal of Remote Sensing*, **18**, 711-725.

Kavzoglu, T., and P. M. Mather 1999, Pruning artificial neural networks: An example using land cover classification of multi-sensor images, *International Journal of Remote Sensing*, **20**, 2787-2803.

Kavzoglu, T., and P. M. Mather 2002, The role of feature selection in artificial neural network applications, *International Journal of Remote Sensing*, **23**, 2919-2937.

Kavzoglu, T., and P. M. Mather 2003, The use of backpropagating artificial neural networks in land cover classification, *International Journal of Remote Sensing*, **24**, 4907-4938.

Lillesand, T. M., and R. W. Kiefer 2000, *Remote Sensing and Image Interpretation*, 4 ed., Jonh Wiley & Sons, Inc., New York, 724 pp.

Lobell, D. B., and G. P. Asner 2004, Cropland distributions from temporal unmixing of MODIS data, *Remote Sensing of Environment*, **93**, 412-422.

Marsh, S. E., P. Switzer, W. S. Kowalik, and R. J. P. Lyon 1980, Resolving the percentage of component terrains within single resolution elements, *Photogrammetric Engineering & Remote Sensing*, **46**, 1079-1086.

Mather, P. M. 1999, *Computer processing of remotely sensed images: an introduction*, 2 ed., Jonh Wiley & Sons, Inc., Chichester, 292 pp.

MERIS-Envisat 2005, MERIS level 2: geophysical data product, European Space Agency, Access date: 01-03-2006, http://envisat.esa.int/instruments/meris/data-app/dataprod.html#top.

Moody, A., and C. E. Woodcock 1994, Scale-dependent errors in the estimation of land-cover proportions: Implications for global land-cover datasets, *Photogrammetric Engineering and Remote Sensing*, **60**, 585 -594.

Moody, A., S. Gopal, and A. H. Strahler 1996, Artificial neural network response to mixed pixels in coarse-resolution satellite data, *Remote Sensing of Environment*, **58**, 329-343.

Mücher, C. A., C. Heunks, K. T. Steinnocher, and F. P. Kressler 2000, Land cover characterization and change detection for environmental monitoring of pan-Europe, *International Journal of Remote Sensing*, **21**, 1159-1181.

Murakami, T., S. Ogawa, N. Ishitsuka, K. Kumagai, and G. Saito 2001, Crop discrimination with multitemporal SPOT/HRV data in the Saga Plains, Japan, *International Journal of Remote Sensing*, **22**, 1335-1348.

Murthy, C. S., P. V. Raju, and K. V. S. Badrinath 2003, Classification of wheat crop with multi-temporal images: Performance of maximum likelihood and artificial neural networks, *International Journal of Remote Sensing*, **24**, 4871-4890.

Paola, J. D. 1994, Neural Network classification of multispectral imagery, University of Arizona, Tucson.

Paola, J. D., and R. A. Schowengerdt 1995, A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery, *International Journal of Remote Sensing*, **16**, 3033-3058.

Paola, J. D., and R. A. Schowengerdt 1997, The effect of neural-network structure on a multispectral land-use/land-cover classification, *Photogrammetric Engineering and Remote Sensing*, **63**, 535-544.

Quarmby, N. A., J. R. G. Townshend, J. J. Settle, and K. H. White 1992, Linear mixture modelling applied to AVHRR data for crop area estimation, *International Journal of Remote Sensing*, **13**, 415-425.

Rempel, R. S. 2003, Point Analyst extension for Arc View, Access date: 10-01-2006, http://flash.lakeheadu.ca/~rrempel/pointmap/index.html.

Ripley, B. D. 1993, Statistical aspects of neural networks, in *Networks and Chaos: statistical and probabilistic aspects*, edited by O. E. Barndorff Nielsen, J. L. Jensen and W. S. Kendall, Chapman & Hall, London, pp. 40-123

Rumelhart, D. E., and J. L. MacClelland 1986, *Parallel distributed processing: explorations in the microstructure of cognition*, The MIT Press, Cambridge, 547 pp.

Sakamoto, T., M. Yokozawa, H. Toritani, M. Shibayama, N. Ishitsuka, and H. Ohno 2005, A crop phenology detection method using time-series MODIS data, *Remote Sensing of Environment*, **96**, 366-374.

Santer, R., V. Carrere, D. Dessailly, P. Dubuisson, and J. Roger 2000, Atmospheric corrections over land, Université du Littoral-Côte d'Opale, Report: atbd_2_15, Wimereux, 89 pp.

Swain, P. H., and R. C. King 1973, Two effective feature selection criteria for multispectral remote sensing. In Proceedings of the First international joint

conference on pattern recongnition, November, Washington, IEEE, pp. 536-540.

