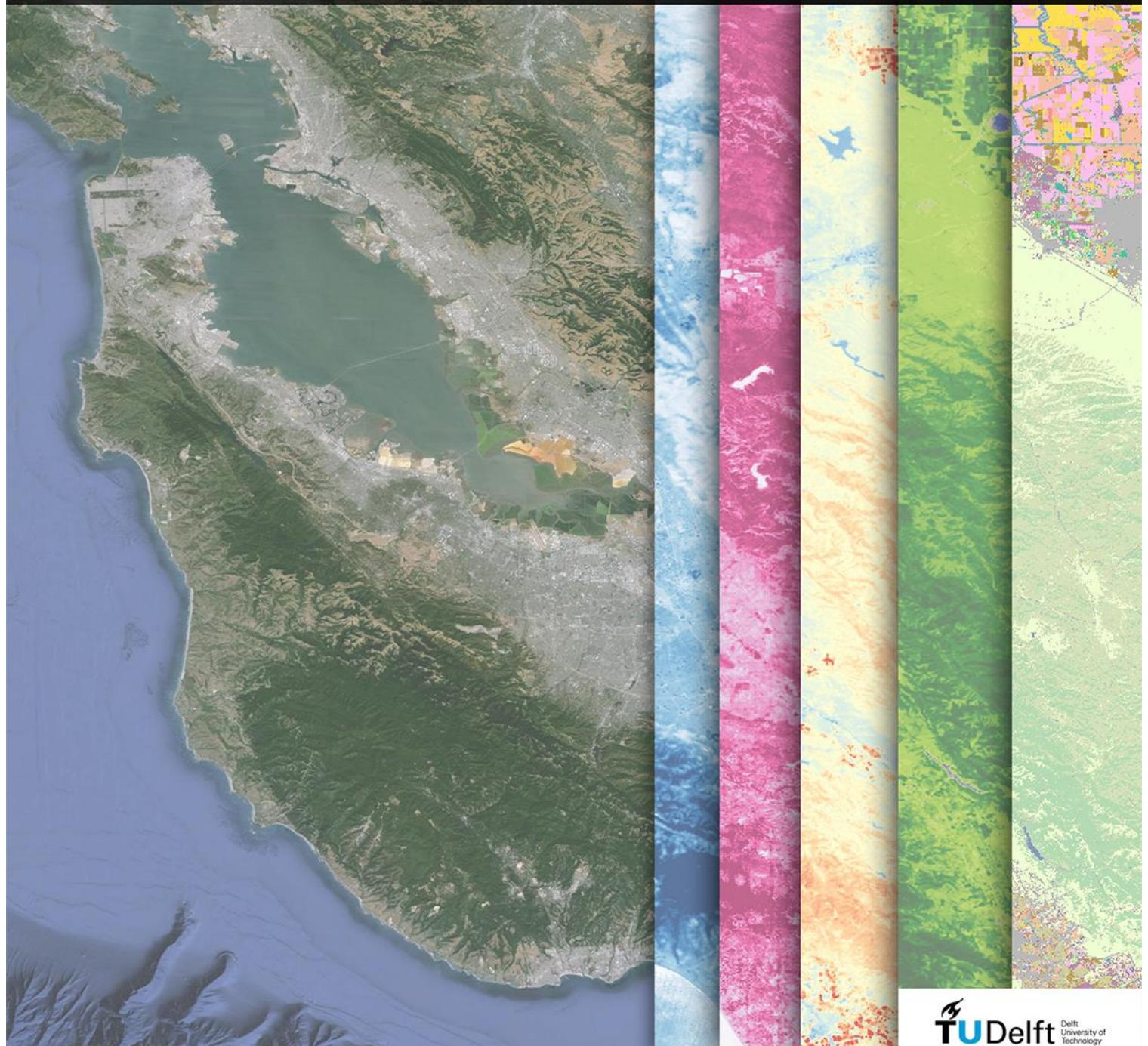




Land and water use classification by means of spectral index based time series analysis

Damien Allen



Land and water use classification by means of spectral index based time series analysis

By
Damien Allen

MASTER OF SCIENCE THESIS

For the degree of Master of Science in Civil Engineering

26 June 2015

Graduation Committee:

Prof. dr. Wim G.M. Bastiaanssen
Prof. dr. Massimo Menenti
dr.ir. Martine Rutten

Faculty of Civil Engineering and Geosciences
Delft University of Technology



Abstract

Currently available land and water use maps don't meet the needs of hydrologists and water managers in many cases, particularly in developing countries. Existing maps often lack the resolution and detail required by hydrological models, and those that do meet requirements are limited in spatial coverage. The research presented in this report attempts detailed land and water use classification in the Netherlands and California using time series analysis of spectral indices generated from reflectance data of the European Space Agency's PROVBA-V. The satellite combines high temporal resolution (2-3 day revisit time) with moderate spatial resolution (100 m pixels) which is especially attractive for coverage at the scale of river basins.

The Harmonic Analysis of Time Series (HANTS) algorithm was used to perform Fourier transformation of the selected spectral indices resulting in frequency domain harmonic component maps and gap-free, cloud-corrected reconstructed time series images. Two approaches for classification were attempted: unsupervised classification of the harmonic component maps after undergoing principle component analysis and supervised classification of time series index curves using a root-mean-square (RMSE) error minimization approach. Neither approach was able to reproduce the ground truth maps used in either location with an accuracy much higher than 50%. Both approaches were also outperformed by traditional unsupervised classification using reflectance data from 3 cloud-free images acquired during different seasons. Thus, the hypothesis that time series with very high frequency would be a suitable source for the determination of water and land use classes in river basins appeared false for the approaches used in this report.

The results found that accuracy of classification using time series data was limited by low variation of time series index curves between classes and high variation within classes. While the high temporal resolution data used in this research may be useful in vegetation monitoring of one particular agro-ecosystem, the limitations posed by time series curve variability suggest that applications in land and water use classification of river basins with a heterogeneous land use is limited and perhaps a higher spectral resolution would prove more useful. The results also found that accuracy was also lowered by strong seasonality in the Netherlands and that classification performed better in larger contiguous areas less affected by the mixed-pixel effect.

In some instances the RMSE classifier output seemed to exhibit correlation to vegetative and water features in richer detail than the ground truth map it was compared to. Thus, recommendations for alternate ground truth sources were suggested. Overall, the method showed some promise but will likely need to be fine-tuned for better performance and may be better suited for detailed differentiation of hydrological behavior within a certain land use classes, rather than as a first-pass classifier. It may be worth investigating if this technique can be used to improve the performance of existing classification techniques, perhaps by contextual application within a decision tree classifier. Notably, the mapping of vegetation cover, soil moisture and development intensity by spectral indices could help to classify the heterogeneous landscape of a river basin into certain land use classes with similar hydrological characteristics.

Table of Contents

Abstract	5
List of Figures	8
List of Tables	10
1 Introduction	13
1.1 Remote Sensing.....	13
1.2 Land and Water Use Classification for Water Accounting	14
1.2.1 Importance of Land Use Data.....	14
1.2.2 Growing Need for Detail.....	15
1.2.3 Land Use in the Hydrological Cycle.....	17
1.2.4 Transparency and Objectivity.....	18
1.3 Sensor Selection	19
1.3.1 PROBA-V.....	19
1.3.2 Other Sensors	20
1.4 Spectral Indices	21
1.5 Land Cover Classification.....	21
1.6 Research Objective.....	22
2 Study Areas	23
2.1 Bay Area & Upper San Joaquin Valley, California.....	23
2.2 The Netherlands.....	23
2.3 Water Availability and Climate.....	24
3 Methodology.....	25
3.1 Preliminary Work and Literature Review	25
3.2 Materials	25
3.2.1 Software	25
3.2.2 Data	25
3.3 Spectral Index Calculations	26
3.3.1 NDVI.....	26
3.3.2 NDWI	26
3.3.3 CI.....	27
3.3.4 MSAVI2	27
3.3.5 VNIROI	27
3.4 Ground Truth Data	28
3.4.1 The Netherlands	28
3.4.2 California	29
3.5 Harmonic Analysis	30
3.5.1 Fourier Transformations.....	30
3.5.2 Harmonic ANalysis of Time Series Algorithm and Application	30
3.6 Classification	32
3.6.1 Unsupervised Maximum Likelihood Classification	32
3.6.2 Spectrot temporal RMSE Classification	34
4 Results.....	37

4.1	HANTS Performance.....	37
4.1.1	Cloud Removal.....	37
4.1.2	Error Compared to Raw Index Values	38
4.2	Unsupervised Classification	39
4.2.1	Using Combinations of Indices	41
4.2.2	Attempts to Improve Accuracy.....	42
4.2.3	Comparison to Traditional Unsupervised Classification	44
4.3	Spectrotemporal RMSE Classification	45
4.3.1	Attempts to Improve Accuracy.....	47
5	Discussion & Conclusions.....	49
5.1	Harmonic Analysis.....	49
5.1.1	Number of Frequencies	49
5.1.2	Principal Component Analysis	49
5.1.3	Fourier Transforms for Gap-Filling	49
5.2	Classification	50
5.2.1	Unsupervised Classification.....	50
5.2.2	Supervised RMSE Classification	51
5.3	Uniqueness of Time-Series Curves both Within and Between Classes	53
5.3.1	Differences within Classes.....	53
5.3.2	Similarities between Classes	56
5.4	Experimental 100m PROBA-V Data versus Synthesis Products	56
6	Recommendations	57
6.1	Higher Spectral Resolution.....	57
6.2	A Focus on the Details.....	57
6.3	Selection of Appropriate Study Areas	57
6.4	RMSE Classifier.....	58
6.5	Decision Trees	58
6.6	More Relevant Ground Truth Data	59
7	References	61
8	Appendix	63
8.1	Parameters for Unsupervised Classification	63
8.2	Ground Truth Legends	63
8.2.1	The Netherlands	64
8.2.2	California	68
8.3	Performance of Different Index Combinations	70
8.4	In-class Variation	72
8.5	Similarity Matrices	74
8.6	GRASS Modules	90
8.6.1	probav_prep.py.....	90
8.6.2	unsup_classifier.py and error_kappa.py	90
8.6.3	sig_classifier.py.....	91
8.6.4	similarity_matrix.py.....	91
8.6.5	calc_class_var.py	91

List of Figures

Figure 1. Common geocentric satellite orbits (NASA Earth Observatory)	13
Figure 2. Comparison of detail between GLOBCOVER 2009 (left) and CDL 2014 (right) in Stockton, CA.....	15
Figure 3. Land use classes categorized in respect to vegetation cover, moisture and development intensity....	17
Figure 4. Examples of interaction between land use and the hydrological cycle.....	17
Figure 5. Artist's depiction of the PROBA-V satellite (VITO)	19
Figure 6. Spatial resolution and revisit time of select sensors	20
Figure 7. Reflectance of various tree species (Clark et al. 2007)	21
Figure 8. California study area	23
Figure 9. The Netherlands	23
Figure 10. Monthly climate data from Modesto, CA (CIMIS 2015)	24
Figure 11. Monthly climate data from De Bilt, NL (KNMI 2015).....	24
Figure 12. Frequency distribution of class areas as a percentage of total area for The Netherlands (left) and the combined BRP/BBG dataset (right)	28
Figure 13. Frequency distribution of class areas as a percentage of total area for California (left) and the CDL dataset (right)	29
Figure 14. Decomposition of an original time-domain signal into the frequency-domain (Stack Exchange).....	30
Figure 15. HANTS curve fitting using 1st, 2nd and 3rd order harmonics (left, center, and right respectively) of the NDVI for example corn pixel near Stockton, CA	31
Figure 16. Transformation of data from original variables (VAR 1, 2 & 3) in 3-D space to principal components (PC 1 & 2) in component space (Scholz 2006)	32
Figure 17. Example mean value plots for barley/corn double cropping plots in California. The shaded bands show ± 1 standard deviation.	35
Figure 18. Demonstration of cloud removal capabilities of HANTS in the Netherlands on 20 September 2014 NDVI map.....	37
Figure 19. Example HANTS curve fitting of NDVI for a grassland pixel in Limburg, NL using 3 rd order harmonic components.....	38
Figure 20. Frequency distributions for day of year experiencing peak vegetation. This is determined by the phase of the first order harmonic component for NDVI in both the Netherlands (left) and California (right) ...	40
Figure 21. Example of masking contiguous areas. Maps from left to right are the CDL ground truth, 100m mask, 300m mask and 500m mask. Areas in white are includes and those in black are excluded.	42

Figure 22. Example of class reduction in an area near Stockton, CA. Note that most class combinations are agricultural areas.....	43
Figure 23. Difference in classification accuracy for top result (left) compared to top 5 results (right) in South Holland. Green indicates ground truth is matched and black indicated non-match.....	45
Figure 24. Example NDVI RMSE classification results of a winter wheat pixel near Stockton, CA.	46
Figure 25. Examples of variation in standard deviation between time series NDWI curves for linseed (left, NL) and broccoli (right, CA). The shaded bands show ± 1 standard deviation.	47
Figure 26. Poor performance in discrimination different crops in Zeeland, NL. The combined BRP/BBG dataset (left), traditional unsupervised classification (center) and harmonic component based classification (right) are shown.	50
Figure 27. Maps from the CDL (left), Google Earth (center) and the RMSE classifier are compared for a golf course and a suburban area in Stockton, CA. The results of the RMSE classification don't match the CDL closely but may show richer details linked to vegetation and water features.....	52
Figure 28. Higher resolution Google Earth images of area between the golf course and suburban area.	52
Figure 29. Examples of variation in standard deviation between time series NDWI curves for garlic (left, CA), flax (center, CL) and uncultivated area (right, NL). The shaded bands show ± 1 standard deviation.	54
Figure 30. Examples of variation within classes in California for corn (left) and grapes (right)	54
Figure 31. Mean class index values for the Netherlands (left) and California (right)	55
Figure 32. PROBA-V 5-day coverage for 100m sensor (VITO).....	56
Figure 33. Summary of import parameters in land and water use classification.	59

List of Tables

Table 1. Water consumption related services.....	14
Table 2. Land and water use classifications as defined by the Water Accounting+ framework. (Karimi 2014) ...	16
Table 3. PROBA-V spectral bands	19
Table 4. Spectral indices selected for time series analysis as found in additional thesis (Allen 2014)	26
Table 5. Default HANTS parameters used for each spectral index.....	31
Table 6. HANTS values compared to cloud free days in the Netherlands. Values are averaged for each index for non-water areas in the study area.	38
Table 7. Unsupervised classification results for the Netherlands (using 53 classes). A comparison is made between using different orders of harmonics and PCA.....	39
Table 8. Unsupervised classification results for California (using 87 classes) A comparison is made between using different orders of harmonics and PCA.....	39
Table 9. Results of combining indices in the Netherlands (top 5 combinations)	41
Table 10. Results of combining indices in the California (top 5 combinations)	41
Table 11. Results of mixed-pixel effect reduction using various thresholds for minimum contiguous class areas. The percentage of land area meeting these thresholds is also supplied.	42
Table 12. Results of class reduction and number of classes used in each classification attempt	43
Table 13. Results from supervised classification with training	44
Table 14. Comparison of results with a traditional unsupervised classification approach	44
Table 15. Results of spectrotemporal RMSE approach using NDVI, NDWI and VNIROI	45
Table 16. Results using standard deviation based normalization	47
Table 17. Results with weighting for in-class vegetation peaks	48
Table 18. Results using first derivative curves compared to original curves.....	48
Table 19. Summary of unsupervised classification results and attempts to improve accuracy.....	50
Table 20. Summary of RMSE classification results and attempts to improve accuracy	51
Table 21. Limited CDL color legend	52
Table 22. Extent of in-class variation in both space (monthly standard deviation) and time (vegetation peak) for the Netherlands. Only the top 10 classes with lowest mean monthly standard deviation are listed.	53
Table 23. Extent of in-class variation in both space (monthly standard deviation) and time (vegetation peak) for California. Only the top 10 classes with lowest mean monthly standard deviation are listed.	53

Table 24. Default input parameters for cluster analysis.....	63
Table 25. Original BBG classes and reclassified categories for combined BRP/BBG	64
Table 26. Original BRP classes and reclassified categories for combined BRP/BBG	65
Table 27. Reclassification from combined BRP/BBG dataset to reduced class version.....	67
Table 28. Reclassification from CDL to reduced class version	68
Table 29. Full unsupervised classification results of different index combinations in the Netherlands	70
Table 30. Full unsupervised classification results of different index combinations in California	71
Table 31. Full table detailing in-class variation of NDVI images for the Netherlands.....	72
Table 32. Full table detailing in-class variation of NDVI images for California.	73
Table 33. NDVI similarity matrix for BRP/BBG dataset (NL)	75
Table 34. NDWI similarity matrix for BRP/BBG dataset (NL)	76
Table 35. CI similarity matrix for BRP/BBG dataset (NL).....	77
Table 36. MSAVI2 similarity matrix for BRP/BBG dataset (NL)	78
Table 37. VNIROI similarity matrix for BRP/BBG dataset (NL)	79
Table 38.NDVI similarity matrix for CDL dataset (CA), continued on following page.....	80
Table 39. NDWI similarity matrix for CDL dataset (CA), continued on following page.....	82
Table 40. CI similarity matrix for CDL dataset (CA), continued on following page	84
Table 41. MSAVI2 similarity matrix for CDL dataset (CA), continued on following page	86
Table 42. VNIROI similarity matrix for CDL dataset (CA), continued on following page.....	88

1 Introduction

1.1 Remote Sensing

From the perspective of a single human being, analyzing our surroundings is a fairly simple process. Using one or more of our senses we can determine the contents of our surroundings and can infer what use they are being put to. We can easily distinguish between urban and suburban areas just as we can between corn fields and grape vineyards. Similarly, using measurement techniques of varying accuracy and complexity, we can record information about things ranging from precipitation to energy flux to elevation. The limitation of our measurement abilities as humans is quantification of such information on a large scale, both in space and time.

If you want to know the areal extent of all corn fields in California, for example, then the task is already beyond the abilities of a single person. Even a small army would have trouble measuring the size of each field accurately over such a large area. When looking at processes that vary significantly in time and space, such as precipitation, it becomes clear that point measurements are not the best option. Achieving a reasonable spatial resolution using point measurements will quickly become prohibitively expensive and time consuming.

Thus, the usefulness of remote sensing data becomes immediately apparent. Satellite-based sensors can gather data over vast areas, reliably and unobtrusively. Areas can be measured continuously using geo-stationary orbits or periodically using other geocentric orbits.

Earth observing sensors have been gathering data for more than half a century and will likely continue to do so indefinitely. As time passes satellite based sensors improve in both capability and complexity. Furthermore, our archive of largely continuous data continues to grow, year on year, giving us the ability to look back through time. However, the data itself doesn't tell the whole story. We can't directly take the measurements we need from space, we must first find a way to link the spectral and radiometric data we receive into quantifiable measurements of the processes under study. So the question becomes: how can we process such a plethora of data into the usable information that we need?

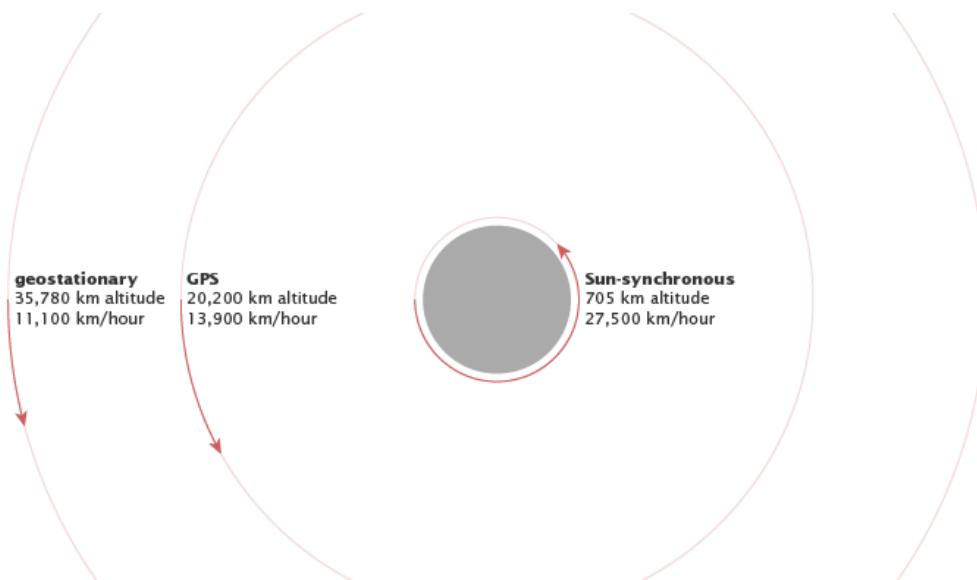


Figure 1. Common geocentric satellite orbits (NASA Earth Observatory)

1.2 Land and Water Use Classification for Water Accounting

The Water Accounting Plus (WA+) framework is being developed by the International Water Management Institute (IWMI), UNESCO-IHE and the Food and Agriculture Organization (FAO). It aims to provide a valuable and reliable source of information for policy makers in the water sector based largely on data from remote sensing. The framework communicates information on water storage and flux within a system using a number of intuitive resource sheets designed to be understood by those with technical and non-technical backgrounds alike (www.wateraccounting.org).

As we enter uncertain times where the effects of climate change are already being felt around the world, competent water management policies will become critical in maintaining sustainable water use. Even as we only begin to experience such effects, it's clear that our climate is shifting towards extremes in temperature, precipitation and weather. Just as sea level rise will test our abilities to keep coastal areas safe, water scarcity will test our ability to maintain food security. The successful interfacing between science and policy will prove invaluable in addressing these challenges and Water Accounting+ aims to help bridge the current divide.

1.2.1 Importance of Land Use Data

Land use classification is a common use for data from earth observing sensors. Disciplines ranging from climatology to ecology to hydrology all benefit from such data since it describes the services and benefits (e.g. sports park) instead of the physical land cover alone (e.g. grass). Hydrologists and water managers rely on land use data as an input for various models and in more subjective analysis of water use practices. It's worth noting that determining land use, in addition to cover type, is especially important for these users compared to the scientific community at large. Water managers look at land use in terms of water use (as well as storage and transmission) in relation to the rendered services as exemplified in Table 1.

For example, it's not always enough to know that a specific area is being used for agricultural production. Ideally, accurate estimates of evaporation, transpiration and food production depend on accurate determination of cropping patterns as well as information on whether the area is irrigated or rain-fed. Even more specifically whether it is flooded irrigation system or a modernized drip irrigation system. Since evaporation and transpiration are two of the most significant hydrological processes in terms of quantity, inaccurate data or assumptions about land use can ultimately lead to significant error propagation in the evapotranspiration sheet related to water accounting.

Table 1. Water consumption related services.

Agriculture	food, feed, fiber, fish, timber, nutrition
Environment	biodiversity, habitats, greenhouse gas mitigation, carbon sequestration, inundation
Economy	industry, harbors, navigation, small holder enterprises
Domestic	drinking, sanitation, cooking, gardening
Energy	hydropower, firewood, mines, shales
Leisure	lakes, sport fields, urban parks, silent areas

For open water, distinction between storage reservoirs, evaporation ponds and managed wetlands is equally important because they have different functions: aquaculture occurs in evaporation ponds and biodiversity is a benefit from having wetlands. Assessing successes and failures in river basin water management thus requires information on the rendered services, which are related to very specific land use categories. In urban areas, acquiring more detailed information about the surface can provide invaluable information on water runoff characteristics, flood risks and the various functions of urban areas (e.g. living, leisure, economy, transport, and environment). Furthermore, information about density of residential areas can provide better water use estimates for domestic purposes and identifying industrial areas can help address potential issues with water quality.

1.2.2 Growing Need for Detail

Karimi (2014) outlined the development of the Water Accounting+ framework which separates land use and land cover (LULC) into four distinct categories: managed water use, modified land use, utilized land use and protected land use. The framework emphasizes the importance in differentiating land cover (such as cropland) with land use (such as irrigated banana plantations). Managed water use involves manipulation of the water cycle by physical infrastructure, such as irrigation works and domestic water use. Modified land use describes areas greatly modified by human activity including rain-fed agriculture and parks. Anthropogenic changes in vegetation cover and soil fertility will create a change in evaporation and runoff processes, hence the amount of water consumed and amount of water available for downstream users without transporting water through structures. Utilized land use includes land with natural land cover that provides service with little interference including grassland used for grazing and sustainable timber harvest. Areas set aside for conservation, such as national parks, are examples of protected land use.

Existing land cover maps tend to have a number of trade-offs for this particular application. Most global land cover datasets provide a limited number of distinct classes, usually focused on vegetative cover. Highly detailed maps, both in resolution and number of classes, are typically only available in limited areas where affordability of generating such data is not an issue. Additionally, such maps may be held as proprietary information and offered at significant cost.

Figure 2 provides a visual comparison of GLOBCOVER, a commonly used global land cover dataset and the Crop Data Layer (CDL) from the US Department of Agriculture. The level of detail in the CDL is achieved using a combination of threshold-based remote sensing classification trained with ground survey data and classified government data sets. Such detailed datasets are only available in a limited number of developed countries who have enough resources to produce the data. Many other countries are limited to the class variety of GLOBCOVER or similar products.