Townshend, J. R. G., C. Huang, S. N. V. Kalluri, R. S. Defries, S. Liang, and K. Yang 2000, Beware of per-pixel characterization of land cover, *International Journal of Remote Sensing*, **21**, 839 -843.

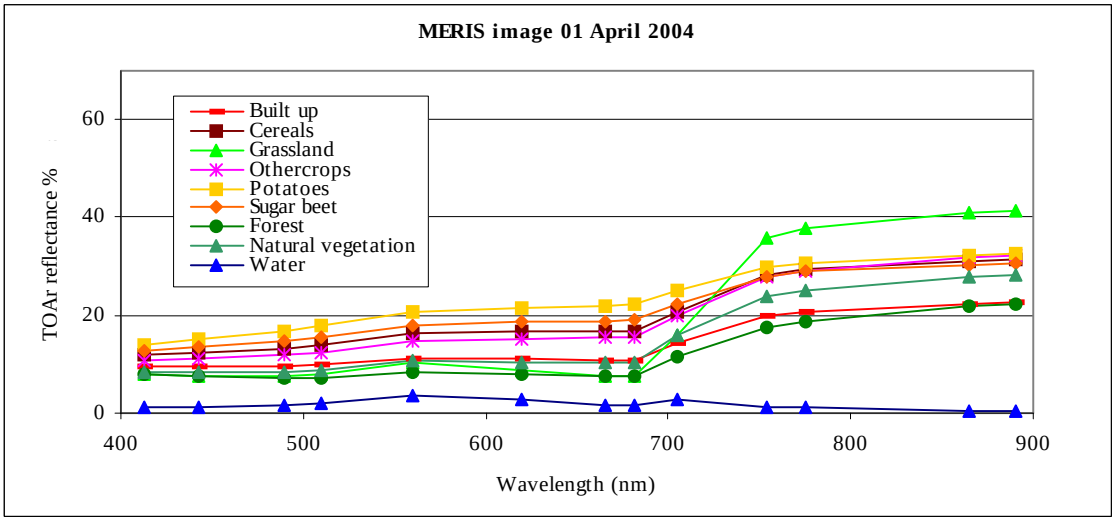
Van Kootwijk, E. J., H. Van Der Voet, and J. J. M. Berdowski 1995, Estimation of ground cover composition per pixel after matching image and ground data with subpixel accuracy, *International Journal of Remote Sensing*, **16**, 97-111.

Van Vliet, A. 2005, De Natuurkalender, Department of Environmental Sciences and System Analyses, Wganingen University, Wageningen, Access date: 20-09-2005, http://www.natuurkalender.nl.

Verstraete, M. M., B. Pinty, and P. J. Curran 1999, MERIS potential for land applications, *International Journal of Remote Sensing*, **20**, 1747-1756.

Wang, F. 1990, Fuzzy supervised classification of remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 194-201.