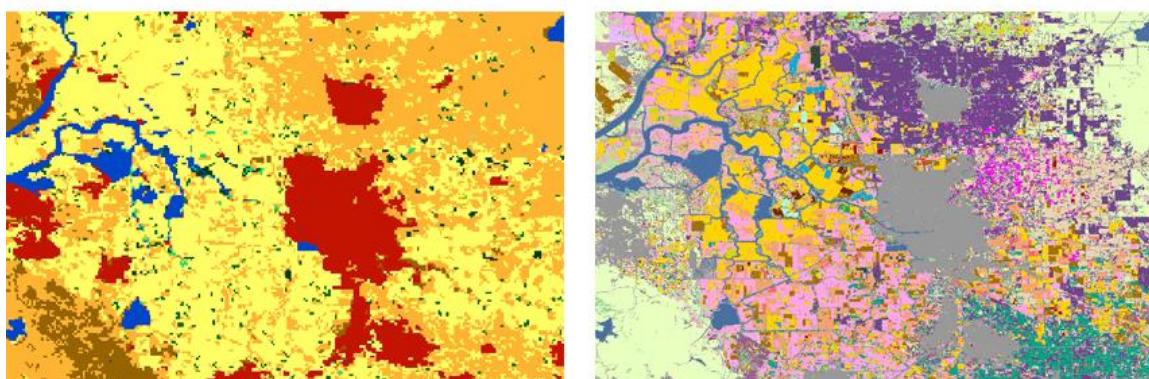


Figure 2. Comparison of detail between GLOBCOVER 2009 (left) and CDL 2014 (right) in Stockton, CA

Although such land and water use maps are not yet universally available, the Water Accounting+ framework defines a number of classes that would be most useful under its applications. Table 2 below lists specific land and water use types and their categorization under the major land use types defined by WA+. This categorization continues to evolve along with the development of the WA+ framework, but the need for a reliable data source capable of such detail remains.

Table 2. Land and water use classifications as defined by the Water Accounting+ framework. (Karimi 2014)

Protected land use	Utilized land use	Modified land use	Managed water use
<ul style="list-style-type: none"> • Protected forests • Protected shrubland • Protected natural grasslands • Protected natural waterbodies • Protected wetlands • Glaciers • Protected other 	<ul style="list-style-type: none"> • Closed natural forests • Tropical rainforests • Open natural forests • Woody savanna • Open savanna • Sparse savanna • Shrub land & mesquite • Herbaceous cover • Meadows & open grassland • Riparian corridors • Deserts (low rainfall) • Wadis • Natural alpine pastures • Rocks, gravel, stones & boulders • Permafrosts • Brooks, rivers & waterfalls • Natural lakes • Flood plains & mudflats • Saline sinks, playas & salinized soil • Bare soil permanent • Wasteland • Moore fields • Wetland & swamps • Mangroves • Alien invasive species 	<ul style="list-style-type: none"> • Forest plantations • Mixed species agro-forestry • Rainfed production pastures • Rainfed food crops – cereals • Rainfed food crops – non-cereals • Rainfed fruits • Rainfed vegetables • Rainfed beverage crop (coffee, tea) • Rainfed non-food crop (biofuels, cotton, quat, oilpalm) • Rainfed parks (leisure & sports) • Fallow & idle land • Dump sites & deposits • Oasis • Rural paved surfaces (lots, roads, lanes) • Rainfed industry parks – outdoor • Rainfed homesteads and gardens (urban cities) – outdoor • Rainfed homesteads and gardens (rural villages) - outdoor 	<ul style="list-style-type: none"> • Irrigated production pastures • Irrigated food crops – cereals • Irrigated food crops – non-cereals • Irrigated fruits • Irrigated vegetables • Irrigated forest plantation • Irrigated non-food crops (biofuels, cotton, quat, oilpalm) • Managed water bodies (reservoirs, canals, harbors, tanks) • Greenhouses – indoor • Aquaculture • Domestic households – indoor (sanitation) • Manufacturing & commercial industry – indoor • Irrigation homesteads and gardens – outdoor • Irrigated industry parks – outdoor • Irrigated parks (leisure & sports) • Livestock & domestic husbandry • Managed wetlands & swamps • Managed other inundation areas • Mining/quarry & shale exploration • Evaporation ponds • Water treatment plants • Hydropower plants • Power plants (thermal, coil, nuclear)

1.2.3 Land Use in the Hydrological Cycle

The unmet need for detailed land and water use data by the Water Accounting+ Framework highlights the importance of the link between land use and the hydrological cycle.

Consider how land and water use can be categorized by the following: moisture, vegetation cover and development intensity. Figure 3 provides some examples of how land use classes would fit into these categories. Development intensity forms a tertiary axis which could correlate to the increased potential water demand and stress on water quality due to heavier industrialization or population density. Moisture is largely dependent on water storage in the area, irrigation supply and precipitation. Vegetation cover depends both on the availability of water and the energy needed for crop growth. Vegetation patterns will have a large effect on evaporation through transpiration, a key component of the hydrological cycle.

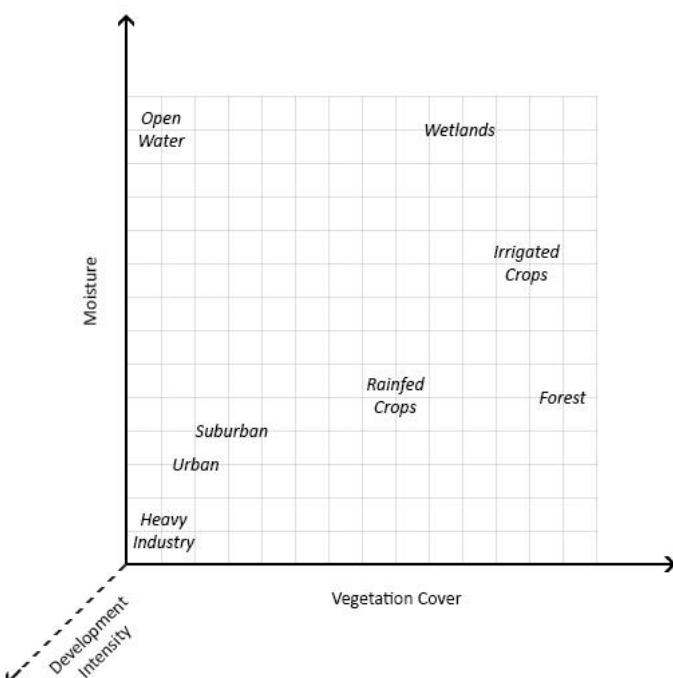


Figure 3. Land use classes categorized in respect to vegetation cover, moisture and development intensity

There is also a temporal component to this link between land use and hydrology which serves as a motivation for the time series analysis approach applied in this research. Figure 4 displays a few examples of how this may manifest. Some areas will be permanent “wet zones” characterized by high groundwater such wetlands and riparian zones which lie adjacent to perennial water bodies. On the other hand, the wetness of irrigated crop land is largely dependent on irrigation practices, especially in areas with low precipitation. Wetness of natural areas further from open water and shallow groundwater will experience variation in wetness more directly linked to climate. The levels will vary as precipitation and potential evaporation fluctuate seasonally. By mapping vegetation cover, soil moisture and development intensity, it will be feasible to classify the heterogeneous landscape of a river basin into certain land use classes with similar hydrological characteristics.

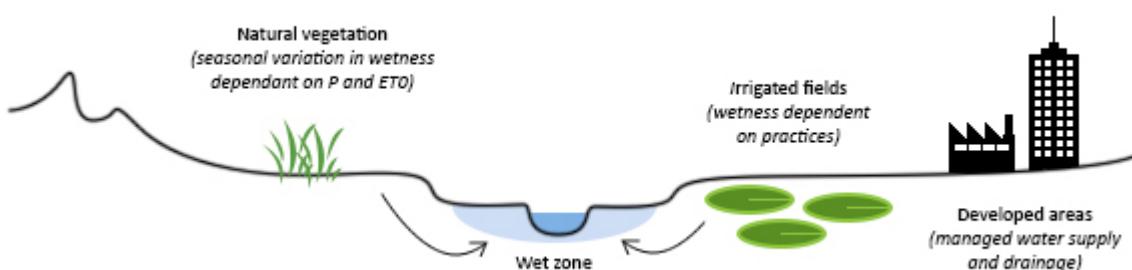


Figure 4. Examples of interaction between land use and the hydrological cycle

1.2.4 Transparency and Objectivity

The Water Accounting+ framework focuses on the use of public access data in an effort to maintain a high level of transparency. This is especially important considering that the framework is applied on a river basin scale which can often span multiple countries. In areas experiencing water scarcity, tensions over water use can run high. Some countries may be unwilling to share data with their neighbors, or data that is provided may be seen as untrustworthy since conflicts of interest could potentially lead to data falsification.

Luckily, remote sensing can be a reliable source of data as far as objectivity is concerned. Imagery from government satellites is often publically available, thus alleviating concerns over proprietary commercial data which may be inaccessible to some parties. Data products from the National Aeronautics and Space Administration (NASA) and European Space Agency (ESA) are often provided completely free of charge for all users regardless of nationality or intended application. This allows for creation of precipitation, evapotranspiration, net primary production and land use datasets, among others, that can also be provided as public domain.

1.3 Sensor Selection

1.3.1 PROBA-V

The PROBA-V is a recent earth observation mission commissioned by the European Space Agency with the Flemish Institute for Technological Research (VITO) responsible for data processing and distribution. Based on the Project for On-Board Autonomy (PROBA) platform developed for ESA's MicroSat program, the satellite is designed to serve as a successor to the vegetation instruments aboard the SPOT-4 and SPOT-5 satellites. Its visible and infrared bands, summarized in Table 3, closely match those of the SPOT VEGETATION sensors and thus allows for data continuity between the two satellite programs and future missions such as the Sentinel series. Although the variety of available bands is limited, the most critical bands for vegetation monitoring are present.

Table 3. PROBA-V spectral bands

Band	Spectrum	Wavelength (nm)	Bandwidth (nm)
1	BLUE	447-493	46
2	RED	610-690	80
3	NIR	777-893	116
4	SWIR	1570-1650	80

The PROBA-V provides daily measurements at 333 m resolution, an improvement over the 1km resolution of the SPOT VEGETATION sensor. Additionally, a 100m product provides full coverage every 2-5 days depending on altitude. The variation in revisit times between the sensors is due to the difference in swath width. Thus, the 100 m sensor with a smaller swath needs more passes to capture the same area as the 333 m sensor. Although the 100 m PROBA-V product possesses an ideal balance of resolution and revisit time, the first data was released in March 2014 which limits the possible applications of the data to short time series. The current amount of available PROBA-V data is inadequate for the purpose of land and water use change over multiple years, but can provide insights for the 2014 growing season. For the scope of the research presented in this report a full year of data is sufficient and the balance of resolution and revisit time make it a prime candidate for time series analysis compared to other available sensors.

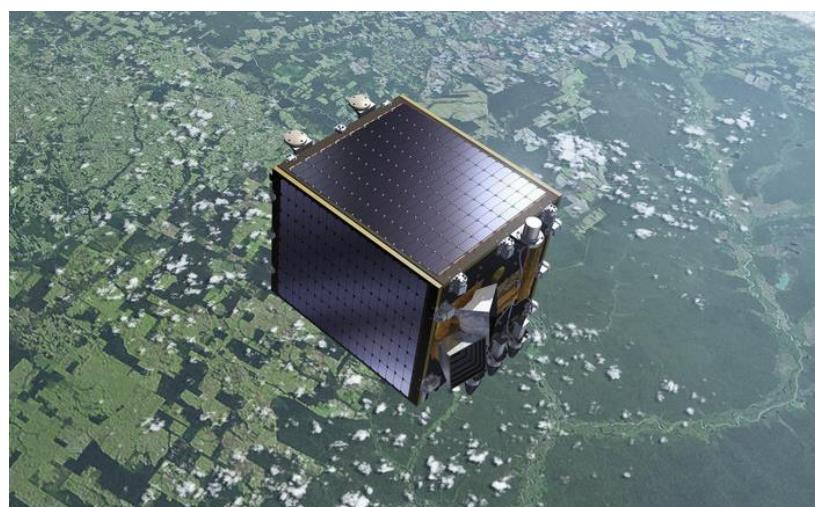


Figure 5. Artist's depiction of the PROBA-V satellite (VITO)

1.3.2 Other Sensors

There are a number of different earth observing sensors with similar roles to the PROBA-V. The main differences between the sensors are spectral properties, revisit time and spatial resolution. Figure 6 below, presented on a semi-log plot, illustrates the typical balance between spatial resolution and revisit time and highlights the PROBA-V's niche among currently available sensors.

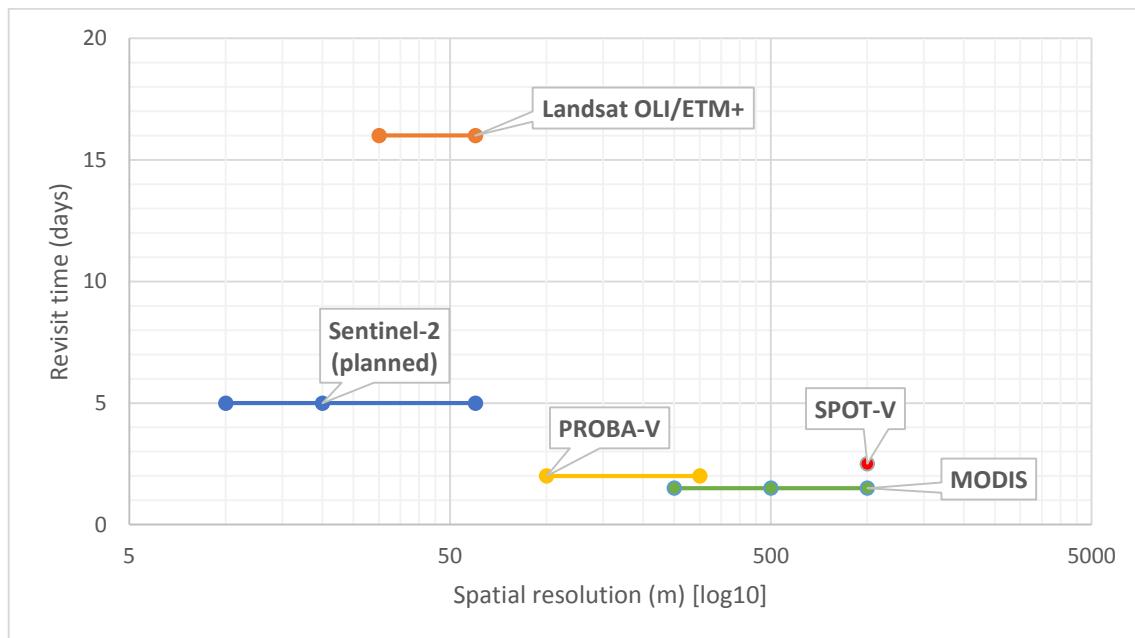


Figure 6. Spatial resolution and revisit time of select sensors

Sentinel-2

The ESA has planned the launch of two Sentinel-2 Earth observation satellites in the years 2015-2016, the first of which is expected in June 2015. The satellites will continue to provide free, high-quality data for scientific research as well as emergency planning purposes. These missions will provide spectral data across 13 bands in the visible, near infra-red and shortwave infra-red parts of the spectrum. The various sensors will provide data at either 10 m, 20 m or 60 m resolution with a revisit time of 2-5 days depending on the latitude of the capture area (ESA 2010). Since the PROBA-V was designed to provide data continuity between earlier SPOT sensors and Sentinel-2, methods developed for the PROBA-V bands can be applied once Sentinel-2 data is available.

Landsat Program

The Landsat program consists of eight satellites which have captured millions of images since the first launch in 1972. Of the eight satellites launched, two remain active as of 2015: Landsat 7 and Landsat 8. The pair of satellites provide 15 to 100m resolution images with a revisit time of 16 days. Their sensors operate in the visible, near-infrared, shortwave infrared and thermal infrared spectrums with Landsat 7 and Landsat 8 having 8 and 11 bands respectively.

MODIS

The Moderate-resolution Imaging Spectroradiometer (MODIS) instrument is also an important component of NASA's Earth Observing System. The sensor was installed onboard the Terra and Aqua satellites which were launched in 1999 and 2002 respectively. The draw of these sensors is that they record data over 36 spectral bands at resolutions varying from 250m to 1km. The pair of sensors is able to capture the entire Earth's surface every 1 to 2 days.

1.4 Spectral Indices

Spectral indices are transformations of raw multi-spectral data into indicators which have correlation to measurable properties and processes. Traditional spectral indices include linear combinations and normalized differences between various spectral bands. Numerous indices have been proposed that correct for anomalies introduced by atmospheric interference and soil background noise. In the case of vegetation, the difference between low reflectance in the red band and very high reflectance in the near-infrared band is often exploited. The relationship between these two bands (the second and third shaded areas) can be seen in the first rapid increase in reflectance shown in Figure 7.

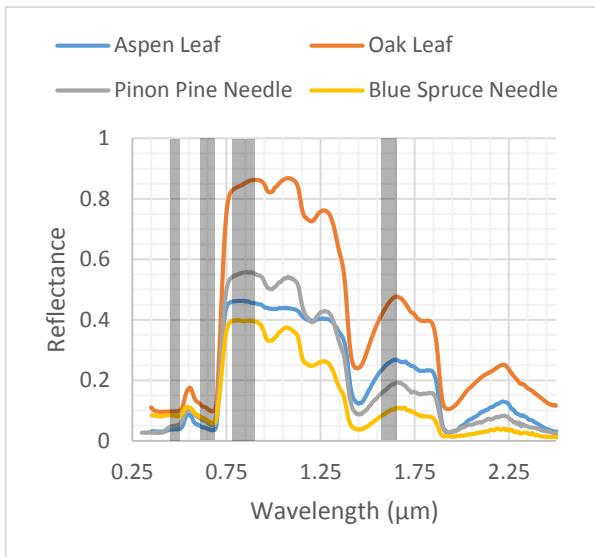


Figure 7. Reflectance of various tree species (Clark et al. 2007)

In this research, spectral indices were chosen over raw band reflectance values due, in part, to the limited spectral bands available on the PROBA-V which are shaded grey in the figure. The indices were seen as a means to maximize the usefulness of information from the four available bands.

1.5 Land Cover Classification

There are currently three dominant approaches to land cover classification by means of remote sensing: statistical methods, artificial neural networks and decision trees. Inputs for land cover classification algorithms are largely based on multi-spectral reflectance data from earth observing sensors, however supplemental data such as digital elevation maps are commonly used. These algorithms are usually trained using point sourced data, land parcel information and survey data.

Statistics based classifiers often make use of a maximum likelihood estimation. First, an image must be separated into a number of similar areas which is usually performed using a clustering algorithm, training areas or manually using expert judgement. Signatures for these areas are then calculated and passed on to the maximum likelihood classifier. The classifier assumes a chosen probability density function and determines which class each pixel has the highest probability of belonging to. Thus, data lacking a distinct and unified probability density function will likely see lower classification accuracy using statistical methods (Pal & Mather 2003).

Artificial neural networks are statistics based machine learning methods used across a variety of different fields. The general concept of such models is inspired by biological neural networks such as brains. A modeled network of “neurons” is finely tuned based on the application resulting in an adaptive network which is capable of learning. In the case of remote sensing based classification the method is limited by the need for detailed network configuration, parameter specification and extended training (Civco 1993).

The use of decision tree based classifiers is becoming increasingly popular. Decision trees are hierarchical rule sets with flow-chart like structures. Dating mining techniques allow for the creating of decision trees based on a set of inputs and accurate training data. However, such classifiers may be hindered by their dependence on the training set which, in the case of land use classification, limits their effectiveness to smaller geographic areas (Friedl & Brodley 1997).

1.6 Research Objective

This report seeks to explore the usefulness of time series analysis of spectral indices in the discrimination and classification of different land and water use areas. The hypothesis is that different land and water use types will exhibit unique time series spectral signatures that will allow for detailed classification. Such signatures are expected to result from plant phenology and interactions with hydrological processes as well as differential spectral response to seasonal and environmental forcing. Variation of moisture in both space and time needs to be examined, thus, consideration of climatic factors is a must since they have a large impact on the precipitation and potential evaporation.

This report aims to investigate the following research question:

Is land and water use classification by means of Fourier time series analysis of multiple spectral indices based on high temporal resolution PROBA-V images viable for detailed, unsupervised land and water use classification?

2 Study Areas

This report will focus on two different study areas: the California Bay Area including the upper San Joaquin Valley and the entirety of the Netherlands. These two areas were selected due to the variety in climate, hydrological processes, and land use patterns as well as access to reasonably reliable existing land use data.

2.1 Bay Area & Upper San Joaquin Valley, California

The California Bay Area contains some of the largest cities in California including San Francisco, Oakland, and Sacramento. This area includes the river deltas for both the San Joaquin and Sacramento River systems. It lies adjacent to the upper end of the San Joaquin Valley, a major source of agricultural production in California and the country as a whole. The major crops in the area include grapes, nuts, tomatoes and corn. The average plot size is 312 ha. The Central California Valley, which encompasses the San Joaquin Valley is one of the most productive agricultural areas in the world with more than 230 varieties of cultivated crops (USDA 2012).



Figure 8. California study area

The study area, which is generally water stressed, is now facing one of the worst droughts in recorded history in 2015. Irrigation in the valley is a very common process to ensure sufficient moisture availability for crop evaporation. The landscape is composed of rolling hills with natural vegetation ranging from grasslands and shrub lands to redwood forests. The selected study area lies between 39° and 36° north latitude and -123° to -119° east longitude.

2.2 The Netherlands

The climate of the Netherlands stands in stark contrast to that of Central California. A combination of multiple river deltas, a high groundwater table and moderate precipitation provides the country with an abundance of water. A water layer of 300 mm must be pumped to the sea each year to prevent the lowlands from flooding. The landscape consists of dense urban areas nestled between grasslands, forests and agricultural areas.



Figure 9. The Netherlands

The Netherlands is the second largest exporter of agricultural products in the world behind the United States. Major crops grown on arable land include potatoes, wheat, corn, sugar beets and onions. Additionally, a large number of crops are grown in greenhouses including tomatoes, peppers and cucumbers. The average plot size is 25 ha (Martins 2008). Furthermore, the Netherlands is a worldwide leader in the export of cut flowers and bulbs. The images of the Netherlands fall between 53.75° and 50.75° north latitude and 3.25° to 7.25° east longitude. The area is later masked only to include the territory of the Netherlands.

2.3 Water Availability and Climate

There are two major components of the water balance that provide a clear picture of water availability and climate in a given area: precipitation and evaporation. Figure 10 and Figure 11 show monthly total values for precipitation (P) and reference evaporation (ET_0) for each study area over the study period.

Data for the California study area was obtained by the CIMIS station in Modesto, California, a location which is fairly central in the study area. The station recorded a total reference evaporation of 1412mm over the study period with only 104 mm of precipitation (CIMIS 2015). This is in high contrast to 921 mm of precipitation recorded in De Bilt, Netherlands, over the same time period with 606 mm of reference evaporation (KNMI 2015). These figures are consistent with the general knowledge about these locations: California faces water scarcity and relies heavily on irrigation while the Netherlands has a surplus of water which must be drained from the country using pumping stations. This contrast will allow for some contemplation on the effects of climate on this study.