ANNEX I: SPECTRAL SIGNATURES OF LAND COVER TYPES



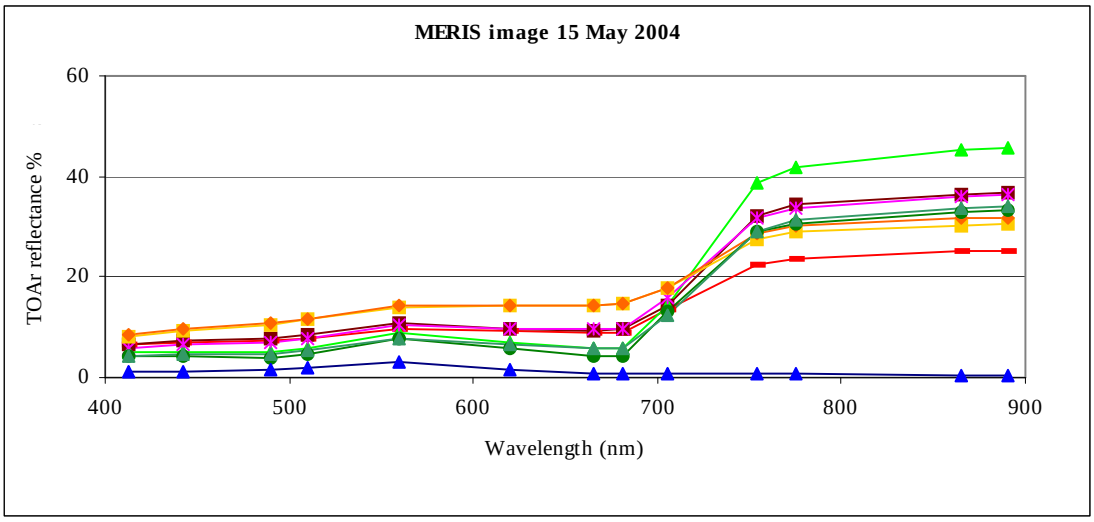
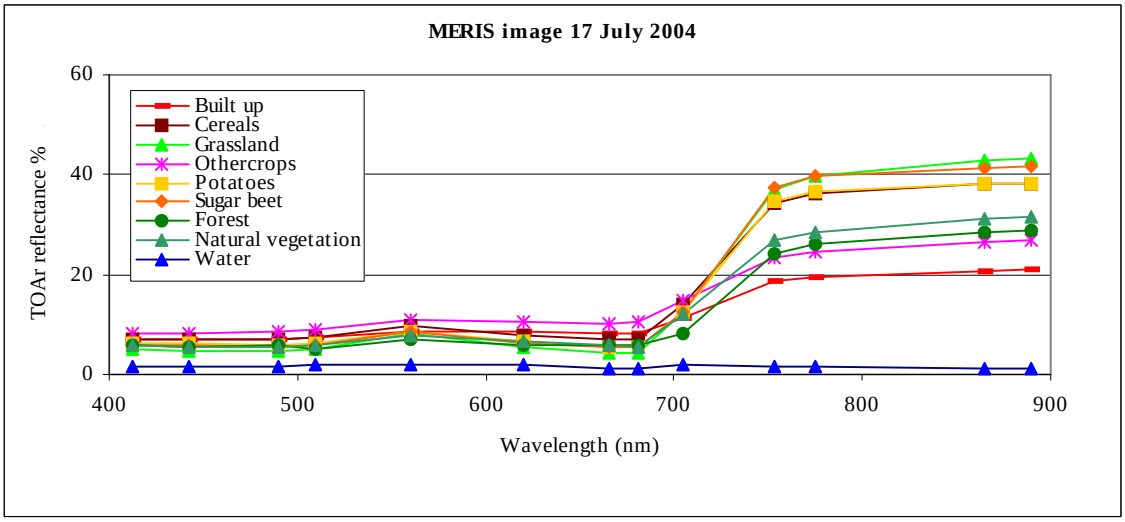


Figure 22a. Spectral signatures for land cover types derived from MERIS level 2 images of April 1st and May 15th, 2004.



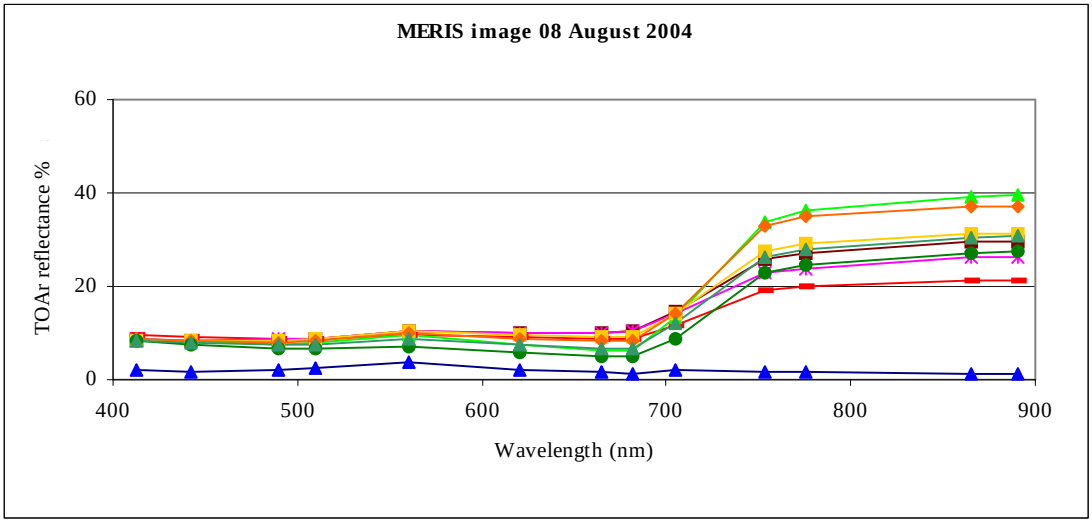


Figure 22b. Spectral signatures for land cover types derived from MERIS level 2 images of July 17th and August 08th, 2004.

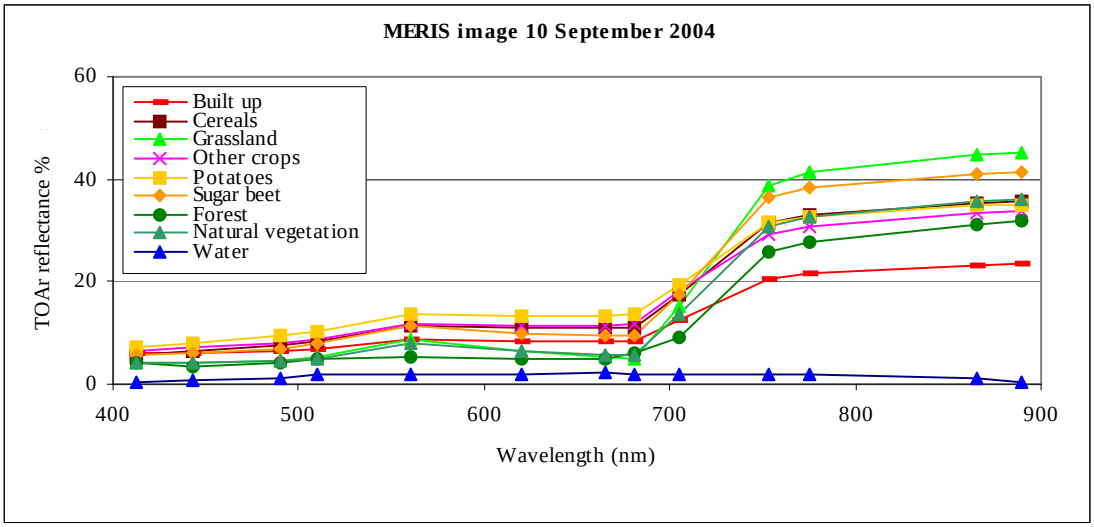


Figure 22c. Spectral signatures for land cover types derived from MERIS level 2 images of September 10th, 2004.

ANNEX II: PRINCIPAL COMPONENT ANALYSIS FOR THE MERIS IMAGE OF APRIL 1st, 2004

Table 15. Principal component analysis for the MERIS image of April 1st, 2004.

Principal component	Explained variance (%)	Cumulative variance (%)
1	93.58	93.58
2	3.29	96.87
3	2.91	99.78
4	0.16	99.94
5	0.04	99.98
6	0.01	99.99
7	0.01	100.00
8	0.00	100.00
9	0.00	100.00
10	0.00	100.00
11	0.00	100.00
12	0.00	100.00
13	0.00	100.00
14	0.00	100.00
15	0.00	100.00
16	0.00	100.00
17	0.00	100.00

ANNEX III: JM DISTANCE VALUES BETWEEN CLASES FOR TEMPORAL PROFILES

Table 16. JM Distances values between classes. Very good separability: 1.414 and Low separability: <1.3.