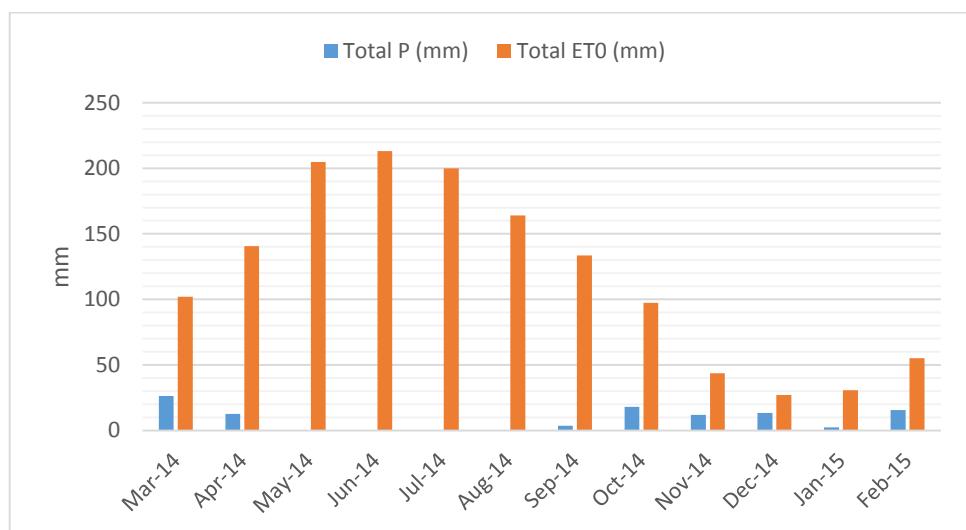


Figure 10. Monthly climate data from Modesto, CA (CIMIS 2015)

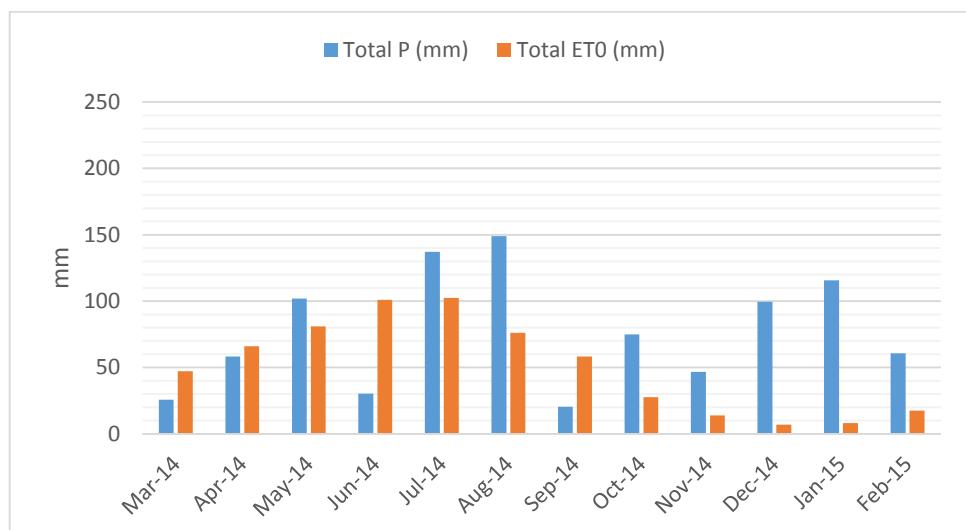


Figure 11. Monthly climate data from De Bilt, NL (KNMI 2015)

3 Methodology

3.1 Preliminary Work and Literature Review

This report was preceded by a preliminary literature review titled *Identifying appropriate spectral indices for land and water use classification by means of time series analysis* (see Allen 2014). This review outlined the selection of spectral indices used in this report while also providing further background information on spectral analysis. The review found that the selected spectral indices, shown in the table below, had strong correlation to bio- and soil-physical parameters and were thus appropriate for further study. The report also provided justification for use 100 m PROBA-V data and concluded that no other freely available satellite-based spectral data met the balance of resolution and revisit time useful for this research.

3.2 Materials

3.2.1 Software

This research was conducted using GRASS GIS and Python, both free and open-source software platforms. GRASS version 7 was released as stable in February 2015 and provided time series analysis tools which proved extremely useful for the calculations used in this report. GRASS includes a Python API which allows for execution of GRASS commands and manipulation of returned data. More importantly, this interfacing allowed for scripted commands and batch processing which were crucial when dealing with large amount of time series data. Some classification methods required the processing of thousands of raster images, including maps required for intermediate calculation steps and those used for temporary data storage. Various plots included in this report were generated using the matplotlib plotting library for Python.

3.2.2 Data

Spectral data from the PROBA-V was acquired from VITO as a part of their limited release of experimental 100m data from the satellite's front facing sensor. This research uses data from 16 March 2014 until 27 February 2015 which is the entirety of the 100m experimental product testing period. Given the satellite's 2-3 day revisit time at the moderate latitudes of the study areas, this amounted to 150 and 160 days for the California and Netherlands study areas respectively. This includes both complete and partial images of varying size depending on the sensor swath coverage. Complete raster images of California and the Netherlands contain roughly 10 and 4.5 million pixels respectively.

The data was provided as a level 3 product which is an atmospherically corrected top-of-canopy reflectance of each spectral band which has undergone geometric and radiometric processing. Clouded pixels are flagged during this process but are not removed.

The areas of each study area are masked before being used as input to classification algorithms. The mask excludes areas of open water. This is done to reduce calculation times and since open water is already easy to discriminate using remote sensing techniques. In the case of the Netherlands, the mask limits the area of calculation to only include the territory of the Netherlands.

3.3 Spectral Index Calculations

The top-of-canopy reflectance data was then used to create spectral index maps corresponding to the level 3 PROBA-V images which results in maps matching the temporal and spatial coverage of the level 3 products. At this stage, error from clouding and shadowing is still present so these maps only serve as an intermediate for future calculations. Table 4 below summarizes the selected indices which the following section explains in further detail.

Table 4. Spectral indices selected for time series analysis as found in additional thesis (Allen 2014)

Index	Purpose	Equation	Reference
NDVI	Vegetative cover	$NDVI = \frac{NIR - R}{NIR + R}$	Rouse et al. (1974)
NDWI/ NDBI	Vegetative moisture, built-up areas	$NDWI = -NDBI = \frac{NIR - SWIR}{NIR + SWIR}$	Goa (1996) Zha (2007)
CI	Soil crust composition	$CI = 1 - \frac{R - B}{R + B}$	Karnieli (1997)
MSAVI ₂	Leaf composition	$MSAVI_2 = \frac{2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - R)}}{2}$	Qi (1994)
VNIROI	Soil moisture, surface permeability	See below	Allen (2014)

3.3.1 NDVI

The Normalized Difference Vegetation Index (NDVI) was proposed by Rouse et al. (1974) for use in their Great Plains study using Landsat 1 data. The index remains one of the most commonly used vegetation indices and is based on previous work on normalized difference spectral indices by Kriegler et al. (1969).

NDVI is an excellent indicator of photosynthetic activity (Huete et al. 1997) and time series NDVI can provide some information on crop types if growing season duration and timing are known. However, the index has some drawbacks such as susceptibility to saturation which limits its usefulness in densely vegetated areas, and additionally, it is sensitive to canopy background variation such as altered soil conditions (Huete 1988).

3.3.2 NDWI

The normalized difference index using NIR and SWIR bands is commonly referred to as the Normalized Difference Water Index (NDWI) as proposed by Goa (1996). This combination of bands has shown correlation to vegetative moisture content. The NDWI benefits from the differential absorption of vegetative moisture in the NIR and SWIR bands and was found to be less susceptible to atmospheric variation than NDVI due to the reduced scattering of infrared light compared to visible light (Gao 1996).

The inverse form of the equation is known as the Normalized Difference Build-up Index (NDBI) was proposed by Zha et. al (2007) as means for delineating urban areas. Li and Liu (2008) found correlation between NDBI and land surface temperature and suggest that the NDBI could be used to measure urban heat island effects which suggests that the index could prove useful in classification of inhabited areas.

3.3.3 CI

Karnieli (1997) discovered that organic crusts that form on sand in arid and semi-arid regions can be detected by spectral means. One of the primary organisms in organic crusts are cyanobacteria which contain blue-sensitive phycobilin pigments and the Crust Index (CI) was devised to exploit the differential spectral response of these organisms. The Crust Index has proved useful in differentiating between dune sand, organic crusts and physical crusts, however, the visible spectrum is especially susceptible to atmospheric interference. However, for vegetated and other non-soil surfaces, the index provides no documented use.

3.3.4 MSAVI2

The second Modified Soil Adjusted Vegetation Index (MSAVI2) was introduced by Qi (1994) and eliminated the need to calculate the slopes of a soil line or the need to specify a soil brightness correction factor. This was accomplished by choosing a seed value and iterating the original MSAVI equation until soil effects were minimized.

The lack of empirical calculations makes the MSAVI2 more attractive for use across a variety of sensors. Since the soil line and soil brightness correction factor are integrated into the equation it will theoretically perform well over land cover types with varying vegetative cover. MSAVI2 was found to have greater dynamic range than NDVI with less potential for saturation (Qi 1994).

3.3.5 VNIROI

Previous studies have found direct correlation between soil moisture content and surface reflectivity (Menenti, 1984; Biehl and Stoner 1985). As soil moisture content increases, there is a decrease in reflectivity across the spectrum as previously mentioned. The proposed offset index seeks to reflect the change of moisture conditions of non-vegetated surfaces.

Since satellite data is only available for limited wavelength bands of various bandwidths, this proposed index focuses on the 400-900 nm range which is typically linear for soil and is usually covered by land observation sensors. In this study, a least squares linear regression is performed on three data points: reflectance values in the red, blue and near-infrared regions provided by the corresponding spectral bands. The offset of the resulting plot is used as new spectral indices which will be referred to as the visual and near infrared offset index (VNIROI). Wavelength in nanometers and top-of-canopy reflectance as reported by the sensor are used for the x- and y-axes respectively. The VNIROI is then scaled by a factor of 1/600 giving it a range of approximately -3.0 to 1.0 with arbitrary units. The change in overall reflectance, and possibly soil moisture content, is estimated by increases or decreases in offset. VNIROI seems to be useful for areas with bare soil and sparse vegetation.

3.4 Ground Truth Data

In order to judge the performance of classification attempts resultant maps must be compared to ground truth data. In the case of this report, the ground truth data isn't necessarily the "truth". The datasets, outlined below, are simply the best publically accessible data. This includes remote sensing data and plot survey data of varying levels of accuracy.

It's worth noting that both of these datasets are defined by traditional land use classes and, although very detailed, do not meet the needs of Water Accounting+ as outlined in Table 2. The needs of WA+ are more closely linked to hydrological processes so, in a more ideal scenario, the ground truth data would incorporate spatial information related to those processes.

3.4.1 The Netherlands

Ground truth data used for the Netherlands was based on two public domain datasets provided by the Dutch government. The Basis Registratie Percelen (BRP) Gewaspercelen dataset is published annually and provides the results of crop surveys reported by farmers to the government for the sake of environmental and tax regulations. This data is combined with the Bestand Bodem Gebruik (BBG) dataset provided by the Centraal Bureau voor de Statistiek (CBS) which provides ground cover information in non-agricultural areas not covered by the BRP. Some BRP classes which provided detailed information on soil type and crop usage were combined for the same crop types which resulted in 53 total classes for the combined dataset as detailed in appendix section 8.2.1.

The 2014 BRP and 2010 BBG data sets provide a good overview of the Dutch landscape. Roughly 40% of the country is covered by vast stretches of grassland. Developed areas account for about 15% of the country while another 15% is characterized by forest, wetlands and other natural areas. The remainder of the land is mostly used for agriculture, both in open fields and greenhouses. Much of the Netherlands is composed of polders which are low-lying areas surrounded by dikes with strictly controlled ground water levels. The polder system uses a series of ditch and canal networks interconnected with pumping stations in order to regulate water levels.

Figure 12 shows the frequency distribution of relative class sizes as well as a map of the combined BRP/BBG dataset. The plot shows the distribution of class areas as a percentage of total area for the 53 classes included in the dataset.

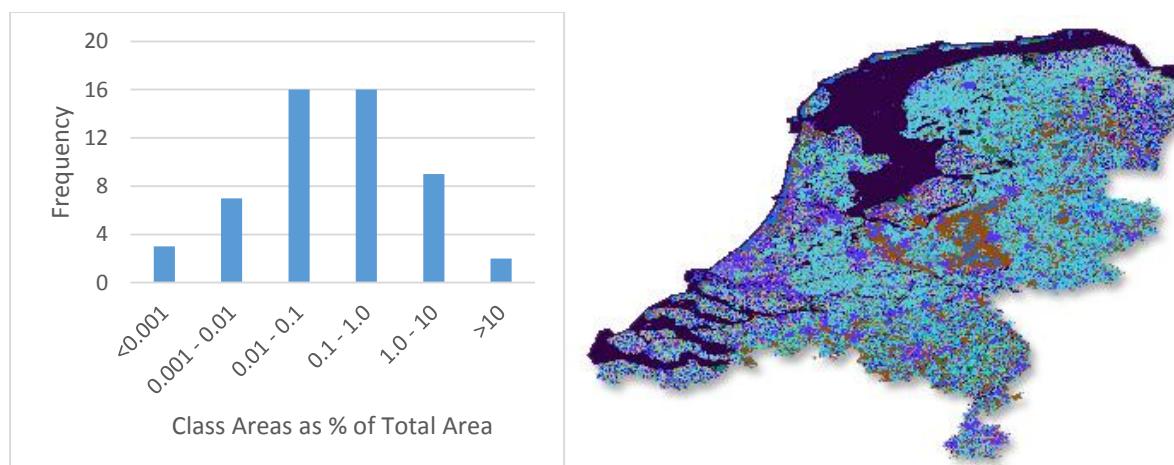


Figure 12. Frequency distribution of class areas as a percentage of total area for The Netherlands (left) and the combined BRP/BBG dataset (right)

3.4.2 California

The Crop Data Layer (CDL) was selected as the ground truth dataset for the California study area. The dataset is released on a yearly basis by the US Department of Agriculture. It is created using a decision tree classifier based on remote sensing data from various sensors including those from the Landsat program. It is worth noting that this data-mining based approach requires high accuracy training data which was sourced from farmer surveys and classified government datasets. The accuracy of the CDL in California is 85%, a figure that should be considered in terms of error propagated into the results of this report. For non-agricultural areas, the CDL draws from data in the National Land Cover Database (NLCD) which is produced every five years by the US Geological Survey. Again, a full list of CDL classes can be found in appendix section 8.2.2.

The landscape of central California, including the Bay Area and the Upper San Joaquin Valley, stand in strong contrast to that of the Netherlands. The drastic differences in water availability are immediately highlighted by the dry landscape in California. Non-irrigated areas are often comprised of arid grass- and scrublands as well as a variety of different forested areas. The bulk of agricultural land is located in the San Joaquin Valley which lies between the Sierra Nevada Range to the east and the Diablo and Temblor Ranges to the west. Large tracks of forested areas line the mountain ranges with reservoirs and vineyards intermingled in some areas. The Sacramento-San Joaquin River Delta empties into the Suisun Bay, an extension of the San Francisco Bay. This is the only major part of the study area which contains persistent wetlands. Developed areas can be found in the San Francisco Bay, along the Pacific coast and dispersed throughout the San Joaquin Valley.

Figure 13 shows the frequency distribution of relative class sizes along with the CDL dataset. The plot of frequency distribution is noticeably different than that of the Netherlands in Figure 12. There are more classes that encompass a small percentage of land area. This is likely due to the higher variety of crops in the study area and higher number of classes: a total of 87 compared to the 53 of the BRP/BBG dataset. Although the CDL provides land cover data for the 2014-2015 it is worth noting that training data may be from previous growing seasons as well as data from the 2011 NLCD.

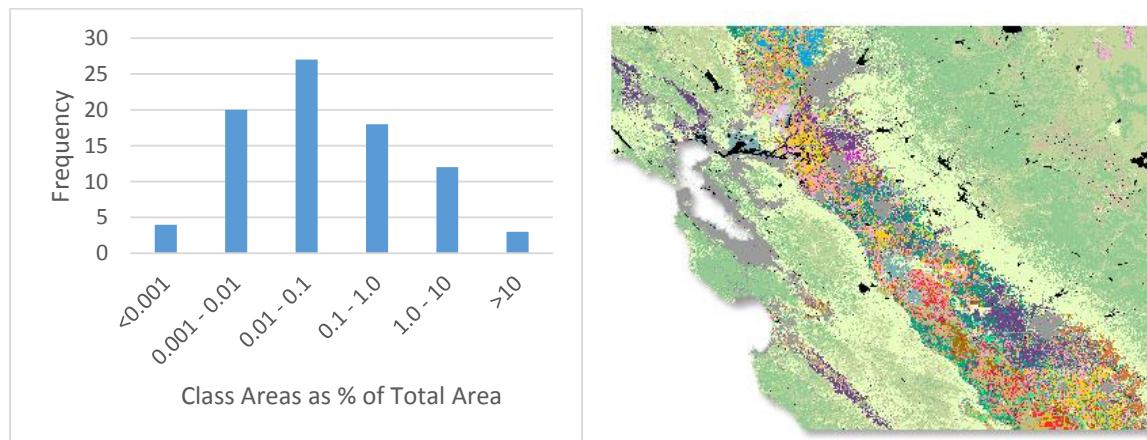


Figure 13. Frequency distribution of class areas as a percentage of total area for California (left) and the CDL dataset (right)

3.5 Harmonic Analysis

3.5.1 Fourier Transformations

This research relies on Fourier transformations which were first developed by Joseph Fourier (1822). A Fourier transformation is a decomposition of a function in time into a number of harmonics which can be summed to recreate the original curve (Bracewell 1965). In Figure 14, the red curve on the left is the original curve in the time domain and the blue plot on the left shows the approximation in the frequency domain which results from a Fourier transformation. The harmonic components in the frequency domain can be described in terms of their amplitude, periodicity and phase offset in comparison to a reference time. The number of harmonic components must be chosen depending on the application: details from the original curve will be lost with too few components or error from over-fitting may be introduced with too many components.

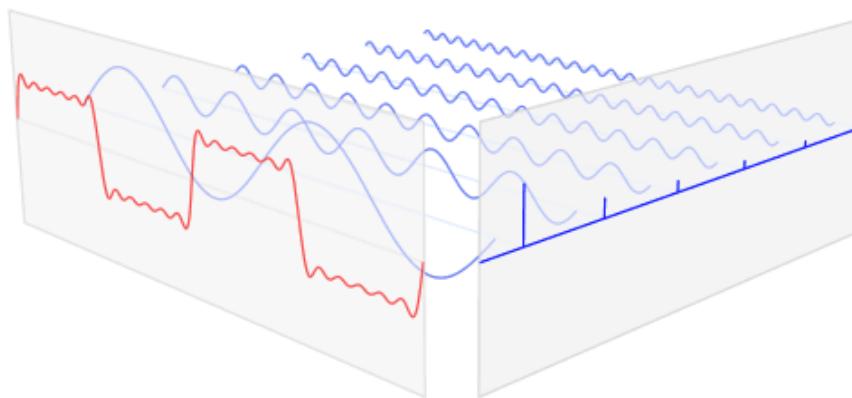


Figure 14. Decomposition of an original time-domain signal into the frequency-domain (Stack Exchange)

3.5.2 Harmonic ANalysis of Time Series Algorithm and Application

Fourier transforms were used in this study to fill PROBA-V data gaps caused by the limited sensor swath and to remove outlier values caused by clouding or other anomalies. For this application the Harmonic ANalysis of Time Series (HANTS) algorithm (Menenti et al. 1993, Verhoef et al. 1996) proved very capable as the selected Fourier algorithm. Compared to traditional Fourier transform algorithms, such as the widely popular Fast Fourier Transform, HANTS has the advantage of not requiring discrete time steps. This makes it much more useful in remote sensing applications where data can be non-uniform in both temporal and spatial coverage. Furthermore, the HANTS algorithm has more comprehensive outlier management which has already proven effective at cloud removal in NDVI images (Roerink et al. 2000). Error introduced by clouding is generally systematic so being able to selectively suppress how or high outliers is especially helpful. The ability of the HANTS algorithm in both gap filling and cloud removal made it invaluable for this research.

The curve fitting and gap-filling abilities of HANTS are largely dependent on the various parameters that serve as input to the algorithm. The most important is the number of frequencies (NF) to calculate harmonic components for which are summed in order to recreate the fitted curve. This parameter must be determined on a case to case basis but largely depends on the periodicity of the data. A pixel-by-pixel analysis determined that three harmonics were able to curve fit the index values which typically experience one or two major peaks per year. An attempt was made to minimize the number of harmonics since higher frequencies are likely to highlight differences within classes rather than between classes. This choice is specific to this application since these high frequency details could be desirable for analysis within the same land cover class.

The fit error tolerance specifies the minimum distance a point must be away from the current iteration of the reconstructed curve in order to be considered an outlier. The HANTS algorithm uses 2^*NF-1 data points to fit the curve and the degree of over-determination specifies how many extra data points should be used. The set of HANTS parameters used in the initial analysis are listed below in Table 5.

Table 5. Default HANTS parameters used for each spectral index

Index	Number of Harmonics	Outlier Rejection	Fit Error Tolerance	Degree of Over-determination
NDVI	3	Low	0.2	0
NDWI	3	High	0.2	0
CI	3	Low	0.2	0
MSAVI2	3	Low	0.2	0
VNIROI	3	High	0.2	0

Figure 15 below gives an example of the results from HANTS in comparison to the raw index values. Results were examined as the highest order harmonic was increased from one to three. The figure portrays an NDVI time series so low outliers are rejected since clouding causes high reflectance in the visual spectrum (and thus low NDVI values). This is most clearly seen in the plot of the 3rd order harmonic. For other indices outlier rejection was selected based on the expectance index value change in clouded conditions.

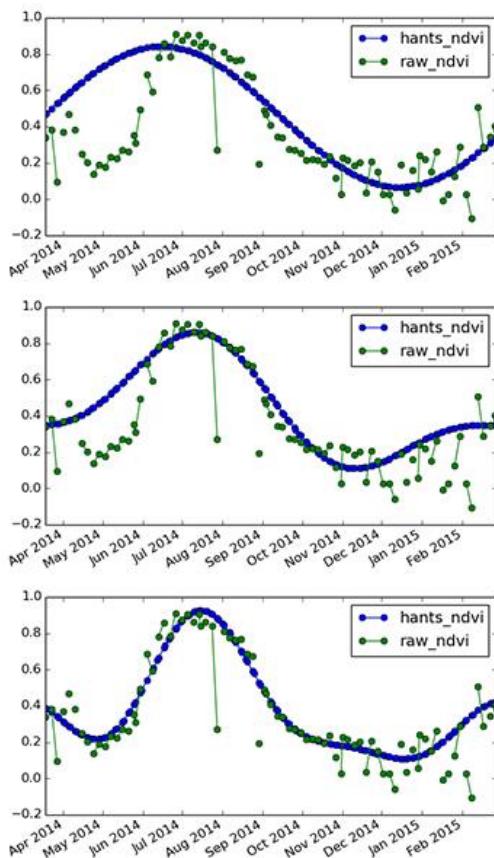


Figure 15. HANTS curve fitting using 1st, 2nd and 3rd order harmonics (left, center, and right respectively) of the NDVI for example corn pixel near Stockton, CA

3.6 Classification

The results of the HANTS algorithm serve as input to multiple classification efforts tested in this report. This is either in the form of gap-free index maps generated for each raw input map or in the form of harmonic component maps. In the case of the parameters outlined in Table 5, this amounts to 35 maps across the 5 indices: amplitude and phase maps for each harmonic of order one and higher and the amplitude of the zero harmonic which is the average index value. Two different approaches were used to define land and water use classes with similar amplitudes and phases: an unsupervised classification using clustering and maximum likelihood, and supervised classification use a method based on root-mean-square error to reference curves. Both are explained in further detail in this section.

3.6.1 Unsupervised Maximum Likelihood Classification

Principal Component Analysis

It was hypothesized that there might be relatively high redundancy in the HANTS harmonic component maps, so a principal component analysis was performed before maximum likelihood classification. The principal component analysis is a statistical procedure which uses a series of orthogonal transformations to shift a set of N-dimensional input variables to a lower dimension set of less correlated components. A visual example of this process is shown in Figure 16 below. This sort of analysis removes redundancies due to correlation between input variables by definition since each component gives the highest possible variance while remaining orthogonal to previous components. The principal components are eigenvectors of the covariance matrix and as such are orthogonal by nature (Jolliffe 2002).

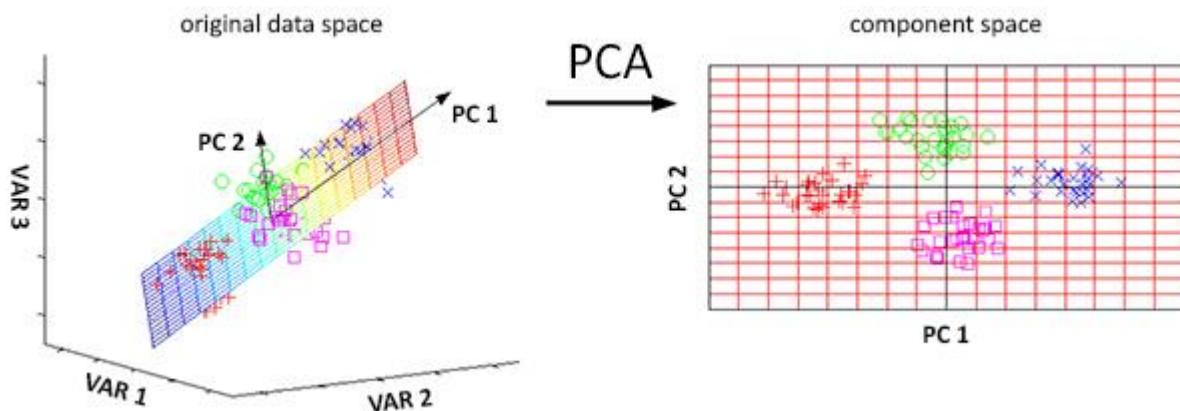


Figure 16. Transformation of data from original variables (VAR 1, 2 & 3) in 3-D space to principal components (PC 1 & 2) in component space (Scholz 2006)

Unsupervised classification was performed using the harmonic component maps for each index. This amounts the average value map provided by the amplitude of the 0th harmonic in addition to phase and amplitude maps for each subsequent harmonic. A principal component analysis is performed on the Fourier components and resulting principal components with less than one percent importance, as determined by the algorithm, were not used in subsequent calculations. Thus, the resulting number of principal component maps varied depending on the inputs of each classification attempt.

Clustering and Maximum Likelihood Analysis

Next, a standard clustering algorithm is applied which serves as the first pass in unsupervised classification. This algorithm attempts to create pixel clusters with the initial number of clusters set by the user. In this research, an attempt was made to maximize the number of classes by selecting a high number of classes by using the parameters outline in appendix section 8.1. In GRASS, the maximum number of potential cluster classes is limited to 255. Clusters are created using an iterative process and the resulting signatures are saved.

The second pass of unsupervised classification is performed with a maximum likelihood classifier. The algorithm classifies pixels by selecting the best statistical match out of classes defined by the clustering algorithm by assuming a probability density function.

Data-Assisted Labeling Approach

The clustering and maximum likelihood analysis results are only based on the input data from HANTS and thus classes in the output are numerical only and unlabeled. The unsupervised classes must first be labeled to allow for comparison to ground truth data. A data-assisted labeling approach (DALA) as described by Lang et al. (2008) is applied in order to make accuracy analysis possible.

In this technique a subset of each area delineated by unsupervised classification is used to guess which ground truth reference class best matches the area. Ideally, this reference set is different than the ground truth data used for the later accuracy assessment, however, for this research this was not possible since no other detailed ground truth data sets were freely available in the study areas and field observations were outside the scope of this research.

Instead, a subset of 10,000 and 100,000 pixels were chosen at random for the Netherlands and California study areas respectively for sampling. Next, the number of pixels in each reference class were counted for each of the unsupervised classes. A majority rule is applied so that each class is labeled as the ground truth class with the highest pixel count. Finally, the unsupervised classification map is reclassified according to the results of the DALA and error analysis is performed. Since the unsupervised classification is performed with a very high number of output classes, it's worth noting that labels from the ground truth dataset will be assigned to multiple classes in many cases.

Attempts to Improve Accuracy

There were also several attempts to improve the accuracy of classification. These variations in the classification process, which are further detailed in results section, also serve as a means to determine which factors could be affecting classification accuracy and to better understand the limitations of the methods used in this research. The following approaches were attempted:

1. Limiting classification to areas that meet certain thresholds for contiguous areas of the same class. This was in an attempt to reduce the mixed-pixel effect due to finer landscape structure and to focus on larger land cover areas.
2. Combining ground truth classes in order to reduce the overall number of classes and thus the classification complexity.
3. Using a fraction of the ground truth map for training signature generation which serves as input to the maximum likelihood classifier, thus introducing supervision.
4. Comparing results to a more traditional approach of using band reflectance values for 3 hand-selected, could-free images.

3.6.2 Spectrotemporal RMSE Classification

Root-Mean-Square Error Calculations and Classification Algorithm

Classification based on the spectrotemporal signatures of ground truth classes was also performed. In this technique, time series mean index values were computed for each class and served as reference curves, thus making this approach supervised. Classification is performed based on a root-mean-square error (RMSE) analysis of each pixel compared to a reference library of classes. In mathematical terms root-mean-square error for n intervals is calculated between two curves $\hat{\theta}$ and θ as follows for n time steps:

$$RMSE = \sqrt{\frac{1}{n} \sum_{1}^{n} (\hat{\theta}_n - \theta_n)^2}$$

The algorithm used in this research follows these steps:

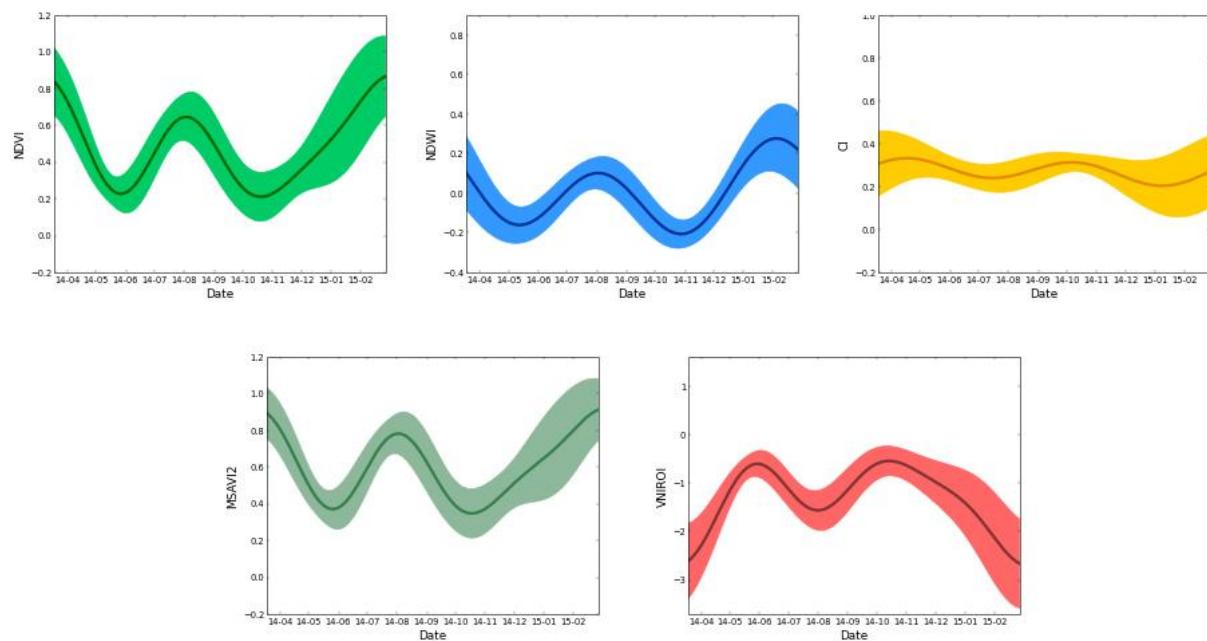
1. Masking calculation area to exclude open water and foreign territory.
2. For each spectral index considered:
 - a. Univariate statistical analysis of all ground truth classes at each time step, from which the mean and standard deviation values of each class are later used. Note that time steps of 10 days were used since this processes is very computationally expensive.
 - b. Creation of mean value maps for each class at each time step. Thus, all pixels are assigned the same mean values in these maps.
 - c. Mean-squared error is found by calculating the difference between each pixel from HANTS and the previously generated mean value maps. This results in one mean-squared error map per class per time step. This calculation was altered to apply weighting or to use first derivative values in some cases as explained in the following sub-section.
 - d. Mean-square error maps are first summed across all time steps for each class and are then divided by the number of time steps and rooted to produce one per-pixel RMSE error map per ground truth class.
3. RMSE values are averaged between considered indices per-pixel. This results in one RMSE value map for each ground truth class.
4. The minimum RMSE value across all classes is determined per-pixel and the associated ground truth class is assigned to the pixel. This is the final classification result.

Further Attempts to Increase Accuracy

There were also attempts to increase the accuracy of the RMSE classifier. The first two approaches were based on the hypothesis that some parts of the time series curves contain more useful information than other parts, in terms of classification. Thus, differential weighting of time steps would lead to improved accuracy if the weighting method was valid. This is particularly relevant for The Netherlands where most vegetation goes dormant during the winter, and all land surfaces are relatively wet due to rainfall and snowfall and absence of evaporation. Such situation is not favorable for distinguishing land use classes having different water management conditions.

The first approach normalized the mean error by the standard deviation at each time step. This was in an attempt to focus on time windows where in-class variability was limited. The next approach assigned weights based on the distance of a time step from the vegetation peak based on a normal distribution.

The final attempt to improve accuracy using the RMSE method involved calculating first derivative curves of each reference curve for each index. The first derivative curves were then used in place of the original curves for the RMSE method. This should give some indication of the usefulness of the time rate-of-change for class differentiation.



*Figure 17. Example mean value plots for barley/corn double cropping plots in California.
The shaded bands show ± 1 standard deviation.*

4 Results

4.1 HANTS Performance

Finally, the performance of the HANTS algorithm is examined. As with any gap-filling technique, the HANTS algorithm serves as a source of error. The section aims to determine how much error is introduced and how successfully the algorithm dealt with cloud cover by means of outlier removal. HANTS parameters outlined in Table 5 were used in this section.

4.1.1 Cloud Removal

The cloud removal and outlier removal abilities of HANTS were critically important for our approach. Instead of selecting cloud-free images, either by use of a cloud detection algorithm or by hand, this research incorporates all data collected in the study areas over the examination period. Thus, it is important that the gap-filling algorithm not only bridges gaps caused by orbital characteristics but also proves effective at outlier removal.

Figure 18 below shows the results of the HANTS algorithm in the Netherlands on 20 September 2014 with the raw image above and the HANTS image below. The image is an NDVI map with vegetation highlighted in green. The high reflectance of clouds in the visible spectrum gives them low NDVI values where are shown in blue in the image. A visual inspection of the cloud removal provided by HANTS is in agreement with the effectiveness found by Roerink et al. (2000).

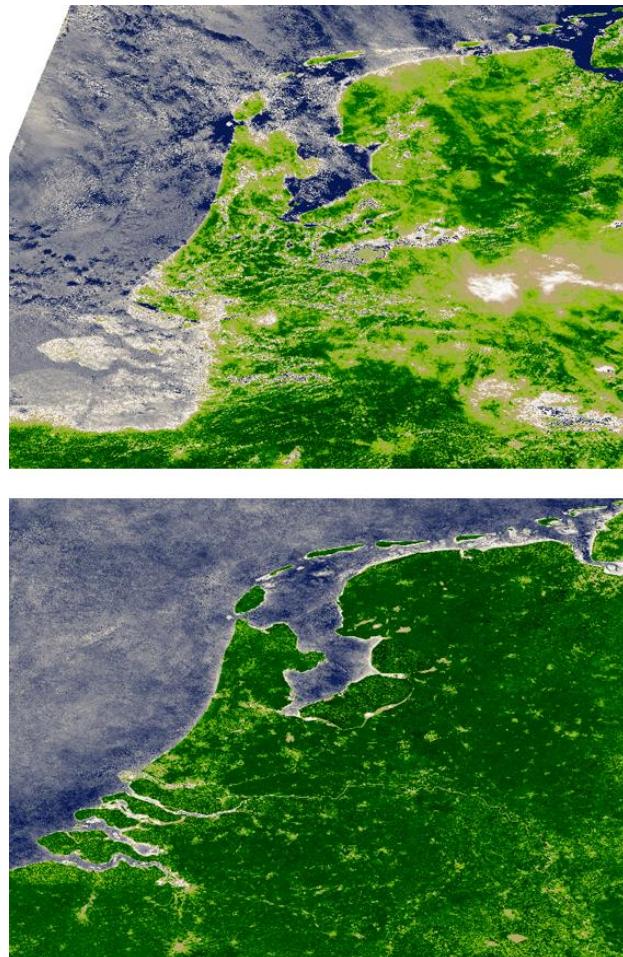


Figure 18. Demonstration of cloud removal capabilities of HANTS in the Netherlands on 20 September 2014 NDVI map.

4.1.2 Error Compared to Raw Index Values

A more objective approach was applied to compare values from HANTS to hand-picked images with low clouding. The values were compared to two cloud free days: one in the summer (23 July 2014) and one in the winter (7 January 2015). This analysis was performed in the Netherlands since it is the study area more affected by clouds. For each index, average values were computed for the extent of the Netherlands study area for both the raw index values and HANTS. The difference between these two values gives some indication of how much error is introduced by curve-fitting with HANTS over the entire study area. It's worth noting that this error is not uniformly distributed as some areas will inevitably experience more cloud cover than others.

*Table 6. HANTS values compared to cloud free days in the Netherlands.
Values are averaged for each index for non-water areas in the study area.*

Index	23 July 2014				7 January 2015			
	Measured	HANTS	Abs. Difference	Rel. Difference	Measured	HANTS	Abs. Difference	Rel. Difference
NDVI	0.724	0.685	0.039	0.054	0.519	0.455	0.064	0.123
NDWI	0.263	0.247	0.016	0.061	0.219	0.188	0.031	0.141
CI	0.281	0.231	0.050	0.178	0.774	0.826	-0.052	0.067
MSAVI2	0.826	0.788	0.038	0.046	0.663	0.622	0.041	0.062
VNIROI	-1.542	-1.677	0.135	0.088	-0.954	-1.087	0.133	0.139

While the results in Table 6 show a relatively low error introduced by HANTS on the summer day, results were slightly worse in winter. A per-pixel analysis of error in January revealed that some areas had only one cloud free data-point in January, particularly in the southern Netherlands province of Limburg. Unfortunately, the outlier-removal capabilities of HANTS has limitations which were fully tested under these circumstances.

Figure 19 below shows an example of HANTS curve fitting of a time series NDVI curve of a grassland pixel located in Limburg. Note that the low NDVI values in January and February 2015 are the result of clouds with only one cloud free observation (on 7 January 2015). The HANTS algorithm, understandably, has trouble fitting this time period in the example pixel and surrounding areas.

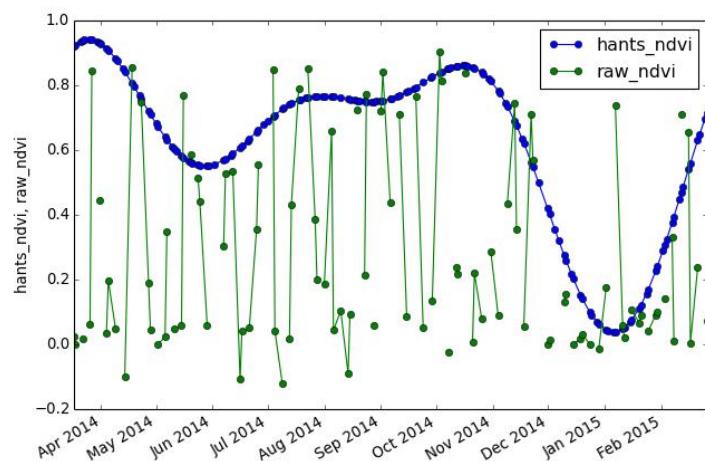


Figure 19. Example HANTS curve fitting of NDVI for a grassland pixel in Limburg, NL using 3rd order harmonic components

4.2 Unsupervised Classification

Unsupervised classification was performed using the statistical techniques outlined in the methods section. Table 7 below shows the results of the classification for the Netherlands which uses 53 different classes from the combined BRP/BBG ground truth map. Table 8 shows classification results in the California study area which contained 87 classes from the CDL. Results are shown for each spectral index with and without the principal component analysis applied and using between 1 and 3 frequencies in HANTS. The number of frequencies corresponds to the highest order harmonic components used. The accuracy results for all results in the unsupervised classification section were obtained using the data-assisted labeling approach described in the methods section.

Table 7. Unsupervised classification results for the Netherlands (using 53 classes). A comparison is made between using different orders of harmonics and PCA.

Index	% Accuracy					
	With PCA			Without PCA		
	1 st order	2 nd order	3 rd order	1 st order	2 nd order	3 rd order
NDVI	41.1	37.1	37.0	41.1	37.1	37.2
NDWI	41.9	35.6	36.3	40.6	35.7	36.6
CI	41.5	35.5	35.5	35.6	35.4	35.6
MSAVI2	40.7	36.3	36.8	40.6	36.2	36.8
VNIROI	42.1	37.4	38.4	42.1	37.4	38.9

Table 8. Unsupervised classification results for California (using 87 classes) A comparison is made between using different orders of harmonics and PCA.

Index	% Accuracy					
	With PCA			Without PCA		
	1 st order	2 nd order	3 rd order	1 st order	2 nd order	3 rd order
NDVI	50.5	44.6	45.8	40.0	45.2	47.5
NDWI	44.8	42.2	42.5	35.1	42.4	43.1
CI	42.6	40.4	41.0	33.7	40.9	42.7
MSAVI2	50.7	44.3	45.7	40.0	44.8	47.1
VNIROI	49.9	46.4	46.4	42.0	46.62	47.2

These results gave some indication of the performance of each index, the usefulness of PCA and the selection of harmonic components. From this point forward calculations were made using PCA, since it seems to have either a neutral or positive effect on the results. Both the number of harmonics and PCA will be discussed further in discussion section 5.1.

The results from California appear to be systematically higher than those in the Netherlands. It's possible that this is the result of larger agricultural plot sizes and larger contiguous areas in general. The effects of plot sizes and potential implications of the mixed pixel effect will be explored later in this section.

Another explanation is that there is less variation in vegetation and moisture trends in the Netherlands where many land use classes are green, moist and have the same seasonality. This seasonality is demonstrated in the frequency distribution plots in Figure 20 below. The plots show the day of the year on which the NDVI peak is reached, per pixel. These plots were created using maps of phase offset for the first order harmonic component for NDVI. In the Netherlands, many pixels reach peak vegetation around the 200-220th day of the year, which lies in late July. In the California study area, natural vegetated areas reach a peak in the early spring before the hot, dry summer. Irrigated land, on the other hand, experiences peak vegetation over a larger range of time, mostly depending on crop types and planting patterns.

These large differences in the frequency distributions between the two study areas could explain the systematically higher accuracies in California. A less uniform distribution will aid in class discrimination since it will be easier for the clustering algorithm to find unique clusters.

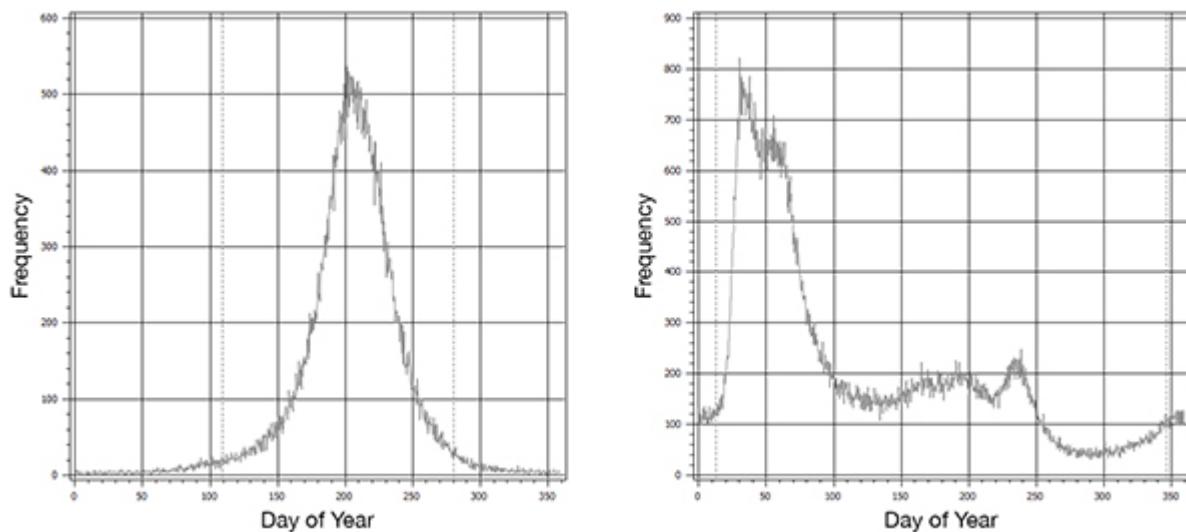


Figure 20. Frequency distributions for day of year experiencing peak vegetation. This is determined by the phase of the first order harmonic component for NDVI in both the Netherlands (left) and California (right)

4.2.1 Using Combinations of Indices

Next, unsupervised classification was performed using a combination of spectral indices. Both Table 9 and Table 10, for the Netherlands and California respectively, show that increases in accuracy between top results is small, however, limiting calculations to a smaller subset of indices has a very positive effect on computation times. Full tables with the results of all possible index combinations are listed in appendix section 8.3.

Table 9. Results of combining indices in the Netherlands (top 5 combinations)

NDVI	NDWI	CI	MSAVI2	VNIROI	% Accuracy
▪	▪	▪	▪	▪	43.8
▪	▪		▪	▪	43.7
▪	▪			▪	43.7
▪	▪	▪		▪	43.6
	▪		▪	▪	43.6

Table 10. Results of combining indices in the California (top 5 combinations)

NDVI	NDWI	CI	MSAVI2	VNIROI	% Accuracy
▪	▪			▪	51.4
▪	▪	▪	▪	▪	51.2
	▪	▪	▪	▪	50.9
▪	▪	▪	▪		49.7
▪		▪		▪	49.7

The combination of NDVI, NDWI and VNIROI was used for all of the following classification results including both unsupervised classification and the RMSE classifier. This combination was selected due to both the relatively high performance and the desire to retain some information from all spectral bands. VNIROI appears to offer a reasonable improvement on the overall classification results compared to CI and MSAVI2. One interim conclusion is that it is worth exploring and developing new spectral indicators that meet the need of water accounting, and not use only standard indices developed for vegetation. Another conclusion is that NDWI with information on leaf water content has positive contribution on the overall results.

With one zero order and two first order harmonic component maps per index, this amounts to an input of 9 maps before the principal component analysis. It's worth noting that gains over using a single index appear to be minimal, so the importance of this selection may be limited.

4.2.2 Attempts to Improve Accuracy

The following section lists the results of multiple attempts to improve the accuracy of unsupervised classification. The first lines of the tables in this section are the original unsupervised results.

Reducing the Mixed Pixel Effect

The mixed-pixel effect is a common issue in remote sensing. The discrete pixel resolution required by sensors dictates that each data point is essentially an integration over the pixel area. In terms of land use classification, this means that PROBA-V pixels may encompass multiple land use classes. Such pixels are more difficult to classify since they may contain spectral characteristics from a number of different sources.

In an attempt to minimize the error introduced by the mixed pixel effect, the classification was limited to contiguous areas of varying sizes containing only one ground truth class. This amounted to choosing areas with no variance when re-sampled to a lower resolution and masking calculations to exclude other areas. A number of such lower resolution thresholds were examined, as detailed in Table 11 below. Although the contiguous pixel areas were chosen at a lower resolution, the classification was still performed at 100m for pixels which fell inside the masked area.

Table 11. Results of mixed-pixel effect reduction using various thresholds for minimum contiguous class areas. The percentage of land area meeting these thresholds is also supplied.

Pixel Threshold	% Accuracy		% of Area Meeting Threshold	
	<i>Netherlands</i>	<i>California</i>	<i>Netherlands</i>	<i>California</i>
100 m (original)	43.7	51.4	100	100
200 m	52.4	60.1	49.4	45.7
300 m	57.1	67.5	27.4	27.6
400 m	58.6	72.6	16.4	18.3
500 m	58.5	76.7	10.2	12.8

This results show that the accuracy of classification improves as the threshold for contiguous pixels is increased in both study areas. This suggests that the mixed pixel effect does contribute some of the classification error. The improvement in accuracy is noticeably higher in California which is likely due to the more “noisy” nature and higher error of the ground truth map which was generated using remote sensing techniques. In comparison, the BRP used in this research is a rasterized version of clearly delineated agricultural plots.

Figure 21 gives an example of how an agricultural area in the San Joaquin Valley is masked. The 100m simply removes water and foreign territory as previously described. In the 300m and 500m masks, mixed-pixel boundaries are removed and only large contiguous plots remain.



Figure 21. Example of masking contiguous areas. Maps from left to right are the CDL ground truth, 100m mask, 300m mask and 500m mask. Areas in white are included and those in black are excluded.

Reduction of Class Variety

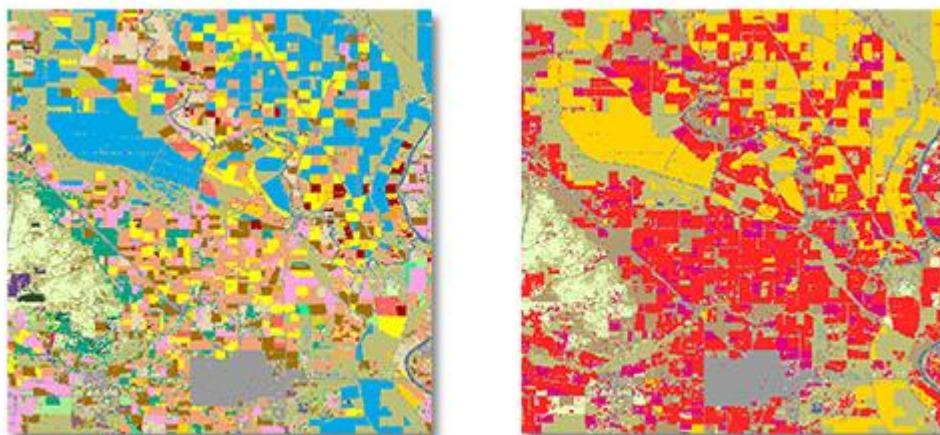
Although the aim of this research was to differentiate between large numbers of land and water use classes, the results indicate that the accuracy of such attempts is very low. Thus, classification was also performed on a lower number of classes in order to determine the effect on accuracy. In order to reduce the number of classes, existing classes were combined manually based both on the land cover type and similarity. The details of these combinations are provided for both study areas in appendix section 8.2 and the results of class reduction are shown in Table 12 below.