Temporal profile 1	JM	Temporal profile 2	JM	Temporal profile 3	JM	Temporal profile 4	JM	Temporal profile 5	JM
Cereals & Potatoes	0.429	Potatoes & Sugar beet	0.653	Forest & Nat Veg	1.139	Forest & Nat Veg	1.077	Forest & Nat Veg	1.254
Cereals & Sugar beet	0.469	Built up &Other crops	0.956	Potatoes & Sugar beet	1.192	Potatoes & Sugar beet	1.147	Built up &Other crops	1.349
Potatoes & Sugar beet	0.699	Forest & Nat Veg	0.988	Cereals & Other crops	1.239	Cereals & Potatoes	1.228	Potatoes & Sugar beet	1.351
Built up & Other crops	0.778	Cereals & Potatoes	1.096	Cereals & Potatoes	1.259	Built up & Other crops	1.262	Forest & Other crops	1.367
Nat Veg & Potatoes	0.781	Cereals & Forest	1.186	Built up & Other crops	1.286	Forest & Other crops	1.316	Cereals & Potatoes	1.369
Cereals & Forest	0.956	Forest & Other crops	1.221	Other crops & Potatoes	1.291	Forest & Grassland	1.324	Nat Veg & Other crops	1.389
Cereals & Nat Veg	0.984	Forest & Grassland	1.261	Nat Veg & Potatoes	1.295	Nat Veg & Other crops	1.339	Grassland & Nat Veg	1.395
Forest & Sugarbeet	1.078	Nat Veg &Other crops	1.300	Forest & Other crops	1.323	Built up & Forest	1.370	Cereals & Sugar beet	1.397
Nat Veg & Sugar beet	1.093	Grassland & Nat Veg	1.308	Built up & Potatoes	1.344	Grassland & Nat Veg	1.371	Forest & Grassland	1.398
Grassland & Potatoes	1.113	Cereals & Sugar beet	1.309	Built up & Cereals	1.347	Built up & Nat Veg	1.372	Built up & Forest	1.403
Cereals & Grassland	1.131	Built up & Forest	1.319	Grassland & Nat Veg	1.363	Cereals & Sugar beet	1.380	Built up & Nat Veg	1.409
Forest & Potatoes	1.139	Built up & Nat Veg	1.323	Forest & Grassland	1.366	Forest & Potatoes	1.393	Forest & Potatoes	1.410
Grassland & Sugar beet	1.146	Forest & Potatoes	1.362	Nat Veg & Other crops	1.376	Nat Veg & Potatoes	1.403	Nat Veg &Potatoes	1.411
Grassland & Nat Veg	1.206	Cereals & Grassland	1.369	Forest & Potatoes	1.381	Cereals & Forest	1.407	Cereals & Forest	1.413
Forest & Grassland	1.237	Cereals & Nat Veg	1.382	Built up & Forest	1.381	Other crops & Potatoes	1.408	Other crops & Potatoes	1.414
Nat Veg & Other crops	1.247	Nat Veg & Potatoes	1.388	Cereals & Forest	1.383	Cereals & Other crops	1.412	Forest & Sugar beet	1.414
Built up & Nat Veg	1.261	Other crops & Potatoes	1.404	Cereals & Sugar beet	1.389	Forest & Sugar beet	1.414	Cereals & Other crops	1.414
Forest & Nat Veg	1.287	Nat Veg & Sugar beet	1.406	Other crops & Sugar beet	1.395	Built up & Potatoes	1.414	Cereals & Nat Veg	1.414
Built up & Potatoes	1.339	Forest & Sugar beet	1.407	Cereals & Nat Veg	1.396	Cereals & Nat Veg	1.414	Built up &Potatoes	1.414
Other crops & Potatoes	1.346	Cereals & Other crops	1.410	Nat Veg & Sugar beet	1.400	Grassland & Sugar beet	1.414	Cereals & Grassland	1.414
Cereals & Other crops	1.363	Built up & Potatoes	1.410	Built up & Nat Veg	1.403	Cereals & Grassland	1.414	Nat Veg & Sugar beet	1.414
Built up & Cereals	1.363	Other crops & Sugar beet	1.412	Built up & Sugar beet	1.406	Built up & Cereals	1.414	Other crops & Sugar beet	1.414
Forest & Other crops	1.389	Built up & Cereals	1.413	Forest & Sugar beet	1.409	Grassland & Potatoes	1.414	Built up & Cereals	1.414
Other crops & Sugar beet	1.391	Built up & Sugar beet	1.414	Cereals & Grassland	1.414	Nat Veg & Sugar beet	1.414	Grassland & Sugar beet	1.414
Built up & Sugar beet	1.403	Grassland & Sugar beet	1.414	Grassland & Other crops	1.414	Other crops & Sugar beet	1.414	Built up & Sugar beet	1.414
Built up & Forest	1.412	Grassland & Potatoes	1.414	Grassland & Sugar beet	1.414	Built up & Sugar beet	1.414	Grassland & Other crops	1.414
Grassland & Other crops	1.414	Grassland & Other crops	1.414	Built up & Grassland	1.414	Grassland & Other crops	1.414	Built up & Grassland	1.414
Built up & Grassland	1.414	Built up & Grassland	1.414	Grassland & Potatoes	1.414	Built up & Grassland	1.414	Grassland & Potatoes	1.414

ANNEX IV: MSEs OF TRAINED NETWORKS WITH VIs AND ρ

Pixel by Pixel FFNN Matlab code trained with ρ and VIs.

```
net_sample = newff(minmax(ρ_data), [9, 8] , products:, 'trainrp');
net_sample.trainParam.show = 50;
net_sample.trainParam.epochs = 5000;
net_sample.trainParam.goal = 0.01;
net_sample.trainParam.max_fail = 20;
net_sample.layers347.initFcn = 'initwb';
net_sample.inputWeights{1,1}.initFcn = 'myinit';
net_sample = init(net_sample);
net_trained = train(net_sample, ρ_data, t_data, [], [], VV);
estimated_test = sim(net_trained, ρ_data);
input_profile = t_data - estimated_test;
perf_input_profile = mse(input_profile);
```

Table 17a. MSE values for pixel by pixel FFNNs trained with ρ.

	Iterations	MSE
	468	0.033
	339	0.033
	347	0.033
	558	0.033
	426	0.034
	416	0.033
	292	0.033
	392	0.033
	339	0.033
	631	0.033
Mean	421	0.033
s	106	0.000316

s = Standard deviation

Pixel by Pixel FFNN Matlab code trained with VIs.

```
net_sample = newff(minmax(dataVI), [13, 8] ,{'logsig', 'logsig'}, 'trainrp');
net_sample.trainParam.show = 50;
net_sample.trainParam.epochs = 5000;
net_sample.trainParam.goal = 0.01;
net_sample.trainParam.max_fail = 20;
net_sample.layers{1}.initFcn = 'initwb';
net_sample.inputWeights{1,1}.initFcn = 'myinit';
net_sample = init(net_sample);
net_trained = train(net_sample, dataVI, t_dataVI, [], [], VV_IV);
estimated_test = sim(net_trained, dataVI);
input_profile = t_dataVI - estimated_test;
perf_input_profile = mse(input_profile);
```

Table 17b. MSE values for pixel by pixel FFNNs trained with VIs.

	Iterations	MSE
	1203	0.033
	257	0.034
	395	0.034
	621	0.034
	841	0.033
	763	0.034
	785	0.034
	734	0.034
	705	0.033
	1786	0.034
Mean	809	0.034
s	427	0.000483

s = Standard deviation

Window 3 by 3 FFNN Matlab code trained with ρ.

```
net_sample = newff(minmax(ρ_data_n), [49, 8] ,{'logsig', 'logsig'}, 'trainrp');
net_sample.trainParam.show = 50;
net_sample.trainParam.epochs = 5000;
net_sample.trainParam.goal = 0.01;
net_sample.trainParam.max_fail = 20;
net_sample.layers{1}.initFcn = 'initwb';
net_sample.inputWeights{1,1}.initFcn = 'myinit';
net_sample = init(net_sample);
net_trained = train(net_sample, ρ_data_n, t_data_n, [], [], VV_neighbor);
estimated_test = sim(net_trained, ρ_data_n);
input_profile = t_data_n - estimated_test;
perf_input_profile = mse(input_profile);
```

Table 17c. MSE values for window 3 by 3 FFNNs trained with ρ.

	Iterations	MSE
	178	0.025
	150	0.027
	237	0.024
	163	0.026
	174	0.026
	300	0.024
	182	0.025
	134	0.026
	224	0.024
	167	0.025
Mean	191	0.025
s	49.34	0.00103

s = Standard deviation

Window 3 by 3 FFNN Matlab code trained with VIs.

```
net_sample = newff(minmax(dataVI_n), [49, 8] ,{'logsig', 'logsig'}, 'trainrp');
net_sample.trainParam.show = 50;
net_sample.trainParam.epochs = 5000;
net_sample.trainParam.goal = 0.01;
net_sample.trainParam.max_fail = 20;
net_sample.layers{1}.initFcn = 'initwb';
net_sample.inputWeights{1,1}.initFcn = 'myinit';
net_sample = init(net_sample);
net_trained = train(net_sample, dataVI_n, t_dataVI_n, [], [], VV_IV_neighbor);
estimated_test = sim(net_trained, dataVI_n);
input_profile = t_dataVI_n - estimated_test;
perf_input_profile = mse(input_profile);
```

Table 17d. MSE values for window 3 by 3 FFNNs trained with VIs.

	Iterations	MSE
	213	0.028
	255	0.026
	175	0.028
	212	0.026
	178	0.026
	242	0.028
	210	0.028
	235	0.026
	293	0.025
	238	0.025
Mean	225	0.027
s	35.48	0.0012

s = Standard deviation

ANNEX V: MSEs OF TRAINED NETWORKS WITH DIFERENT NUMBER OF HIDDEN LAYER NODES

Pixel by pixel FFNN Matlab code trained with 9, 13 and 30 hidden layer nodes.

```
net_sample = newff(minmax(ρ_data), [(9,13, and 30), 8] ,{'logsig', 'logsig'}, 'trainrp');
net_sample.trainParam.show = 50;
net_sample.trainParam.epochs = 5000;
net_sample.trainParam.goal = 0.01;
net_sample.trainParam.max_fail = 20;
net_sample.layers{1}.initFcn = 'initwb';
net_sample.inputWeights{1,1}.initFcn = 'myinit';
net_sample = init(net_sample);
net_trained = train(net_sample, p_data, t_data, [], [], VV);
estimated_test = sim(net_trained, p_data);
input_profile = t_data - estimated_test;
perf_input_profile = mse(input_profile);
```

Table 18a. MSE values for pixel by pixel FFNNs with 9 hidden layer nodes.

	Iterations	MSE
	468	0.033
	339	0.033
	347	0.033
	558	0.033
	426	0.034
	416	0.033
	292	0.033
	392	0.033
	339	0.033
	631	0.033
Mean	421	0.033
s	106	0.000316

s = Standard deviation

Table 18b. MSE values for pixel by pixel FFNNs with 13 hidden layer nodes.

	Iterations	MSE
	208	0.032
	345	0.032
	579	0.032
	329	0.033
	373	0.032
	347	0.032
	374	0.032
	352	0.032
	345	0.032
	345	0.032
Mean	360	0.032
s	90.24	0.00032

s = Standard deviation

Table 18c. MSE values for pixel by pixel FFNNs with 30 hidden layer nodes.