Table 12. Results of class reduction and number of classes used in each classification attempt

Class Complexity	% Accuracy		Number of Classes	
	<i>Netherlands</i>	<i>California</i>	<i>Netherlands</i>	<i>California</i>
Default	43.8	51.4	53	87
Reduced Classes	44.5	55.7	13	14

The results show a slight increase in accuracy for both study areas. The small magnitude of accuracy improvement is likely due to the size of the areas being combined. Classes which cover a large portion of the landscape, such as forests in California and grasslands in the Netherlands, saw little change in accuracy during class reduction while still accounting for a relatively large portion of the classification accuracy. Still, these results suggest that class confusion between the initially large numbers of classes is affected more by mixed pixels than by the class variety itself.

For both study areas, most class combinations were for agricultural classes which made up the majority of total classes. An example of the reduction of classes can be seen in Figure 22 which shows a mix of developed and agricultural land near Stockton, CA. Crops were grouped by growing seasons in some cases including plots with double cropping practices. Classes of tree crops, forests, natural areas and developed areas were also grouped accordingly.



*Figure 22. Example of class reduction in an area near Stockton, CA.
Note that most class combinations are agricultural areas.*

Using Some Ground Truth for Training

Classification was attempted again using various amounts of data for training, thus shifting to a supervised classification which still uses a maximum likelihood classifier. Ideally, training data should come from a ground truth source different than that used for verification but such data was not available in this case. Instead, a percentage of random pixel from the ground truth maps were used as inputs for signature generation prior to maximum likelihood analysis. The results of this attempt are listed in Table 13 below and show a negligible change to accuracy in the Netherlands and a mild decrease in California. The decrease of accuracy in California could possibly be explained by error propagation from the CDL dataset. Furthermore, the ground truth datasets (and thus training data) may be forcing the results into classes which are not well aligned with the input data with a focus on land cover and lacking a connection to hydrological processes. A better approach for ground truth data is given in recommendations section 6.6.

Table 13. Results from supervised classification with training

Training	% Accuracy	
	<i>Netherlands</i>	<i>California</i>
0% (unsupervised)	43.7	51.4
5%	43.3	46.9
10%	43.4	48.1
15%	43.3	47.5

4.2.3 Comparison to Traditional Unsupervised Classification

The results of the unsupervised classification using the zero and first order harmonic component maps for NDVI, NDWI and VNIROI were compared to a more traditional unsupervised classification approach using band reflectance data from 3 hand-picked, cloud-free days in each locations.

In the Netherlands the cloud-free images were from 23 July 2014, 13 November 2014 and 7 January 2015. In California the images were from 23 March 2014, 1 July 2014 and 18 October 2014. HANTS maps generated using 3rd order harmonic components as outlined in Table 5 were used and principle component analysis was not applied. The results of the comparison are shown in Table 14 .

Table 14. Comparison of results with a traditional unsupervised classification approach

Input Data	% Accuracy	
	<i>Netherlands</i>	<i>California</i>
Using HC sets	43.8	51.4
Using 3 cloud-free images	58.2	61.6

In both the Netherlands and California the more traditional unsupervised classification approach outperforms the harmonic component based classification by more than 10%. This suggests that unsupervised time series classification by means harmonic component spectral index maps is not an effective approach. This finding will be detailed further in the discussion and conclusions section, however, it is already clear that the unsupervised classification approach put forth in this report is severely underperforming compared to traditional approaches. The implication of this finding is that additional information from time series analysis will not necessarily improve the overall mapping accuracy, as was the case with this approach.

4.3 Spectrotemporal RMSE Classification

Classification based on the RMSE method outlined above was performed on the two study locations. As previously explained, classification is performed per-pixel based on the class with the lowest average RMSE between NDVI, NDWI and VNIROI curves. Table 15 shows the results of this classification using one to three frequencies in HANTS. Note that the table also includes the accuracy of the technique if the top N results for each pixel are considered. Thus, if the correct class is among the top N results for a given pixel (the N lowest RMSE values), then it will be counted as correct. These results also provide some insight into how much the accuracy of the technique is affected by different classes sharing similar time series behavior. This similarity between classes also varies with time, so some time periods show more uniqueness between classes compared to others.

Table 15. Results of spectrotemporal RMSE approach using NDVI, NDWI and VNIROI

Using top N results	% Accuracy					
	Netherlands			California		
	1 st order	2 nd order	3 rd order	1 st order	2 nd order	3 rd order
1	41.5	40.9	37.8	32.6	33.1	24.6
2	53.6	51.3	48.0	41.6	42.0	33.0
3	58.2	56.5	54.2	49.2	49.5	41.2
4	61.2	60.0	58.5	54.0	54.3	46.0
5	63.4	62.7	62.0	58.9	59.0	49.8

The results of the RMSE classifier show decreased accuracy compared to unsupervised classification. Since the 1st order harmonics performed best in the Netherlands and negligibly worse than 2nd order harmonics in California, HANTS maps using 1st order harmonics were used in the remainder of this results section. The better performance of the 2nd order harmonics in California are likely due to double cropping and lower seasonality compared to the Netherlands. It is important to note that this varies from the HANTS maps using 3rd order harmonic components as listed in Table 5.

When the top 5 results are considered, more reasonable accuracies are obtained. This suggests that class confusion may only be occurring between small numbers of classes. Figure 23 gives a graphical representation of the difference between the accuracy of the correct class being listed among the top 5 results compared to the top class alone. The province of South Holland in The Netherlands is shown in this example.

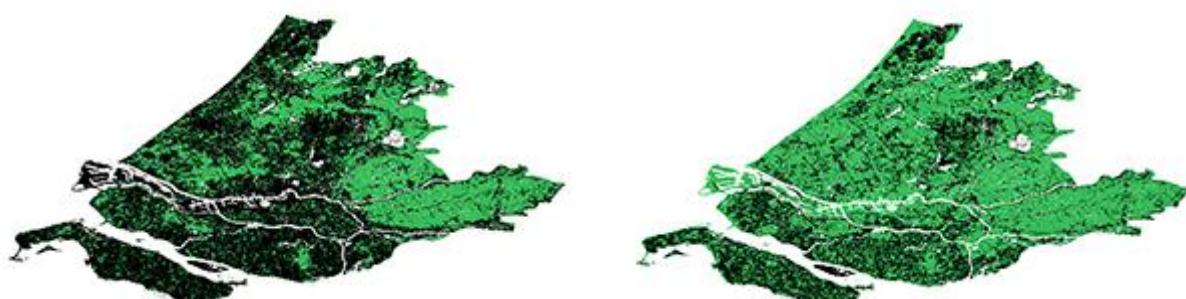


Figure 23. Difference in classification accuracy for top result (left) compared to top 5 results (right) in South Holland. Green indicates ground truth is matched and black indicated non-match.

Another example is shown in Figure 24 which demonstrates the classification process of the RMSE classifier. The figure shows a plot of the ground truth according to the CDL followed by the top 5 results for the NDVI time series of the example pixel. In this case, the second guess identifies the example pixel as winter wheat. The incorrect top result classified the pixel as triticale which is a hybrid between wheat and rye. Class confusion between wheat, rye and triticale in this example is due to low variation in the NDVI curves between the three classes as shown in the plots below.

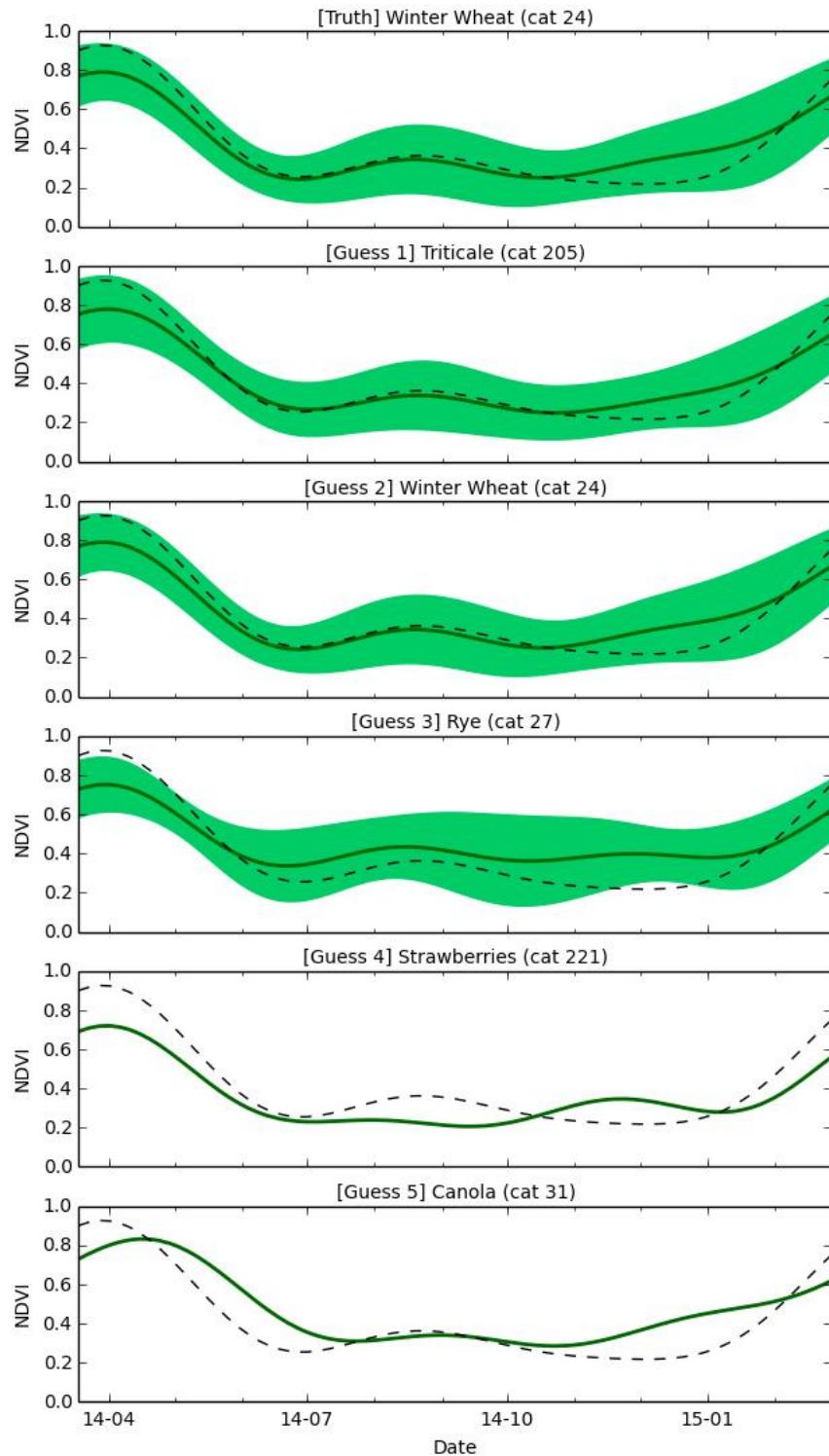


Figure 24. Example NDVI RMSE classification results of a winter wheat pixel near Stockton, CA.

4.3.1 Attempts to Improve Accuracy

The following section outlines a number of attempts that were made to increase the accuracy of the results using the RMSE method. These results are all in comparison to the original results using 1st order harmonic components for HANTS.

Normalization to Standard Deviation

The first attempt to increase the accuracy of the RMSE approach was to weigh the RMSE based on the standard deviation at each time step. This effectively punishes high error when the standard deviation is low and thus focuses on time spans with low in-class variation. Curves exhibiting standard deviation variation with time, such as broccoli in Figure 25 below, were the original motivation for this approach. The results of this approach are summarized in Table 16 using the modified RMSE equation is as follows with standard deviation at each time step designated as σ_n :

$$RMSE_W = \sqrt{\frac{1}{n} \sum_1^n \left(\frac{\hat{\theta}_n - \theta_n}{\sigma_n} \right)^2}$$

Table 16. Results using standard deviation based normalization

Normalization	% Accuracy	
	<i>Netherlands</i>	<i>California</i>
None	41.5	32.6
Standard deviation	45.6	21.6

The results are mixed and are different in each study area. This suggests that the effectiveness of normalization will largely be dependent on the input data. It is worth noting that the California study area had a higher number of classes, 87 in comparison to the 53 in the Netherlands. In addition, there are a larger number of classes with smaller relative area in California as evidenced by the frequency distribution plots in Figure 12 and Figure 13. Classes with fewer pixels often had lower standard deviations and were thus biased. An example of such a class is linseed in Figure 25 below.

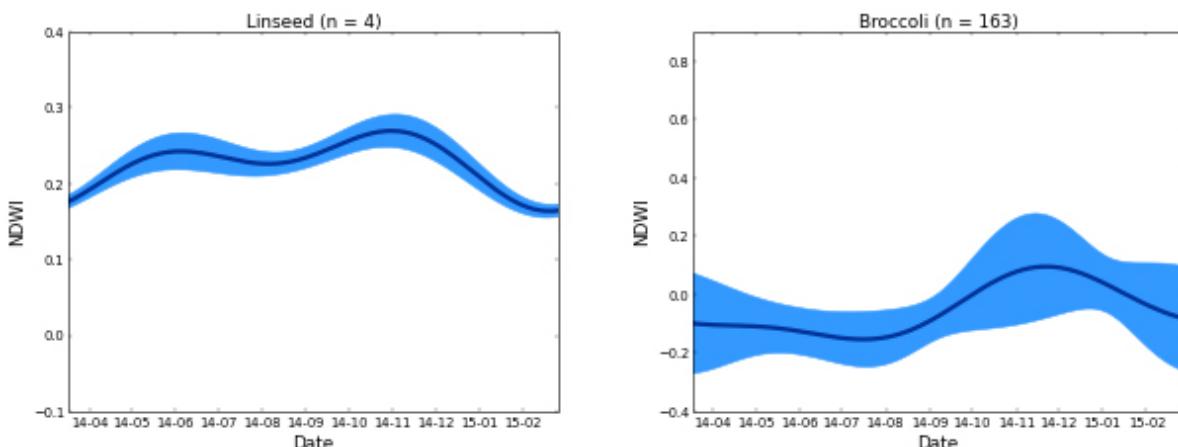


Figure 25. Examples of variation in standard deviation between time series NDWI curves for linseed (left, NL) and broccoli (right, CA). The shaded bands show ±1 standard deviation.

Weighting for Peak In-Class Vegetation

This peak was determined by the date of the maximum NDVI value. In this case a coefficient was applied at each time step based on the normal distribution. Two different time periods were used to describe the distribution and further described in the results section.

Next, the classifier was modified to weigh for peak in-class vegetation. This was in the form of a normal distribution centered on the date of the maximum NDVI value with various standard deviation values tested. The probability density function of a normal distribution is as follows:

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

In this case, the peak index value is used for μ , x represents the day of the year, and values for the standard deviation (σ) are listed in Table 17 along with a summary of the results. For both study areas, classification accuracy declined significantly. The outcome of these results weren't expected to be this drastic however there is a possible explanation. Figure 20 which showed the frequency distributions of the time of the vegetation peak for both study areas sheds some light on this outcome. If vegetation peaks are concentrated at specific times of the year, then the time period surrounding the peak likely experiences high similarity between classes which leads to more class confusion and lower accuracy.

Table 17. Results with weighting for in-class vegetation peaks

Standard deviation (days)	% Accuracy	
	<i>Netherlands</i>	<i>California</i>
Without weighting	41.5	32.6
15	10.1	18.8
30	13.0	25.3

Using First Derivative

Another attempt at classification was made by using first derivative curves. Table 18 below compares the results of the first derivative curves to the originals. Using the first derivative of index values emphasizes the time-rate-of-change of index values over the amplitudes. Table 18 shows that this approach showed decreased accuracy in both study areas, significantly so in the Netherlands.

It is likely that only very specific parts of the first derivative curves actually provide useful information. Time periods with the highest magnitude first derivatives, both positive and negative, emphasize times when the index value is undergoing the most change. For NDVI values of agricultural areas, this would represent peak growth followed by harvest.

Table 18. Results using first derivative curves compared to original curves

Curves used	% Accuracy	
	<i>Netherlands</i>	<i>California</i>
Original	41.5	32.6
First derivative	9.7	26.7

5 Discussion & Conclusions

5.1 Harmonic Analysis

5.1.1 Number of Frequencies

The best results for both unsupervised classification and the RMSE classifier found that using one frequency produced the best results in the Netherlands as shown in Table 7 and Table 15. This includes the phase and amplitude of the first harmonic component as well as the mean annual value (the amplitude of the zeroth component). It appears that subsequent frequencies, while helpful in terms of more accurate curve fitting, only added to class confusion. The increase in accuracy gained by using was generally 4-9% for unsupervised classification and 2-4% using the RMSE method.

However, in California, there were conflicting results. Table 15 shows similar performance when using first and second order harmonics. It's possible that this is caused by the mild climate in the Upper San Joaquin valley allowing more flexibility in crop selection based on growing seasons. A reasonable number of areas were identified by the CDL as being used for double cropping, a practice that will introduce a minimum of two vegetation peaks. This could explain this difference in performance between first and second order harmonics in comparison to the Netherlands.

5.1.2 Principal Component Analysis

The principal component analysis was performed in order to reduce the number of inputs to the clustering algorithm by eliminating redundant and less useful data. Unsupervised classification was performed both with and without the principal component analysis in order to assess the usefulness of this step. The effects of running a principal component analysis showed mixed results between the Netherlands and California as emphasized in Table 7 and Table 8.

In the Netherlands, the PCA seemed to have little effect on the performance of the unsupervised classification. Differences in accuracy were both positive and negative with a magnitude of less than 1% in most cases. Results in California, however, show PCA leading to an increase of more than 10% accuracy in some cases. This suggests that there were more redundancies between harmonic component maps in California which may have been accentuated by the higher number of ground truth classes considered.

5.1.3 Fourier Transforms for Gap-Filling

A large part of this research depends on the ability of HANTS to perform gap-filling and outlier removal. As with any gap-filling procedure, it is important to ensure that the original data is appropriately represented.

Thus, the drawbacks of using Fourier transforms to curve fit missing data must be considered. With a low number of harmonics, Fourier transforms will fail to match sudden peaks or valleys in data. In terms of land use classification, this is important because crops are often harvested suddenly and that time period could potentially be critical in crop differentiation. On the other hand, using too many harmonics may lead to over-fitting and thus introduction of random error. Even without over-fitting, it's possible that too many harmonics will introduce higher frequencies which are essentially noise in terms of classification. This data could potentially be useful for discrimination of characteristics within a class but might cause problems with differentiation between classes.

The results of the classification methods used in this report suggest that using first or second order harmonics is most appropriate, depending on how strong the seasonality is in the study area. However, more harmonics must be used for reliable curve fitting and outlier removal.

5.2 Classification

5.2.1 Unsupervised Classification

The results of this research, as summarized in Table 19, show that unsupervised classification by means of clustering and maximum likelihood analysis of spectral index harmonic component maps is not an effective approach. It was outperformed by traditional unsupervised classification using band reflectance maps for 3 cloud-free images. Classification was attempted both with and without analysis, using varying numbers of harmonic components and with further attempts to improve accuracy with little success. Although, none of these approaches succeeded in achieving high accuracy, the analysis on mixed-pixels suggest that it is a significant source of error.

Table 19. Summary of unsupervised classification results and attempts to improve accuracy

Unsupervised Classification	% Accuracy	
	<i>Netherlands</i>	<i>California</i>
Using a percentage of ground truth as samples for training	0% (unsupervised)	43.7
	5%	43.3
	10%	43.4
	15%	43.3
Using default classes versus a reduced set	Default	43.8
	Reduced Classes	44.5
Using a contiguous pixel threshold	100 m	43.7
	200 m	52.4
	300 m	57.1
	400 m	58.6
	500 m	58.5
Using different input data	HC sets	43.8
	3 cloud-free images	58.2
		61.6

Figure 26 shows the difficulty that the unsupervised classification algorithms had with agricultural plot discrimination in the Netherlands. Compared to the BRP/BBG dataset (left) traditional unsupervised classification (center) could only distinguish four unique crops: potatoes, sugar beets, corn and winter wheat. The harmonic component based unsupervised classification (right) could only distinguish corn and winter wheat. In both cases, much of the agricultural land was incorrectly classified as grasslands. The misclassification in this example is likely due to low variation in time series curves between agricultural class due to rigid seasonality and to the mixed pixel effect which can play a large role in areas with a high variety of small plots.

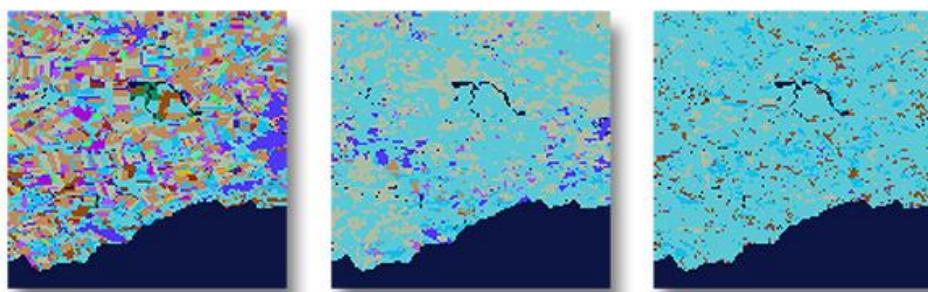


Figure 26. Poor performance in discrimination different crops in Zeeland, NL. The combined BRP/BBG dataset (left), traditional unsupervised classification (center) and harmonic component based classification (right) are shown.

Using a full annual time series introduces a lot of information. Traditional unsupervised classification algorithms such as maximum likelihood classifiers are designed for analysis of multi-spectral information at only a few points in time. It should be taken into consideration that the low performance of these algorithms could be explained by the fact that the supplied input data is too different than that which the algorithms were designed to use, or that the data doesn't conform to probability density functions as required by the algorithm. This is supported by the findings in this report where traditional unsupervised classification using band reflectance from 3 images outperformed the harmonic component based unsupervised classifier. Hence, harmonic component maps should be avoided as input to clustering algorithms. There is also the possibility that the classification was limited by the ground truth maps which were more focused on land cover and lacked the sought after connection to hydrological processes which is further detailed in the recommendations section 6.6.

5.2.2 Supervised RMSE Classification

The RMSE classifier also failed to produce results with a reliable level of accuracy. However, when the top five best guesses were taken into consideration the results suggest that this approach might have some promise as a time series classifier. It's possible that having more spectral bands and thus access to a larger number of spectral indices will increase the viability of the RMSE technique, however, such investigation was outside the scope of this research.

The results of the RMSE classifier and attempts to improve classification accuracy are shown in Table 20 below. The attempts to improve accuracy by means of weighting and using first derivative curves resulting in changes of accuracy ranging from a slight improvement to significant worsening as previously discussed in the results section. Still, there are many more weighting schemes that were not applied in this report which could potentially alter the accuracy of this approach if different parts of the time series curves are more useful in classification. In the results of this report, the only improvement from weighting occurred during normalization of mean error by standard deviation at each time step, but only in the Netherlands.

Table 20. Summary of RMSE classification results and attempts to improve accuracy

RMSE Classification	% Accuracy	
	Netherlands	California
Using top N results	1	41.5
	2	53.6
	3	58.2
	4	61.2
	5	63.4
Normalization using standard deviation	None	41.5
	Normalized	45.6
Weighting to peak-vegetation using normal distribution	None	41.5
	$\sigma = 15$	10.1
	$\sigma = 30$	13.0
Using first derivative curves	Original	41.5
	First derivative	9.7

One of the major drawbacks of this method is computational cost. Both the generation of the original signatures and RMSE calculation involve traversing through an entire year of data, a procedure which is slow given current database structures. Another issue with this approach is the reliance on accurate ground truth data for signature generation. This limits the effectiveness of such techniques to areas with adequate ground truth data, something this research was trying to avoid.

There were some examples of interesting findings using the RMSE classifier which suggest that it might have more merit than the accuracy results suggest. Figure 27, shows a golf course in Stockton, CA neighboring a residential area. The classes assigned by the CDL (left) and found by the RMSE are quite different for the area displayed in the satellite image from Google Earth (center).

Table 21 on the right provides a limited color legend for the example area. The CDL assigns the grass areas of the golf course mostly as “developed” although some dispersed “alfalfa” pixels were also detected. The RMSE classifier, however, clearly delineated the golf course with a mix of alfalfa and other hay pixels, both of which have very similar time series index curves compared to grass. Unlike the CDL, the RMSE results detect the full extent of the grass area.

Table 21. Limited CDL color legend

CDL Class	Color
Developed	Grey
Alfalfa	Pink
Other Hay/Non-Alfalfa	Light Green
Woody Wetlands	Light Blue
Tree Crops & Orchards	Dark Green
Fallow/Idle Cropland	Yellow-Gold

More interestingly, the rows of trees surrounding golf course were classified as tree crops. It’s possible that these rows show similar time series spectral characteristics compared to rows of trees in orchards. Furthermore, the suburban area to the top left of the figure is detected as woody wetlands. A closer inspection of high resolution imagery of this area reveals that it has more back yard pools and higher tree cover compared to the suburban area in the top left of the image. Additionally, an open water body lies between the area and the golf course. This example suggests that it is possible that the RMSE classification results are actually richer than the ground truth data. This supports the call in recommendations section 6.6 to use ground truth data more closely correlated to the categories of vegetative cover, moisture and development intensity proposed in Figure 3 of the introduction.



Figure 27. Maps from the CDL (left), Google Earth (center) and the RMSE classifier are compared for a golf course and a suburban area in Stockton, CA. The results of the RMSE classification don't match the CDL closely but may show richer details linked to vegetation and water features.



Figure 28. Higher resolution Google Earth images of area between the golf course and suburban area.

5.3 Uniqueness of Time-Series Curves both Within and Between Classes

The results of the study indicate that classification through time series analysis of spectral index maps is limited by similarity between classes, especially agricultural areas. In many cases, different crops have very similar growing seasons and thus have similar time series behavior as observed by the vegetation-focused bands of the PROBA-V. On the other hand, there is also a limitation on accurate classification imposed by differences in spectrotemporal behavior within classes.

Ideally, there should be minimal in-class variability with large variability between classes. The performance of classification by means of time series analysis of spectral profiles relies both of these conditions being met.

5.3.1 Differences within Classes

Variation of spectrotemporal behavior within classes is problematic when matching pixels of the same class using either clustering or RMSE minimization. There are various possible reasons for such variation. Differences in phase could be explained by farmers planting crops at different times. Other variation could be caused by differing crop health, field configurations or growing practices. Table 22 and Table 23 below provide some analysis of in-class variability for the Netherlands and California study areas respectively for a selection of classes with full tables given in appendix section 8.4.

Table 22. Extent of in-class variation in both space (monthly standard deviation) and time (vegetation peak) for the Netherlands. Only the top 10 classes with lowest mean monthly standard deviation are listed.

Class ID	Class Description	# of pixels	Monthly Standard Deviation		Veg. Peak (day of year)	
			Mean	St. Dev	Mean	St. Dev
666	Linseed	5	0.059	0.039	198.6	4.3
2297	Walnut	26	0.067	0.022	222.4	6.5
212	Fruit trees	24568	0.128	0.034	213.9	23.6
3736	Flax	2597	0.133	0.044	173.2	40.4
2026	Fallow (forested)	72	0.133	0.035	196.6	17.8
663	Lupin	139	0.139	0.024	205.6	29.2
664	Rapeseed	46	0.142	0.022	203.2	32.9
511	Chicory	4619	0.144	0.027	229.7	17.4
247	Poppy	679	0.144	0.038	208.5	35.0
256	Sugar beets	100435	0.146	0.029	222.8	20.4

Table 23. Extent of in-class variation in both space (monthly standard deviation) and time (vegetation peak) for California. Only the top 10 classes with lowest mean monthly standard deviation are listed.

Class ID	Class Description	# of pixels	Monthly Standard Deviation		Veg. Peak (day of year)	
			Mean	St. Dev	Mean	St. Dev
121	Developed/Open Space	394022	0.004	0.176	82.9	70.9
77	Pears	2986	0.008	0.108	182.7	45.0
229	Pumpkins	2200	0.008	0.170	207.4	56.4
68	Apples	127	0.009	0.135	169.2	60.9
74	Pecans	764	0.009	0.200	138.2	94.1
223	Apricots	565	0.010	0.162	158.9	83.3
204	Pistachios	48968	0.010	0.149	161.5	79.5
218	Nectarines	1954	0.011	0.140	190.2	50.4
190	Woody Wetlands	21761	0.011	0.182	131.9	70.7
123	Developed/Med Intensity	284123	0.011	0.109	111.7	63.3

The tables examine two different values. First is the mean monthly in-class standard deviation which gives an indication of how curves from the same class vary in space (by index value). The standard deviation the monthly standard deviation values, in turn, gives an indication of variation in time.

Figure 29 below provides some graphical examples of what these values mean. The left-most plot for pears in California shows low mean monthly standard deviation as well as low variation in time which matches the data in Table 23. The plot of flax in the Netherlands, on the other hand, shows high variability of monthly standard deviation in time which is displayed as variation in the width of the standard deviation band and a high standard deviation between monthly values in Table 22. The right-most plot for uncultivated areas is an example of a class with high mean monthly standard deviation with a value of 0.196 and a thick standard deviation band in the figure.

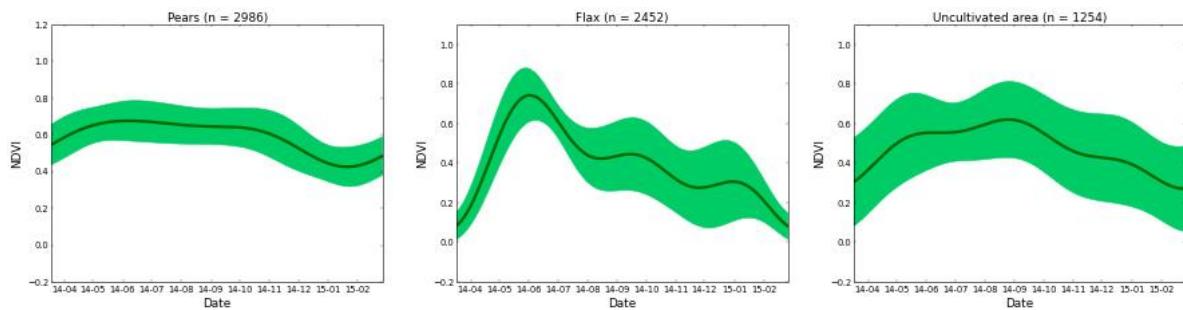


Figure 29. Examples of variation in standard deviation between time series NDVI curves for garlic (left, CA), flax (center, CL) and uncultivated area (right, NL). The shaded bands show ± 1 standard deviation.

Second, the standard deviations of the vegetation peaks give another measure of in-class variability in time. Figure 30 shows 20 randomly selected time series NDVI curves for corn and grapes in California. The plot for corn on the left shows high variation in peak vegetation with a standard deviation of 54 days. This indicates that, statistically, only 68% of corn vegetation peaks will fall within 54 days of the mean and the remaining pixels will have peak outside the roughly 3.5 month window. This exemplifies the difficulty posed by in-class variability in time.

The plot for grapes on the right shows even more troublesome behavior. The time series curves vary greatly in both form and amplitude. This is likely due a wide variety of grape species being grown using various vineyard configurations such as differentially spaced horizontal and vertical trellises.

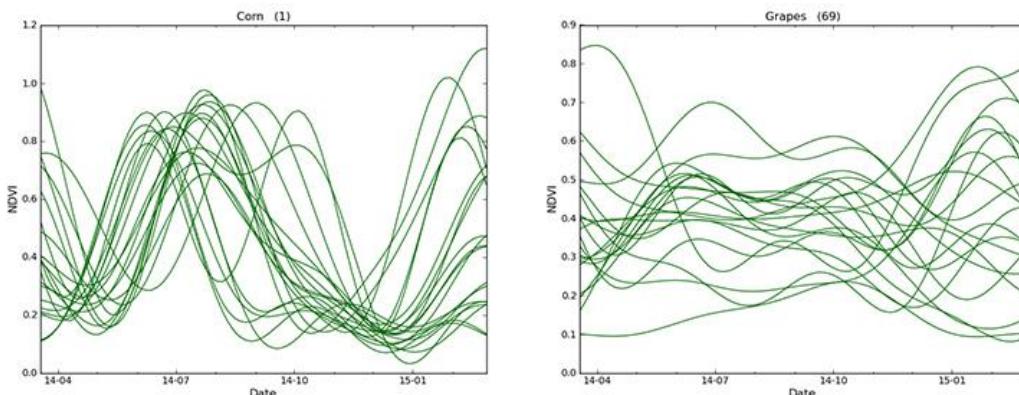


Figure 30. Examples of variation within classes in California for corn (left) and grapes (right)

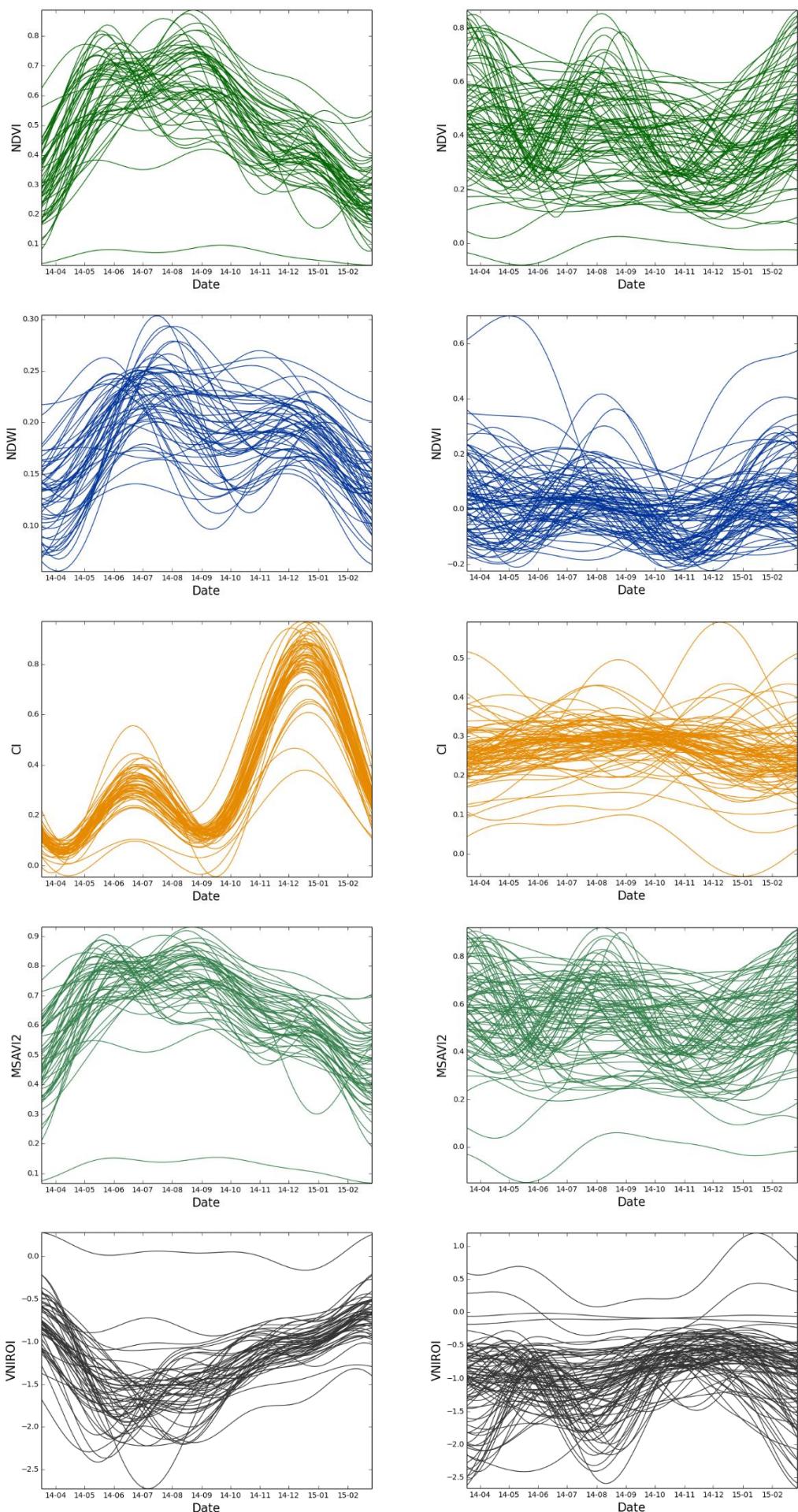


Figure 31. Mean class index values for the Netherlands (left) and California (right)

5.3.2 Similarities between Classes

In order to examine similarities between classes, the results were first compared visually. Figure 31 on the previous page shows the mean value curves for each ground truth class for both the Netherlands (left) and California (right). Ideally, you'd want to see as much variation as possible both in the spacing of peaks and valleys as well as amplitudes of such features.

The soil index which is based solely on visible reflectance stands out in the Netherlands for its lack of variation. This is not surprising since literature suggests the index is useful in distinguishing the composition of bare soils which are rare in the Netherlands. The results from California, where soil background is often visible in the dry climate, show more variability.

Next, similarity matrices were calculated for each index at each location in order to provide a more quantitative indication of similarity between classes. Appendix section 8.5 provides the full similarity matrices for each study area. This data indicates that many classes have similar time series characteristics, often caused by similar growing seasons.

5.4 Experimental 100m PROBA-V Data versus Synthesis Products

After nearly a year of PROBA-V operation the experimental 100m camera data was released to the public in March 2015. Instead of the raw level 3 top-of-canopy data products, VITO reprocessed the data archive into so-called “synthesis products” which are offered as daily or 5-day products. This is achieved by a compositing technique which is described in detail in the PROBA-V user’s manual. In summary, the compositing maximizes coverage while minimizing clouding and viewing angle differences (Wolters et al. 2015).

Coverage of the 100m sensor is limited by its small swath and its 5-day coverage is summarized in Figure 32 below. Thus compositing techniques have more impact at higher latitudes since lower latitudes may only have one observation over the compositing window.

A comparison of attained classification accuracy in the Stockton area, a subset of the California study area, found similar results using the experimental 100m data and the publicly available synthesis products. Thus, it is unlikely that the new synthesis products will have significant effects on the results found in this report.

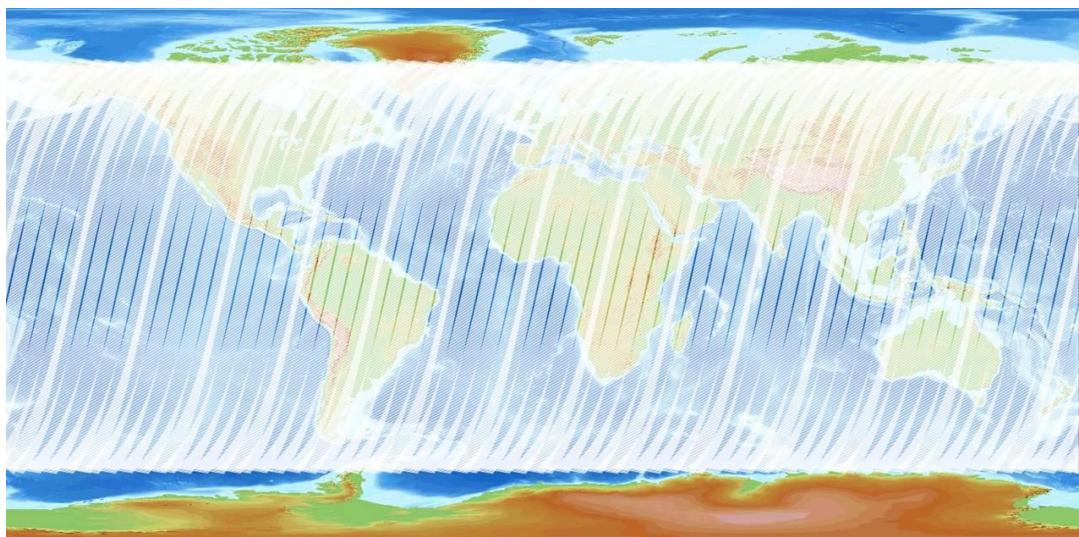


Figure 32. PROBA-V 5-day coverage for 100m sensor (VITO)

6 Recommendations

6.1 Higher Spectral Resolution

While the PROBA-V provides data with excellent spatial and temporal resolution, it is lacking in spectral detail. With the two visible and two infrared bands available on the sensor, this research has found a lack in distinct spectral behavior in time as discussed in section 5.3.2 previously and supported by the similarity matrices in appendix section 8.5.

Furthermore, the results in Table 14 revealed that traditional unsupervised classification using band reflectance data from 3 images outperformed the classification techniques used in this report. This suggests that perhaps hyperspectral data is more useful in land and water use classification compared to hyper-temporal data as provided by the PROBA-V. However, the high temporal resolution of the PROBA-V will likely prove useful in other applications such as vegetation monitoring and change detection.

Unfortunately, hyperspectral data isn't currently available with reasonable temporal and spatial resolution. The future launch of the Sentinel-2 satellite could provide the data needed to test the effectiveness of using this additional spectral data for land and water use classification.

6.2 A Focus on the Details

Due to the scope of this research, these techniques were only tested on a very large scale. This research took all available data over a year into account and the techniques were applied over fairly large study areas with a high variety of land cover types.

It's possible that using time series information on a first pass of classification simply provides too much input information. Perhaps the real power of time series information is in the details. Maybe, on a case-to-case basis, time series information from spectral indices could be much more useful. The effectiveness of the techniques in distinguishing between very specific land cover types could be used in addition to traditional classification techniques. Perhaps this time series information could fill in the gaps in differentiation left by existing classifiers even if it is of limited use on its own. Land use classes with low in-class variation, such as those in Table 22 and Table 23, might be good candidates for this sort of analysis.

6.3 Selection of Appropriate Study Areas

The findings in this report indicate that at least two main factors have some impact on the accuracy of the classification results. Table 11 showed that larger contiguous areas exhibited higher classification performance which suggests that the mixed-pixel effect is responsible for some degradation in accuracy. The lower performance of the Netherlands in unsupervised classification, as shown in Table 19, in combination with the frequency distributions of vegetation peaks shown in Figure 20 suggests that the classification results are likely impacted by the relatively high seasonality in the Netherlands compared to the California study area.

Thus, it is recommended that further attempts at time series based classification take these factors into account when selected study areas. Ideally, the study area should have larger contiguous land cover classes which are not too strongly linked to seasonality.

6.4 RMSE Classifier

The root-mean-square error based classifier didn't perform particularly well in this research but did show some promise. The relatively high performance of the technique when the top 5 best class guesses were taken into account, in addition to the example in Figure 27, show that there is some merit in this approach.

Attempts were made to improve the accuracy by normalization using the standard deviation at each time step, weighting according to the vegetation peaks and by using the first derivative curves. Although these attempts weren't particularly fruitful, they only represent a small subset of different approaches, weighting or otherwise, that could be applied to an RMSE classifier. Based on the findings in this report, further exploration of the RMSE classifier is recommended over unsupervised classification of harmonic components for approaches using time series data.

The idea of a weighted RMSE was only lightly touched upon during this research but may have potential. It is intuitive that not all time series information should be treated equally. The challenge becomes identification of which time periods in particular contain the most useful information for comparative analysis. The shift from the analysis of data from a limited number of observations to a one year time series introduces a lot of additional information so perhaps a more bottom up approach would work better.

A bottom up approach could start by looking at a small number of classes where vegetation, moisture and spectral characteristics are already well understood. Examining the details in these time series curves may help in finding which parts of the time series curves are more important and which are less important. Perhaps specific harmonic components will show correlation to very specific processes that are only present in a limited number of classes. This information could be used to create classifiers which are less applicable in a large scale, but potentially more accurate and useful for discerning more detailed information.

Another approach which would attempt to focus on the more useful parts of the time series curves could involve measuring RMSE over a moving time window. The RMSE could be calculated at given time steps, say every 10 days, using data from the preceding and following 30 day periods. Then the time windows with the lowest RMSE would be compared for each class. Of course, the size of the time steps and the length of the window would need to be optimized.

6.5 Decision Trees

One major flaw of this research was the potentially inappropriate application of spectral indices. Each index was only supported for a limited number of land cover types in literature. However, in the classification performed in this report, the spectral indices were applied to the entirety of both study areas. A better approach could be incorporated with a decision tree based approach. For example, the crust index would only be applied once a pixel was on a "soil" branch.

Decision trees based on time series information could be created manually or built using information gleaned from data mining which has already proven to be useful in land use classification (Pal & Mather 2003). Manual decision tree building should take the important parameters of vegetation cover, moisture and development intensity which are shown in Figure 3, into consideration when deliberating structure.

Still, it is worth noting that decision trees are often limited by their specificity and the original aim of this research was to explore more generally applicable classifiers. This goal seems difficult to obtain since time series behavior of land use classes will always be linked to location in some degree.

6.6 More Relevant Ground Truth Data

The original motivation of this research must be considered once more. There is a continuing need for land and water use classification as outlined back in Table 2. The ground truth data used for both study sites, while detailed, still lacks the link to climate and hydrological processes which are required in water management and hydrology applications such as Water Accounting+.

Thus, it is recommended that future attempts at classification seeking to meet the needs of water managers and hydrologists focus on this link when selecting ground truth data to collect or utilize. Again, the example raised of richer detailed in RMSE classification compared to the ground truth in Figure 27 comes to mind. The importance of the parameters of vegetation cover, moisture and development intensity, as summarized in Figure 33, should be considered when selecting appropriate ground truth data for comparison.

Maps of traditional hydrological parameters might prove useful in meeting water used focused classification needs. Examples of such parameters include evapotranspiration, soil water content, infiltration, and runoff characteristics. Since the interactions between different land use types and these processes are generally understood, this approach could aid in classification and is thus recommended for future classification attempts of this nature.

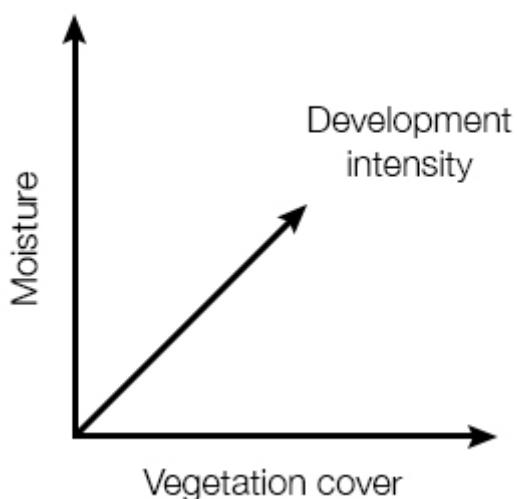


Figure 33. Summary of import parameters in land and water use classification.

7 References

- Allen, Damien. "Identifying appropriate spectral indices for land and water use classification by means of time series analysis" (2014). Additional Thesis.
- Biehl, Larry L., and E. Stoner. "Reflectance properties of soils." *Adv. Agron* 38 (1985): 1-44.
- Bracewell, Ron. "The Fourier Transform and Its Applications." New York (1965).
- Centraal Bureau voor de Statistiek. Bestand Bodemgebruik 2010.
<https://www.pdok.nl/nl/service/wfs-cbs-bestand-bodemgebruik-2010-bbg-2010> (accessed 15/02/10)
- CIMIS (California Irrigation Management Information System). Monthly report for Modesto (Mar 2014 – Feb 2015). <http://wwwcimis.water.ca.gov> (accessed 29.05.15)
- Civco, Daniel L. "Artificial neural networks for land-cover classification and mapping." *International Journal of Geographical Information Science* 7.2 (1993): 173-186.
- Clark, R.N., Swayze, G.A., Wise, R., Livo, E., Hoefen, T., Kokaly, R., Sutley, S.J., 2007, USGS digital spectral library splib06a: U.S. Geological Survey, Data Series 231. (2007)
- Congalton, Russell G., and Kass Green. Assessing the accuracy of remotely sensed data: principles and practices. CRC press, 2008.
- ESA. 2010. GMES Sentinel-2 Mission Requirements Document. esamultimedia.esa.int/docs/S2-Data_Sheet.pdf (accessed 03.11.14).
- Fourier, Joseph. Theorie analytique de la chaleur, par M. Fourier. Chez Firmin Didot, père et fils, 1822.
- Friedl, Mark A., and Carla E. Brodley. "Decision tree classification of land cover from remotely sensed data." *Remote sensing of environment* 61.3 (1997): 399-409.
- GRASS Development Team, 2012. Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation Project. <http://grass.osgeo.org>
- Gao, Bo-Cai. "NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space." *Remote sensing of environment* 58.3 (1996): 257-266.
- Huete, A. R., et al. "A comparison of vegetation indices over a global set of TM images for EOS-MODIS." *Remote sensing of environment* 59.3 (1997): 440-451.
- Huete, A., et al. "Overview of the radiometric and biophysical performance of the MODIS vegetation indices." *Remote sensing of environment* 83.1 (2002): 195-213.
- Jolliffe, I. T. (2002). Analysis, second edition Springer-Verlag. ISBN 978-0-387-95442-4.
- Karimi, Poolad. "Water Accounting Plus for Water Resources Reporting and River Basin Planning." Thesis. Delft University of Technology (2014).
- Karnieli, Arnon. "Development and implementation of spectral crust index over dune sands." *International Journal of Remote Sensing* 18.6 (1997): 1207-1220.