	Iterations	MSE
	417	0.033
	356	0.033
	265	0.032
	177	0.033
	295	0.033
	206	0.032
	208	0.033
	242	0.032
	160	0.033
	393	0.032
Mean	272	0.033
s	90.81	0.00051

Window 3 by 3 FFNN Matlab code trained with 2, 49, 180 and 341 hidden layer nodes.

net_sample = newff(minmax(ρ_data_n), [(2, 49, 180 and 341) , 8] ,{'logsig', 'logsig'}, 'trainrp');

net_sample.trainParam.show = 50;

net_sample.trainParam.epochs = 5000;

net_sample.trainParam.goal = 0.01;

net_sample.trainParam.max_fail = 20;

```
net_sample.layers{1}.initFcn = 'initwb';
net_sample.inputWeights{1,1}.initFcn = 'myinit';
net_sample = init(net_sample);
net_trained = train(net_sample, ρ_data_n, t_data_n, [], [], VV_neighbor);
estimated_test = sim(net_trained, ρ_data_n);
input_profile = t_data_n - estimated_test;
perf_input_profile = mse(input_profile);
```

Table 18d. MSE values for window 3 by 3 FFNNs with 2 hidden layer nodes.

	Iterations	MSE
	2601	0.037
	1347	0.034
	740	0.044
	429	0.040
	306	0.041
	363	0.044
	311	0.044
	427	0.043
	333	0.040
	377	0.044
Mean	723	0.041
s	733	0.003446

s = Standard deviation

Table 18e. MSE values for window 3 by 3 FFNNs with 49 hidden layer nodes.

	Iterations	MSE
	178	0.025
	150	0.027
	237	0.024
	163	0.026
	174	0.026
	300	0.024
	182	0.025
	134	0.026
	224	0.024
	167	0.025
Mean	191	0.025
s	49.34	0.00103

Table 18f. MSE values for window 3 by 3 FFNNs with 180 hidden layer nodes.

	Iterations	MSE
	169	0.025
	190	0.024
	243	0.023
	156	0.025
	157	0.025
	212	0.024
	179	0.026
	124	0.026
	169	0.025
	223	0.023
Mean	182	0.025
s	35.55	0.001075

s = Standard deviation

Table 18g. MSE values for window 3 by 3 FFNNs with 341 hidden layer nodes.

	Iterations	MSE
	214	0.024
	160	0.025
	204	0.025
	179	0.025
	179	0.025
	260	0.024
	258	0.023
	133	0.025
	188	0.025
	223	0.024
Mean	182	0.025
s	40.58	0.000707

ANNEX VI: MSEs OF TRAINED NETWORKS WITH 2400 AND 3300 TRAINING PIXELS

Pixel by pixel FFNN Matlab code trained with 2400 and 3300 training samples.

```
net_sample = newff(minmax(p_data2[2400 or 3300]), [13, 8] ,{'logsig', 'logsig'},
'trainrp');
net_sample.trainParam.show = 50;
net_sample.trainParam.epochs = 5000;
net_sample.trainParam.goal = 0.01;
net_sample.trainParam.max_fail = 20;
net_sample.layers{1}.initFcn = 'initwb';
net_sample.inputWeights{1,1}.initFcn = 'myinit';
net_sample = init(net_sample);
net_trained = train(net_sample, p_data2, t_data2 [2400 or 3300] , [], [], VV);
estimated_test = sim(net_trained, p_data2);
input_profile = t_data2 - estimated_test;
perf_input_profile = mse(input_profile);
```

Table 19a. MSE values for pixel by pixel FFNNs trained with 2400 training samples.

	Iterations	MSE
	208	0.032
	345	0.032
	579	0.032
	329	0.033
	373	0.032
	347	0.032
	374	0.032
	352	0.032
	345	0.032
	345	0.032
Mean	360	0.032
s	90.25	0.000316

s = Standard deviation

Table 19b. MSE values for pixel by pixel FFNNs trained with 3300 training samples.

	Iterations	MSE
	754	0.032
	667	0.032
	442	0.032
	769	0.033
	364	0.032
	283	0.032
	492	0.032
	749	0.032
	667	0.032
	584	0.032
Mean	577	0.032
s	174	0.000316

s = Standard deviation

Window 3 by 3 FFNN Matlab code trained with 2400 and 3300 training samples.

```
net_sample = newff(minmax(p_data2_n [2400 or 3300]), [49, 8] ,{'logsig', 'logsig'},
'trainrp');
net_sample.trainParam.show = 50;
net_sample.trainParam.epochs = 5000;
net_sample.trainParam.goal = 0.01;
net_sample.trainParam.max_fail = 20;
net_sample.layers{1}.initFcn = 'initwb';
net_sample.inputWeights{1,1}.initFcn = 'myinit';
net_sample = init(net_sample);
net_trained = train(net_sample, p_data2_n, t_data2_n [2400 or 3300], [], [],
VV_neighbor);
estimated_test = sim(net_trained, p_data2_n);
input_profile = t_data2_n - estimated_test;
perf_input_profile = mse(input_profile);
```

Table 19c. MSE values for window 3 by 3 FFNNs trained with 2400 training samples.