- Kauth, Richard J., and G. S. Thomas. "The tasseled cap--a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat." *LARS Symposia*. 1976.
- KNMI (Royal Netherlands Meteorological Institute). De Bilt Daily Weather Data.
<http://www.knmi.nl/klimatologie/daggegevens/download.html> (accessed 29.05.15)
- Kriegler, F. J., et al. "Preprocessing transformations and their effects on multispectral recognition." *Remote Sensing of Environment*, VI. Vol. 1. 1969.
- Lang, Ruili, et al. "Optimizing unsupervised classifications of remotely sensed imagery with a data-assisted labeling approach." *Computers & Geosciences* 34.12 (2008): 1877-1885.
- Li, Hua, and Qinhuo Liu. "Comparison of NDBI and NDVI as indicators of surface urban heat island effect in MODIS imagery." *International Conference on Earth Observation Data Processing and Analysis*. International Society for Optics and Photonics, 2008.
- Martins, Carla. "Farm Structure in the Netherlands" Eurostat, 2008.
- Menenti, M. (1984): Physical aspects and determination of evaporation in deserts applying remote sensing techniques. Report 10, Inst. Land Water Manag. Res. (IWC). Wageningen, Niederlande; 202 Seiten, 117 Abb., 8 Farbilder, 55 Tab., 45,- Dfl
- Menenti, M., et al. "Mapping agroecological zones and time lag in vegetation growth by means of Fourier analysis of time series of NDVI images." *Advances in Space Research* 13.5 (1993): 233-237.
- Ministerie van Economische Zaken. Basisregistratie Percelen (BRP) Gewaspercelen 2014.
<https://data.overheid.nl/data/dataset/brp-gewaspercelen> (accessed 10/02/2015)
- Pal, Mahesh, and Paul M. Mather. "An assessment of the effectiveness of decision tree methods for land cover classification." *Remote sensing of environment* 86.4 (2003): 554-565.
- Roerink, G. J., M. Menenti, and W. Verhoef. "Reconstructing cloudfree NDVI composites using Fourier analysis of time series." *International Journal of Remote Sensing* 21.9 (2000): 1911-1917.
- Scholz, Matthias. Approaches to analyse and interpret biological profile data. Diss. Universitätsbibliothek, 2006.
- USDA National Agricultural Statistics Service Cropland Data Layer. 2014. Published crop-specific data layer [Online]. Available at <http://nassgeodata.gmu.edu/CropScape/> (accessed 10/02/2015). USDA-NASS, Washington, DC.
- USDA National Agricultural Statistics Service. 2012. California Agricultural Statistics, Crop Year 2011.
- Wolters, Erwin, Wouter Dierckx, Jan Dries and Else Swinnen. 2015. PROBA-V Products User Manual
http://proba-v.vgt.vito.be/sites/default/files/PROBAV-Products_User_Manual_v1.2.pdf (accessed 20/03/2015)
- Verhoef, W., M. Menenti, and S. Azzali. "Cover A colour composite of NOAA-AVHRR-NDVI based on time series analysis (1981-1992)." *International Journal of Remote Sensing* 17.2 (1996): 231-235.
- Zha, Yong, Jay Gao, and Shaoxiang Ni. "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery." *International Journal of Remote Sensing* 24.3 (2003): 583-594.

8 Appendix

8.1 Parameters for Unsupervised Classification

The following parameters were used as inputs for GRASS's *i.cluster* module. Selection of parameters was chosen in order to maximize the number of unsupervised classes. Resulting signatures are passed to the *i.maxlik* module for maximum likelihood analysis.

Table 24. Default input parameters for cluster analysis

Parameter	Value
Number of unsupervised classes	255
Minimum separation	1.0
Convergence	99.0
Max iterations	50
Min cluster size	17

8.2 Ground Truth Legends

The followings pages contain tables describing the ground truth data used in each study area. Table 25 and

Table 26 show the original classes in the BBG and BRP datasets respectively and also show how classes were reclassified to form the combined BRP/BBG dataset used in this report.

Table 27 and Table 28 show how classes in the combined BRP/BBG and CDL datasets were reclassified into reduced class versions used in section 4.2.2.

8.2.1 The Netherlands

Table 25. Original BBG classes and reclassified categories for combined BRP/BBG

BBG ID	BBG Description	BRP/BBG ID	BRP/BBG Description
10	Spoorterrein	20	Developed
11	Wegverkeersterrein		
12	Vliegveld		
20	Woonterrein		
21	Terrein voor detailhandel en horeca		
22	Terrein voor openbare voorzieningen		
23	Terrein voor sociaal-culturele voorzieningen		
24	Bedrijventerrein		
30	Stortplaats		
31	Wrakkenopslagplaats		
32	Begraafplaats		
33	Delfstofwinplaats		
34	Bouwterrein		
35	Semi verhard overig terrein		
40	Park en plantsoen	40	Parks and recreation
41	Sportterrein		
42	Volkstuin		
43	Dagcreatief terrein		
44	Verblijfscreatief terrein		
50	Terrein voor glastuinbouw	50	Greenhouse
51	Overig agrarisch terrein	51	Agriculture
60	Bos	60	Forest
61	Open droog natuurlijk terrein	61	Dry natural areas
62	Open nat natuurlijk terrein	62	Wetlands
70	IJsselmeer/Markermeer	70	Open water
71	Afgesloten zeearm		
72	Rijn en Maas		
73	Randmeer		
74	Spaarbekken		
75	Recreatief binnenwater		
76	Binnenwater voor delfstofwinning		
77	Vloe- en/of slibveld		
78	Overig binnenwater		
80	Waddenzee, Eems, Dollard		
81	Oosterschelde		
82	Westerschelde		
83	Noordzee		
90	Foreign country	NULL	Foreign country

Table 26. Original BRP classes and reclassified categories for combined BRP/BBG

BRP ID	BRP Description	BRP/BBG ID	BRP/BBG Description
265	Grasland, blijvend	265	Grassland
266	Grasland, tijdelijk		
3718	Grasland, natuurlijk		
3722	Overige natuurterreinen	3722	Other natural areas
3719	Heide	3719	Heather
2029	Braak, Natuur	2029	Fallow (natural)
2026	Braak, met bos (SBL-regeling na 28 juni 1995)	2026	Fallow (forested)
2300	Onbeteelde grond vanwege een teeltverbod/ontheffing	2300	Uncultivated area
2033	Onbeteelde grond, tijdelijk		
2025	Aardappelen, bestrijdingsmaatregel AM	2025	Potatoes
2951	Aardappelen, consumptie, op klei/lössgrond		
3792	Aardappelen, consumptie, op zand/veengrond		
3730	Aardappelen, poot, op klei/lössgrond		
3731	Aardappelen, poot, op zand/veengrond		
3732	Aardappelen, zetmeel		
256	Bieten, suiker-	256	Sugar beets
2651	Bieten, voeder- (inclusief aardperen)		
247	Blauwmaanzaad	247	Poppy
242	Bonen, bruine-	242	Beans
853	Bonen, tuin- (droog te oogsten) (geen consumptie)		
854	Bonen, tuin- (groen te oogsten)		
243	Bonen, veld- (onder andere duive-, paarde-, wierbonen) (droog te oogsten)		
511	Cichorei	511	Chicory
2650	Erwten inclusief schokkers (droog te oogsten)	2650	Peas
244	Erwten, groene/gele, groen te oogsten		
3720	Faunaranden, grasland	3720	Fauna border
3721	Faunaranden, bouwland		
235	Gerst, winter-	235	Barley (winter)
236	Gerst, zomer-	236	Barley (summer)
2652	Granen, overig	2652	Other grain
2653	Graszaad (inclusief klaverzaad)	2653	Grass seeds
1921	Graszoden	1921	Turf
2298	Groenbemesters, vlinderbloemige-	2298	Green manure
2299	Groenbemesters, niet-vlinderbloemige-		
238	Haver	238	Oats
944	Hennep, vezel-	944	Hemp
241	Kapucijners (en grauwe erwten)	241	Marrowfat and chickpeas
246	Karwijzaad (oogst dit jaar)	246	Caraway
1922	Koolzaad, winter (ook boterzaad)	1922	Rapeseed (winter)
1923	Koolzaad, zomer (ook boterzaad)	1923	Rapeseed (summer)
666	Lijnzaad niet van vezelvas (olievas)	666	Linseed
663	Lupinen, niet bittere-	663	Lupin
258	Luzerne	258	Alfalfa
317	Maïs, corncob mix	317	Corn
316	Maïs, korrel-		
259	Maïs, snij-		
814	Maïs, suiker-		
2032	Maïs, energie-		
516	Miscanthus (olifantsgras)	516	Elephant grass

1925	Overige akkerbouwgewassen		1925	Other crops
1930	Tagetes (zand, löss) (geen groene braak)			
314	Triticale			
664	Raapzaad	664	Rapeseed	
237	Rogge (geen snijrogge)	237	Rye	
665	Sojabonen	665	Soy beans	
233	Tarwe, winter-	233	Wheat (winter)	
234	Tarwe, zomer-	234	Wheat (summer)	
1931	Uien, poot en plant (incusief sjalotten)		1931	Onion
262	Uien, zaai-			
263	Uien, zilver-			
3736	Vezelvlas	3736	Flax	
515	Zonnebloemen	515	Sunflower	
2645	Notenbomen	2645	Nut trees	
2297	Woudbomen met korte omlooptijd	2297	Walnut	
1936	Bos, blijvend, met herplantplicht		1936	Forest
863	Bos zonder herplantplicht			
864	Bos (set aside regeling)			
2027	Bos (SBL-regeling)			
176	Bloembollen en -knollen	176	Flower bulbs	
175	Bloemkwekerijgewassen (inclusief bloemzaden)	175	Flower nursery	
3806	Boomkwekerijgewassen en vaste planten, open grond	3806	Tree and plant nursery (open ground)	
294	Boomkwekerijgewassen en vaste planten, pot- en containerveld	294	Tree and plant nursery (container)	
212	Fruit	212	Fruit trees	
672	Groenten open grond (inclusief groentezaden)	672	Vegetables (open ground)	
2620	Poel en klein historisch water	2620	Water (historic)	
2621	Houtwal en houtsingel		2621	Hedge
2622	Elzensingel			
2623	Bossingel en bosje			
2624	Knip- of scheerheg			
2625	Struweelhaag			
2626	Laan	2626	Road	
2627	Knotboom	2627	Knot tree	
2628	Hoogstamboomgaard	2628	Orchard	
2629	Struweelrand	2629	Thickets	
2630	Hakhoutbosje	2630	Coppice (tree stumps)	
2631	Griendje	2631	Willow	
2632	Bomenrij en solitaire boom	2632	Tree rows	
2633	Rietzoom en klein rietperceel	2633	Reeds	
2634	Natuurvriendelijke oever	2634	Ecological bank	
2635	Wandelpad over boerenland	2635	Walking path	

Table 27. Reclassification from combined BRP/BBG dataset to reduced class version

BRP/BBG ID	BRP/BBG Description	Reduced ID	Reduced Description
20	Developed	20	Developed
40	Parks and recreation	40	Parks and recreation
50	Greenhouse	50	Greenhouse
60	Forest	60	Forest
1936	Forest		
2026	Fallow (forested)	61	Other natural areas
61	Other natural areas		
516	Elephant grass		
2029	Fallow (natural)		
3722	Other natural areas		
3719	Heather		
3720	Fauna border		
2298	Green manure		
2300	Uncultivated area		
62	Wetlands	62	Wetlands
70	Water	70	Water
175	Flower nursery	175	Flowers
176	Flower bulbs		
247	Poppy		
515	Sunflower		
212	Fruit trees	212	Tree crops & orchards
294	Tree and plant nursery (container)		
2645	Nut trees		
2297	Walnut		
3806	Tree and plant nursery (open ground)		
233	Wheat (winter)	233	Winter grains
235	Barley (winter)		
234	Wheat (summer)		
236	Barley (summer)		
237	Rye		
238	Oats	234	Summer grains
317	Corn		
265	Grassland		
1921	Turf		
2653	Grass seeds		
246	Caraway	246	Other crops
256	Sugar beets		
258	Alfalfa		
1922	Rapeseed (winter)		
672	Vegetables (open ground)		
241	Marrowfat and chickpeas		
242	Beans		
1931	Onion		
2025	Potatoes		
2650	Peas		
511	Chicory		
663	Lupin		
664	Rapeseed		
665	Soy beans		
666	Linseed		
944	Hemp		
1923	Rapeseed (summer)		
1925	Other crops		
2652	Other grain		
3736	Flax		

8.2.2 California

Table 28. Reclassification from CDL to reduced class version

CDL ID	CDL Description	Reduced ID	Reduced Description
1	Corn	1	Summer grains
3	Rice		
4	Sorghum		
12	Sweet Corn		
29	Millet		
21	Barley	21	Winter grains
22	Durum Wheat		
23	Spring Wheat		
24	Winter Wheat		
27	Rye		
28	Oats		
205	Triticale	54	Other crops
2	Cotton		
6	Sunflower		
10	Peanuts		
31	Canola		
33	Safflower		
36	Alfalfa		
41	Sugarbeets		
42	Dry Beans		
43	Potatoes		
44	Other Crops		
46	Sweet Potatoes		
47	Misc Veggies & Fruits		
48	Watermelons		
49	Onions		
53	Peas		
54	Tomatoes		
57	Herbs		
58	Clover/Wildflowers		
59	Sod/Grass Seed		
69	Grapes		
37	Other Hay/Non Alfalfa		
206	Carrots		
207	Asparagus		
208	Garlic		
209	Cantaloupes		
213	Honeydew Melons		
214	Broccoli		
216	Peppers		
219	Greens		
221	Strawberries		
222	Squash		
224	Vetch		
227	Lettuce		
229	Pumpkins		
242	Blueberries		
244	Cauliflower		

61	Fallow/Idle Cropland	61	Fallow cropland
66	Cherries	71	Tree crops & orchards
67	Peaches		
68	Apples		
71	Other Tree Crops		
72	Citrus		
74	Pecans		
75	Almonds		
76	Walnuts		
77	Pears		
204	Pistachios		
211	Olives		
212	Oranges		
217	Pomegranates		
218	Nectarines		
220	Plums		
223	Apricots		
83	Water	111	Water
92	Aquaculture		
111	Open Water		
112	Perennial Ice/Snow		
121	Developed/Open Space	121	Developed open space
122	Developed/Low Intensity	122	Developed
123	Developed/Med Intensity		
124	Developed/High Intensity		
131	Barren	131	Barren
63	Forest	141	Forest
141	Deciduous Forest		
142	Evergreen Forest		
143	Mixed Forest		
152	Shrubland	152	Shrubland
176	Grassland/Pasture	176	Grassland and pasture
87	Wetlands	190	Wetlands
190	Woody Wetlands		
195	Herbaceous Wetlands		
225	Dbl Crop WinWht/Corn	225	Double crops
226	Dbl Crop Oats/Corn		
234	Dbl Crop Durum Wht/Sorghum		
235	Dbl Crop Barley/Sorghum		
236	Dbl Crop WinWht/Sorghum		
237	Dbl Crop Barley/Corn		
238	Dbl Crop WinWht/Cotton		

8.3 Performance of Different Index Combinations

The performance of different combinations of spectral indices was explored in section 4.2.1 but only the top combinations were provided. Table 29 and Table 30 provide the full results of this analysis for the Netherlands and California study areas respectively.

Table 29. Full unsupervised classification results of different index combinations in the Netherlands

NDVI	NDWI	CI	MSAVI2	VNIROI	% Accuracy
■	■	■	■	■	43.8
■	■		■	■	43.7
■	■			■	43.7
■	■	■		■	43.6
	■		■	■	43.6
■	■	■	■		43.6
	■	■	■	■	43.6
■	■		■		43.5
	■	■	■		43.4
	■		■		43.4
■			■		43.4
	■	■			43.3
	■	■		■	43.3
		■		■	43.3
■			■	■	43.2
■		■	■	■	43.2
		■	■	■	43.2
■	■	■			43.2
■	■				43.2
	■			■	43.2
			■	■	43.1
■		■		■	43.0
■		■	■		43.0
■				■	43.0
■		■			43.0
		■	■		42.9

Table 30. Full unsupervised classification results of different index combinations in California

NDVI	NDWI	CI	MSAVI2	VNIROI	% Accuracy
■	■			■	51.4
■	■	■	■	■	51.2
	■	■	■	■	50.9
■	■	■	■		49.7
■		■		■	49.7
■		■	■	■	49.6
		■	■	■	49.3
■	■	■			49.3
	■	■		■	49.2
	■	■	■		49.0
■	■	■		■	48.2
■	■		■	■	48.1
	■		■	■	47.9
■		■	■		46.9
■		■			46.8
■			■	■	46.7
		■	■		46.5
■				■	46.5
		■		■	46.2
■	■		■		46.0
	■			■	45.8
			■	■	45.8
■	■				44.9
	■		■		44.8
	■	■			42.4
■			■		41.1

8.4 In-class Variation

Table 31. Full table detailing in-class variation of NDVI images for the Netherlands.

Class ID	Class Description	# of pixels	Monthly St. Deviation		Veg. Peak (day of year)	
			Mean	St. Dev	Mean	St. Dev
20	Developed	671490	0.187	0.012	207.1	27.9
40	Parks and recreation	128124	0.156	0.017	207.6	26.1
50	Greenhouse	21135	0.177	0.017	214.9	80.8
60	Forest	437204	0.150	0.033	209.6	23.4
61	Other natural areas	87711	0.219	0.030	218.1	37.0
62	Wetlands	47575	0.149	0.020	208.3	15.7
175	Flower nursery	3248	0.167	0.015	211.9	36.9
176	Flower bulbs	31078	0.182	0.023	214.5	52.2
212	Fruit trees	24568	0.128	0.034	213.9	23.6
233	Wheat (winter)	162588	0.175	0.033	160.2	42.8
234	Wheat (summer)	26379	0.166	0.034	192.9	38.3
235	Barley (winter)	7508	0.176	0.032	162.7	60.3
236	Barley (summer)	29479	0.162	0.032	199.1	37.7
237	Rye	2546	0.158	0.028	183.0	52.7
238	Oats	2375	0.175	0.027	199.6	37.9
241	Marrowfat and chickpeas	354	0.153	0.043	205.3	35.6
242	Beans	4926	0.162	0.034	215.8	38.0
246	Caraway	26	0.156	0.042	183.5	43.8
247	Poppy	679	0.144	0.038	208.5	35.0
256	Sugar beets	100435	0.146	0.029	222.8	20.4
258	Alfalfa	6898	0.176	0.046	203.0	34.9
265	Grassland	1948913	0.157	0.036	201.2	39.7
294	Tree and plant nursery (container)	1558	0.156	0.019	232.6	31.4
317	Corn	327281	0.162	0.041	222.1	24.4
511	Chicory	4619	0.144	0.027	229.7	17.4
515	Sunflower	519	0.172	0.033	227.3	40.1
516	Elephant grass	264	0.150	0.016	222.7	19.1
663	Lupin	139	0.139	0.024	205.6	29.2
664	Rapeseed	46	0.142	0.022	203.2	32.9
665	Soy beans	155	0.159	0.038	216.7	26.5
666	Linseed	5	0.059	0.039	198.6	4.3
672	Vegetables (open ground)	57334	0.168	0.017	238.0	28.8
944	Hemp	2199	0.159	0.039	207.7	27.0
1921	Turf	2596	0.167	0.032	208.2	71.2
1922	Rapeseed (winter)	3262	0.175	0.043	126.7	55.1
1923	Rapeseed (summer)	862	0.178	0.033	216.8	43.1
1931	Onion	39765	0.162	0.037	226.3	41.1
1936	Forest	14207	0.152	0.035	204.6	22.7
2025	Potatoes	210401	0.173	0.028	208.2	39.0
2026	Fallow (forested)	72	0.133	0.035	196.6	17.8
2029	Fallow (natural)	1494	0.159	0.017	208.8	33.4
2297	Walnut	26	0.067	0.022	222.4	6.5
2298	Green manure	2971	0.177	0.015	219.9	42.9
2300	Uncultivated area	1264	0.196	0.019	216.7	53.4
2645	Nut trees	84	0.150	0.039	204.1	24.8
2650	Peas	5146	0.179	0.033	222.6	38.6
2652	Other grain	1073	0.185	0.024	195.0	44.4
2653	Grass seeds	15806	0.162	0.038	174.1	35.9
3719	Heather	17257	0.155	0.019	220.8	24.2
3720	Fauna border	4258	0.175	0.025	197.4	40.5
3722	Other natural areas	21913	0.154	0.010	201.9	21.8
3736	Flax	2597	0.133	0.044	173.2	40.4
3806	Tree and plant nursery (open ground)	19790	0.146	0.027	235.4	30.9

Table 32. Full table detailing in-class variation of NDVI images for California.

Class ID	Class Description	# of pixels	Monthly St. Deviation		Veg. Peak (day of year)	
			Mean	St. Dev	Mean	St. Dev
1	Corn	68983	0.037	0.199	153.6	54.1
2	Cotton	73496	0.046	0.137	212.0	56.5
3	Rice	55042	0.021	0.120	196.3	26.4
4	Sorghum	2721	0.050	0.189	164.1	67.9
6	Sunflower	17463	0.041	0.140	143.1	41.1
12	Sweet Corn	177	0.029	0.136	157.0	51.1
21	Barley	10342	0.065	0.174	83.2	71.1
22	Durum Wheat	4552	0.032	0.119	94.7	27.1
23	Spring Wheat	1617	0.037	0.198	87.9	60.9
24	Winter Wheat	121905	0.055	0.201	86.5	59.1
27	Rye	260	0.030	0.190	109.7	47.0
28	Oats	46620	0.039	0.212	97.4	70.9
31	Canola	259	0.040	0.062	80.2	14.0
33	Safflower	16087	0.045	0.160	128.2	40.1
36	Alfalfa	277334	0.016	0.195	134.0	82.3
37	Other Hay/Non Alfalfa	67588	0.014	0.197	93.3	67.3
41	Sugarbeets	110	0.035	0.133	98.6	17.8
42	Dry Beans	9143	0.046	0.191	146.1	62.4
43	Potatoes	1105	0.035	0.196	170.0	64.6
44	Other Crops	2269	0.038	0.196	159.9	83.9
46	Sweet Potatoes	476	0.032	0.178	174.5	68.6
47	Misc Veggies & Fruits	2123	0.016	0.185	206.2	69.6
48	Watermelons	959	0.041	0.169	200.6	60.4
49	Onions	7019	0.031	0.146	145.4	39.2
53	Peas	1384	0.045	0.176	144.2	78.8
54	Tomatoes	130395	0.034	0.170	179.1	52.9
55	Caneberries	18	0.036	0.138	178.2	63.3
57	Herbs	521	0.060	0.201	119.9	59.9
58	Clover/Wildflowers	17428	0.014	0.136	80.6	56.6
59	Sod/Grass Seed	2025	0.016	0.225	160.6	88.6
61	Fallow/Idle Cropland	311734	0.040	0.164	110.4	80.3
66	Cherries	6676	0.016	0.148	162.3	68.1
67	Peaches	1442	0.023	0.117	192.6	58.5
68	Apples	127	0.009	0.135	169.2	60.9
69	Grapes	380271	0.028	0.138	147.0	88.3
71	Other Tree Crops	1352	0.014	0.170	134.3	97.8
72	Citrus	2153	0.022	0.164	101.5	121.5
74	Pecans	764	0.009	0.200	138.2	94.1
75	Almonds	304997	0.017	0.145	137.3	65.6
76	Walnuts	95788	0.015	0.157	171.1	65.8
77	Pears	2986	0.008	0.108	182.7	45.0
92	Aquaculture	24	0.019	0.132	109.5	40.8
111	Open Water	164106	0.037	0.227	152.6	112.6
121	Developed/Open Space	394022	0.004	0.176	82.9	70.9
122	Developed/Low Intensity	225120	0.014	0.137	100.6	74.1
123	Developed/Med Intensity	284123	0.011	0.109	111.7	63.3
124	Developed/High Intensity	74881	0.012	0.085	129.0	60.3
131	Barren	156895	0.027	0.116	212.1	72.8
141	Deciduous Forest	68823	0.025	0.115	72.7	32.4
142	Evergreen Forest	1888296	0.031	0.202	178.5	111.1
143	Mixed Forest	383647	0.022	0.126	80.5	47.7
152	Shrubland	1991433	0.032	0.200	130.8	90.8
176	Grassland/Pasture	2278582	0.029	0.134	54.8	40.4
190	Woody Wetlands	21761	0.011	0.182	131.9	70.7
195	Herbaceous Wetlands	100327	0.026	0.150	133.2	64.6
204	Pistachios	48968	0.010	0.149	161.5	79.5

205	Triticale	14978	0.033	0.186	92.0	49.2
206	Carrots	593	0.056	0.171	193.0	54.8
207	Asparagus	883	0.015	0.180	206.3	79.2
208	Garlic	5715	0.037	0.105	111.9	27.0
209	Cantaloupes	1963	0.036	0.155	190.2	57.2
211	Olives	7564	0.026	0.124	132.5	120.0
212	Oranges	58631	0.025	0.132	84.1	103.7
213	Honeydew Melons	1392	0.035	0.154	191.7	48.3
214	Broccoli	163	0.063	0.157	126.8	129.8
216	Peppers	409	0.051	0.161	170.8	60.8
217	Pomegranates	4320	0.022	0.131	180.4	91.0
218	Nectarines	1954	0.011	0.140	190.2	50.4
219	Greens	112	0.060	0.132	202.5	144.8
220	Plums	5917	0.018	0.135	151.1	74.2
221	Strawberries	1789	0.020	0.177	182.3	82.1
222	Squash	600	0.035	0.176	140.7	88.5
223	Apricots	565	0.010	0.162	158.9	83.3
224	Vetch	262	0.064	0.147	67.8	39.5
225	Dbl Crop WinWht/Corn	72821	0.027	0.174	117.4	52.9
226	Dbl Crop Oats/Corn	28468	0.025	0.173	117.3	66.8
227	Lettuce	2074	0.051	0.147	118.6	101.4
229	Pumpkins	2200	0.008	0.170	207.4	56.4
234	Dbl Crop Durum Wht/Sorghum	14	0.053	0.094	187.6	64.8
235	Dbl Crop Barley/Sorghum	397	0.046	0.182	112.5	76.1
236	Dbl Crop WinWht/Sorghum	8967	0.036	0.180	120.4	61.4
237	Dbl Crop Barley/Corn	328	0.049	0.162	103.3	41.7
238	Dbl Crop WinWht/Cotton	41	0.069	0.128	167.0	41.3
242	Blueberries	108	0.013	0.151	90.7	44.7

8.5 Similarity Matrices

The following pages contain tables of similarity matrices between classes. The first row/column corresponds to the ID of the class for the combined BRP/BBG and CDL datasets as listed in Table 27 and Table 28 respectively. The similarity is represented by the RMSE between mean index curves between the classes.