	Iterations	MSE
	178	0.025
	150	0.027
	237	0.024
	163	0.026
	174	0.026
	300	0.024
	182	0.025
	134	0.026
	224	0.024
	167	0.025
Mean	191	0.025
s	49.34	0.001033

s = Standard deviation

Table 19d. MSE values for window 3 by 3 FFNNs trained with 3300 training samples.

	Iterations	MSE
	243	0.024
	280	0.023
	166	0.026
	257	0.023
	366	0.022
	271	0.022
	141	0.027
	155	0.026
	201	0.025
	183	0.026
Mean	191	0.024
s	70.02	0.001838

ANNEX VII: DISTRIBUTION OF TRAINED SAMPLES PER FRACTION RANGE

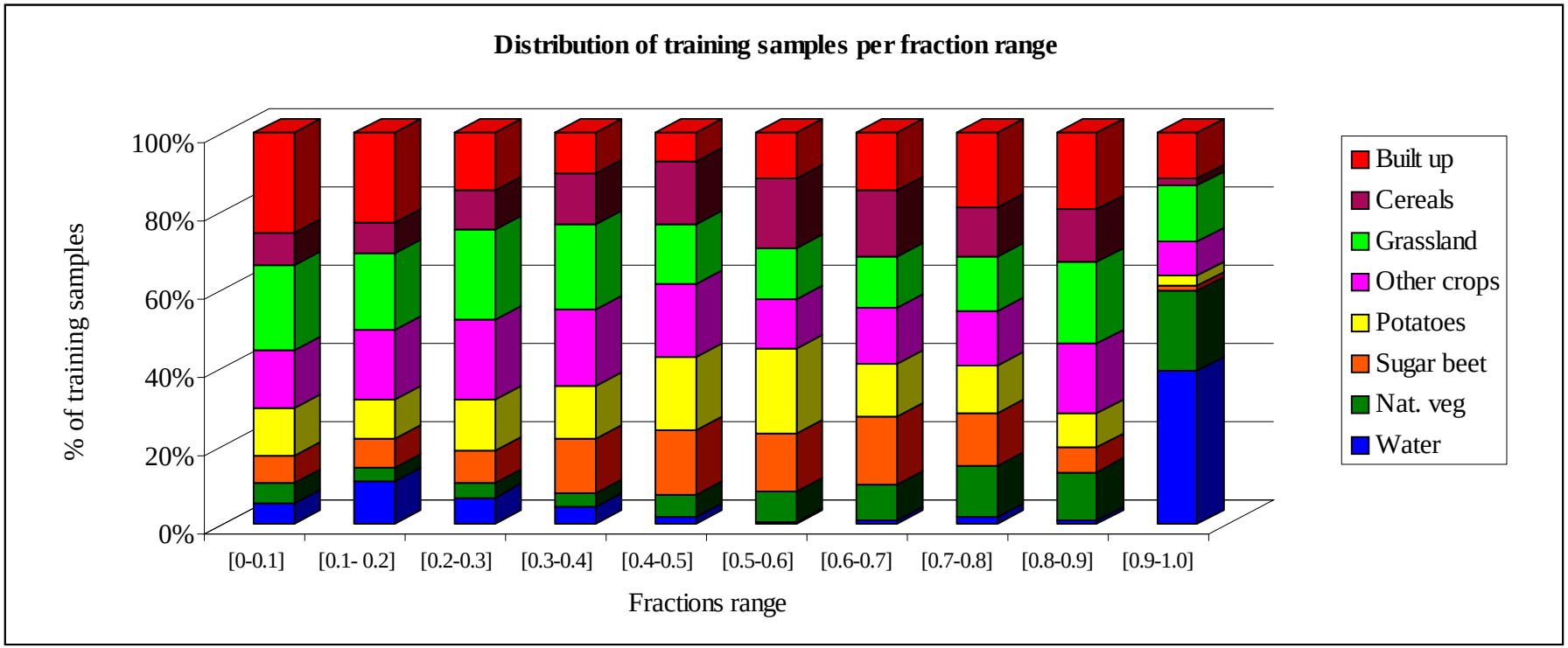


Figure 23. Distribution of training samples per fractions.