Table 33. NDVI similarity matrix for BRP/BBG dataset (NL)

Table 34. NDWI similarity matrix for BRP/BBG dataset (NL)

cat	20	40	50	60	61	62	175	176	212	233	234	235	236	237	238	241	242	246	247	256	258	265	294	317	511	515	516	663	664	665	666	673	944	1921	1922	1923	1924	1925	1926	2029	2297	2298	2300	2645	2650	2652	2653	3719	3720	3722	3736	3806											
20	-	0.49	0.34	0.66	0.18	0.58	0.28	0.17	0.86	0.44	0.11	0.47	0.26	0.54	0.25	0.37	0.28	0.24	0.41	0.66	0.53	0.96	0.48	0.54	0.64	0.19	0.45	0.49	0.27	0.42	0.79	0.47	0.20	0.71	0.77	0.28	0.24	0.71	0.16	0.70	0.49	0.91	0.27	0.22	0.79	0.24	0.35	0.46	0.16	0.41	0.15	0.34	0.51										
40	0.49	-	0.32	0.19	0.61	0.30	0.25	0.33	0.38	0.62	0.43	0.28	0.26	0.15	0.24	0.83	0.69	0.48	0.87	0.57	0.17	0.50	0.30	0.22	0.69	0.46	0.21	0.12	0.25	0.24	0.31	0.52	0.41	0.35	0.81	0.25	0.62	0.23	0.52	0.22	0.10	0.55	0.31	0.28	0.30	0.57	0.17	0.31	0.59	0.09	0.38	0.79	0.30										
50	0.34	0.32	-	0.49	0.42	0.58	0.31	0.27	0.67	0.54	0.37	0.28	0.25	0.28	0.27	0.66	0.60	0.45	0.73	0.75	0.46	0.68	0.37	0.44	0.78	0.37	0.42	0.28	0.30	0.41	0.58	0.55	0.39	0.43	0.64	0.27	0.56	0.53	0.44	0.52	0.37	0.86	0.28	0.20	0.60	0.46	0.24	0.41	0.42	0.29	0.34	0.67	0.37										
60	0.66	0.19	0.49	-	0.78	0.31	0.40	0.51	0.19	0.79	0.60	0.44	0.42	0.27	0.42	0.95	0.83	0.65	1.03	0.56	0.24	0.39	0.35	0.21	0.71	0.61	0.27	0.22	0.42	0.31	0.14	0.58	0.55	0.32	0.91	0.40	0.76	0.05	0.67	0.11	0.18	0.41	0.45	0.46	0.13	0.70	0.36	0.46	0.75	0.27	0.54	0.96	0.34										
61	0.18	0.61	0.42	0.78	-	0.72	0.40	0.34	0.97	0.59	0.28	0.62	0.36	0.66	0.39	0.24	0.27	0.40	0.33	0.73	0.67	1.06	0.53	0.63	0.64	0.20	0.56	0.59	0.42	0.52	0.90	0.46	0.30	0.77	0.87	0.38	0.26	0.82	0.21	0.83	0.60	1.03	0.33	0.33	0.90	0.19	0.50	0.63	0.04	0.55	0.30	0.36	0.55										
62	0.58	0.30	0.58	0.31	0.72	-	0.33	0.43	0.44	0.68	0.48	0.50	0.40	0.45	0.36	0.88	0.67	0.48	0.89	0.39	0.15	0.68	0.44	0.30	0.58	0.54	0.23	0.39	0.33	0.25	0.44	0.53	0.43	0.59	1.02	0.38	0.61	0.33	0.54	0.32	0.26	0.35	0.44	0.38	0.62	0.36	0.41	0.68	0.30	0.43	0.80	0.46											
175	0.28	0.25	0.31	0.40	0.4	0.33	-	0.16	0.59	0.51	0.21	0.38	0.09	0.36	0.07	0.59	0.44	0.30	0.63	0.46	0.28	0.74	0.28	0.28	0.52	0.23	0.18	0.28	0.09	0.15	0.54	0.36	0.15	0.52	0.85	0.07	0.37	0.45	0.27	0.45	0.23	0.64	0.14	0.13	0.53	0.33	0.19	0.36	0.37	0.39	0.14	0.57	0.31										
176	0.17	0.33	0.27	0.51	0.34	0.43	0.16	-	0.70	0.40	0.12	0.33	0.17	0.39	0.10	0.53	0.40	0.19	0.56	0.60	0.36	0.81	0.41	0.42	0.64	0.25	0.32	0.36	0.11	0.31	0.64	0.47	0.18	0.60	0.73	0.17	0.34	0.55	0.24	0.53	0.35	0.77	0.22	0.12	0.63	0.34	0.18	0.31	0.25	0.10	0.47	0.43											
212	0.86	0.38	0.67	0.19	0.97	0.44	0.59	0.70	-	0.97	0.79	0.60	0.61	0.42	0.61	1.17	0.83	1.22	0.65	0.40	0.32	0.49	0.37	0.83	0.79	0.45	0.39	0.61	0.48	0.10	0.72	0.74	0.38	1.10	0.59	0.95	0.16	0.85	0.21	0.37	0.38	0.64	0.65	0.09	0.88	0.55	0.62	0.94	0.47	0.73	1.15	0.48											
233	0.44	0.62	0.54	0.79	0.59	0.68	0.54	0.40	0.97	-	0.41	0.43	0.56	0.63	0.48	0.71	0.61	0.30	0.68	0.93	0.60	1.02	0.80	0.78	0.99	0.61	0.68	0.68	0.46	0.67	0.90	0.86	0.54	0.90	0.53	0.57	0.59	0.81	0.55	0.76	0.67	1.02	0.62	0.50	0.88	0.68	0.46	0.35	0.58	0.54	0.47	0.51	0.81										
234	0.11	0.43	0.37	0.60	0.28	0.48	0.21	0.12	0.79	0.41	-	0.44	0.23	0.51	0.19	0.44	0.29	0.15	0.45	0.59	0.44	0.92	0.46	0.49	0.61	0.21	0.38	0.46	0.19	0.36	0.74	0.45	0.13	0.70	0.80	0.24	0.23	0.64	0.15	0.63	0.43	0.82	0.26	0.20	0.72	0.28	0.30	0.40	0.25	0.35	0.08	0.36	0.49										
235	0.47	0.28	0.28	0.44	0.62	0.50	0.38	0.33	0.60	0.43	0.44	-	0.38	0.21	0.32	0.84	0.73	0.44	0.87	0.80	0.35	0.59	0.52	0.49	0.90	0.55	0.45	0.33	0.32	0.46	0.51	0.72	0.50	0.50	0.54	0.38	0.67	0.45	0.57	0.41	0.37	0.78	0.44	0.34	0.51	0.19	0.19	0.61	0.24	0.43	0.75	0.52											
236	0.26	0.26	0.25	0.42	0.36	0.40	0.09	0.17	0.61	0.56	0.23	0.38	-	0.34	0.10	0.57	0.45	0.35	0.63	0.51	0.33	0.73	0.24	0.29	0.55	0.20	0.22	0.25	0.14	0.20	0.54	0.35	0.18	0.47	0.83	0.03	0.39	0.46	0.27	0.47	0.24	0.70	0.07	0.07	0.54	0.31	0.20	0.40	0.34	0.21	0.17	0.58	0.27										
237	0.54	0.15	0.28	0.27	0.66	0.45	0.36	0.39	0.42	0.63	0.51	0.21	0.34	-	0.33	0.89	0.78	0.55	0.94	0.72	0.31	0.43	0.38	0.34	0.82	0.54	0.35	0.14	0.63	0.51	0.29	0.71	0.34	0.71	0.28	0.60	0.27	0.46	0.67	0.39	0.34	0.64	0.19	0.47	0.86	0.37																	
238	0.25	0.24	0.27	0.42	0.39	0.36	0.07	0.10	0.61	0.48	0.19	0.32	0.10	0.33	-	0.60	0.45	0.27	0.63	0.53	0.28	0.73	0.32	0.32	0.59	0.26	0.23	0.28	0.05	0.21	0.55	0.42	0.18	0.52	0.76	0.09	0.39	0.46	0.28	0.45	0.25	0.68	0.17	0.09	0.54	0.36	0.14	0.55	0.34														
241	0.37	0.83	0.66	0.59	0.24	0.88	0.59	0.53	1.17	0.71	0.44	0.84	0.57	0.89	0.60	-	0.23	0.52	0.13	0.81	0.86	1.30	0.73	0.82	0.68	0.38	0.74	0.82	0.61	0.20	0.45	0.40	1.04	0.59	0.28	1.03	0.33	0.04	0.81	0.18	0.50	0.55	0.11	0.31	0.71	0.82	0.25	0.76	0.47	0.75													
242	0.28	0.69	0.60	0.83	0.27	0.67	0.44	0.40	1.01	0.61	0.29	0.73	0.45	0.78	0.45	0.23	-	0.36	0.22	0.62	0.68	1.18	0.62	0.67	0.53	0.28	0.57	0.70	0.45	0.53	0.97	0.46	0.28	0.91	1.02	0.45	0.07	0.88	0.18	0.86	0.66	0.97	0.43	0.44	0.96	0.24	0.58	0.67	0.24	0.61	0.31	0.23	0.65										
246	0.24	0.48	0.45	0.65	0.40	0.48	0.30	0.19	0.83	0.30	0.15	0.44	0.35	0.55	0.27	0.52	0.36	-	0.51	0.65	0.44	0.96	0.58	0.57	0.70	0.36	0.45	0.54	0.24	0.43	0.78	0.58	0.26	0.78	0.76	0.35	0.32	0.68	0.28	0.65	0.49	0.39	0.31	0.75	0.42	0.33	0.34	0.80	0.21	0.36	0.61	0.36	0.61	0.20	0.83	0.33	0.80	0.50	0.20	0.83			
247	0.41	0.87	0.73	0.33	0.89	0.63	0.56	1.22	0.68	0.45	0.87	0.63	0.94	0.63	0.13	0.22	0.51	-	0.83	0.88	1.36	0.80	0.87	0.72	0.45	0.78	0.88	0.64	0.74	1.17	0.64	0.49	1.08	1.06	0.64	0.29	0.19	0.87	0.61	1.16	0.39	0.75	0.83	0.30	0.80	0.50	0.20	0.83	0.33	0.80	0.50	0.20	0.83	0.33	0.80	0.50	0.20	0.83	0.33	0.80	0.50	0.20	0.83
515	0.19	0.46	0.37	0.61	0.20	0.54	0.23	0.25	0.79	0.61	0.21	0.55	0.20	0.54	0.26	0.38	0.28	0.36	0.45	0.53	0.51	0.93	0.35	0.44	0.48	-	0.37	0.44	0.28	0.33	0.73	0.29	0.14	0.64	0.92	0.21	0.23	0.66	0.12	0.67	0.43	0.33	0.16	0.21	0.74	0.11	0.38	0.55	0.17	0.40	0.17	0.45	0.38										
516	0.45	0.21	0.42	0.27	0.56	0.23	0.18	0.32	0.45	0.68	0.38	0.45	0.22	0.35	0.23	0.74	0.57	0.45	0.78	0.37	0.21	0.65	0.22	0.39	0.13	0.49	0.37	-	0.24	0.23	0.04	0.35	0.30	0.45	0.97	0.19	0.50	0.3																									

Table 35. CI similarity matrix for BRP/BBG dataset (NL)

cat	20	40	50	60	61	62	175	176	212	233	234	235	236	237	238	241	242	246	247	256	258	265	294	317	511	515	516	663	664	665	666	672	944	1921	1922	1923	1924	1925	1926	2029	2297	2298	2300	2645	2650	2652	2653	3719	3720	3722	3736	3806		
20	-	1.01	3.43	1.64	0.40	1.80	0.67	0.97	1.77	1.95	1.87	1.79	2.69	1.89	1.94	0.92	0.64	1.12	0.61	1.08	1.26	1.84	0.51	1.08	0.66	1.17	0.95	1.53	1.44	1.57	2.61	0.97	2.26	1.27	3.45	1.43	1.06	1.52	1.03	2.53	1.79	3.33	0.92	0.98	0.85	0.93	1.77	1.51	1.38	1.92	1.50	0.91	1.50	
40	1.01	-	4.43	0.67	0.62	0.80	0.55	0.27	0.81	1.19	1.14	1.18	2.00	1.07	1.08	1.51	1.02	1.58	0.86	0.53	0.57	0.84	0.69	0.15	0.98	0.32	0.51	0.60	2.23	0.70	2.00	0.73	1.32	0.51	2.75	0.46	0.45	0.51	0.37	1.60	0.91	2.81	0.67	0.39	0.74	0.73	0.87	0.91	0.50	0.98	0.54	1.55	0.61	
50	3.43	4.43	-	5.07	3.83	5.22	3.97	4.34	5.12	5.10	5.04	4.83	5.68	5.14	5.22	3.21	3.59	3.39	3.72	4.34	4.53	5.24	3.83	4.47	3.64	4.54	4.29	4.92	2.59	4.92	5.53	4.13	5.62	4.59	6.40	4.81	4.36	4.93	4.37	5.82	5.08	6.12	4.10	3.30	3.99	4.08	5.09	4.63	4.79	5.25	4.91	3.13	4.86	
60	1.64	0.67	5.07	-	1.24	0.19	1.21	0.86	0.68	1.27	1.19	1.43	1.91	0.99	0.96	2.18	1.69	2.20	1.53	1.05	0.97	0.49	1.29	0.69	1.64	0.70	0.93	0.43	2.88	0.64	1.99	1.26	0.90	0.81	2.53	0.56	0.96	0.36	0.88	1.31	0.91	2.63	1.26	0.98	1.37	1.29	0.78	1.24	0.36	0.81	0.20	2.22	0.63	
61	0.40	0.62	3.83	1.24	-	1.40	0.44	0.63	1.40	1.66	1.59	1.55	2.44	1.57	1.62	1.12	0.69	1.24	0.57	0.81	0.95	1.45	0.36	0.70	0.69	0.82	0.60	1.16	1.76	1.23	2.33	0.79	1.90	0.95	3.15	1.06	0.76	1.13	0.72	2.17	1.46	3.06	0.73	0.67	0.70	0.76	1.43	1.26	1.00	1.56	1.10	1.14	1.15	
62	1.80	0.80	5.22	0.19	1.40	-	1.33	0.96	0.60	1.20	1.14	1.39	1.82	0.93	0.88	2.29	1.80	2.30	1.64	1.11	1.01	0.37	1.43	0.80	1.75	0.78	1.08	0.47	2.99	0.64	1.94	1.34	0.77	0.86	2.48	0.58	1.04	0.38	0.97	1.16	0.83	2.62	1.34	1.07	1.45	1.38	0.72	1.23	0.51	0.71	0.39	2.33	0.64	
175	0.67	0.55	1.21	0.44	1.33	-	0.39	1.16	1.26	1.13	2.11	1.93	1.34	1.00	0.51	1.16	0.36	0.43	0.65	1.28	0.46	0.53	0.43	0.62	0.74	1.01	1.67	1.04	2.17	0.48	1.72	0.69	3.00	0.84	0.43	0.98	0.46	1.93	1.14	3.11	0.33	0.34	0.28	0.42	1.16	0.84	1.00	1.29	1.07	1.02	0.94			
176	0.97	0.27	4.34	0.86	0.63	0.96	0.39	-	0.84	1.06	0.98	1.02	1.85	0.96	1.00	1.38	0.89	1.50	0.73	0.27	0.53	0.93	0.60	0.17	0.80	0.24	0.71	0.65	2.04	0.66	2.09	0.47	1.34	0.35	2.87	0.47	0.19	0.63	0.14	1.66	0.85	3.01	0.41	0.25	0.54	0.47	0.81	0.70	0.65	0.97	0.76	1.40	0.56	
212	1.77	0.81	5.12	0.68	1.40	0.60	1.16	0.84	-	0.68	0.89	0.84	1.64	0.80	0.74	2.01	1.55	1.99	1.41	0.87	0.63	0.25	1.44	0.73	1.51	0.74	1.09	0.67	2.74	0.71	1.58	1.19	0.92	0.75	2.21	0.41	0.86	0.39	0.87	0.83	0.30	2.60	1.12	0.83	1.17	1.20	0.52	0.78	0.84	0.24	0.61	0.27	0.61	
233	1.95	1.19	5.10	1.27	1.66	1.20	1.28	1.06	0.68	-	0.57	0.36	1.24	0.71	0.68	1.96	1.58	2.03	1.48	0.90	0.86	0.89	1.60	1.05	1.49	0.98	1.55	1.04	2.58	0.92	1.84	1.15	1.11	0.87	2.50	0.80	0.94	0.95	1.03	1.09	0.38	3.11	1.08	1.01	1.16	1.15	0.61	0.48	1.30	0.53	1.23	2.00	0.83	
234	1.87	1.14	5.04	1.19	1.59	1.14	1.26	0.98	0.89	0.57	-	0.80	0.88	0.27	0.33	2.06	1.64	2.23	1.54	1.83	0.81	0.99	0.43	0.51	0.84	1.62	0.83	2.56	0.62	2.34	0.95	0.85	0.66	2.98	0.78	0.84	0.98	0.88	1.47	1.66	3.48	1.00	1.22	0.99	0.43	1.11	1.74	1.26	2.07	0.61				
235	1.79	1.18	4.83	1.43	1.55	1.39	1.13	1.02	0.84	0.36	0.80	-	1.50	0.98	0.97	1.66	1.33	1.72	1.25	0.82	0.75	1.09	1.51	1.05	1.26	1.03	1.49	2.22	2.29	1.11	1.77	1.07	1.43	0.91	0.90	1.08	1.02	1.28	0.59	3.11	0.95	0.90	0.95	0.87	0.35	1.43	0.79	1.33	1.71	1.00				
236	2.69	2.00	5.68	1.91	2.44	1.82	2.11	1.85	1.64	1.24	0.88	1.50	-	0.95	0.97	2.83	2.47	3.05	2.38	1.69	1.96	2.70	2.22	1.86	2.31	1.69	2.49	1.57	3.20	1.37	3.00	1.73	1.17	1.51	1.63	1.70	1.76	1.73	1.95	1.43	4.17	1.82	1.94	2.07	1.78	1.20	1.46	1.85	1.43	2.05	2.82	1.42		
237	1.89	1.07	5.14	0.99	1.57	0.93	1.29	0.96	0.80	0.71	0.27	0.98	0.95	-	0.11	2.18	1.73	2.34	1.61	0.88	0.84	1.42	0.93	1.61	0.77	1.55	0.63	2.71	0.44	2.34	1.01	0.61	0.62	2.95	0.69	0.86	0.82	0.86	1.41	0.66	3.38	1.08	1.31	1.06	1.28	0.47	0.93	0.66	1.11	2.19	0.47			
238	1.94	1.08	5.22	0.96	1.62	0.88	1.34	1.00	0.74	0.68	0.33	0.97	0.97	0.11	-	2.23	1.78	2.37	1.65	0.93	1.11	0.76	1.48	0.95	1.66	0.80	1.56	0.63	2.79	0.48	2.47	1.09	0.53	0.67	2.66	2.86	0.68	0.91	0.79	0.91	1.31	0.81	3.11	1.14	1.14	1.24	0.82	0.93	0.58	1.07	2.25	0.48		
241	0.92	1.51	3.21	2.18	1.12	2.29	1.00	1.38	2.01	1.96	2.06	1.66	2.83	2.18	2.23	-	0.49	0.48	0.65	1.32	1.39	2.18	1.19	1.51	0.60	1.61	1.46	2.03	0.88	2.01	2.49	1.27	2.67	1.64	3.39	1.78	1.37	1.92	1.44	2.64	1.94	3.49	1.12	1.23	2.88	1.19	2.07	1.51	2.19	1.98	0.50	1.90		
242	0.64	1.02	3.59	1.69	0.69	1.80	0.51	0.89	1.55	1.58	1.64	1.33	2.47	1.73	1.78	0.45	-	0.70	0.16	0.85	0.95	1.71	0.77	1.02	0.19	1.12	1.03	1.54	1.26	1.53	2.25	0.85	2.19	1.17	3.14	1.30	0.90	1.44	0.96	2.24	2.24	1.51	3.24	0.68	0.75	0.43	0.76	1.61	1.12	1.50	1.69	1.50	0.54	1.42
246	1.12	1.58	3.39	2.20	1.24	2.30	1.16	1.50	1.99	2.03	2.21	1.72	3.05	2.34	2.37	0.48	0.70	-	0.81	1.48	1.36	2.17	1.43	1.60	0.88	1.73	1.39	2.13	2.18	2.15	1.88	2.77	1.21	1.31	1.34	1.31	1.04	2.15	1.63	2.02	2.02	1.51	2.02											
247	0.61	0.86	3.72	1.53	0.57	1.64	0.36	0.73	1.41	1.48	1.54	1.25	2.38	1.61	1.65	0.65	0.16	0.81	-	0.73	0.82	1.56	0.67	0.86	0.21	0.97	0.89	1.40	1.39	2.05	1.15	1.76	1.28	2.12	1.38	3.14	0.57	0.60	0.32	0.67	1.47	1.03	1.34	1.55	1.34	0.70	1.28							
256	1.08	0.53	4.34	1.05	1.11	0.43	0.27	0.87	0.90	0.83	1.69	0.77	0.80	1.61	1.12	1.73	0.97	0.38	0.63	0.78	0.75	0.20	1.02	-	0.83	0.42	2.24	0.43	2.13	0.56	1.11	0.19	2.86	0.32	0.28	0.24	0.18	1.56	0.75	3.04	0.57	0.57	0.30	0.57	0.44	0.76	0.60	0.62	0.72	0.48	0.43	0.66	1.63	0.34
258	1.26	0.57	4.53	0.97	0.95	1.01	0.65	0.53	0.63	0.86	1.09	1.11	1.39	0.95	1.36	0.82	0.55	-	0.82	1.05	0.51	0.94	0.63	0.74	0.90	1.27	0.83	1.05	0.15	2.24	0.45	0.57	0.64	0.63	1.30	0.65	2.65	1.25	0.71	0.34	0.61	0.84	0.88	0.60	0.77	0.77	0.77	0.77	0.77					
265	0.84	1.44	5.09	0.45	1.37	1.24	1.34	0.92	0.11	0.85	1.43	1.16	0.53																																									

Table 36. MSAV12 similarity matrix for BRP/BBG dataset (NL)

cat	20	40	50	60	61	62	175	176	212	233	234	235	236	237	238	241	242	246	247	256	258	265	294	317	511	515	516	663	664	665	666	673	944	1921	1922	1923	1931	1936	2025	2026	2029	2297	2298	2300	2645	2650	2652	2653	3719	3720	3722	3736	3806
20	-	1.54	1.37	2.40	0.18	1.33	0.74	0.59	2.11	1.35	0.95	1.39	1.09	1.84	1.06	0.85	0.71	0.60	0.95	1.57	1.55	2.43	1.07	1.58	1.54	0.62	1.16	1.64	0.91	1.25	1.05	1.27	1.52	2.21	2.03	0.76	0.98	2.30	0.98	2.23	1.25	2.12	0.82	0.55	2.07	0.96	1.25	1.51	1.08	1.21	1.25	0.84	1.48
40	1.54	-	2.86	0.86	1.64	1.01	0.88	1.21	0.58	1.40	1.09	0.63	0.75	0.41	0.79	2.06	1.79	1.41	1.89	1.86	0.71	0.91	0.71	0.60	1.97	1.41	0.93	0.16	0.88	0.77	1.24	1.78	0.97	0.75	1.73	1.01	2.10	0.78	1.11	0.76	0.36	0.94	0.91	1.06	0.54	1.46	0.38	0.83	0.70	0.48	0.40	1.75	0.50
50	1.37	2.86	-	3.72	1.30	2.64	2.09	1.90	3.42	2.45	2.27	2.65	2.45	3.12	2.42	1.52	1.62	1.80	1.83	2.56	2.91	3.71	2.32	2.91	2.42	1.76	2.45	2.98	2.27	2.58	2.32	2.14	2.85	3.46	2.87	2.07	1.54	3.64	2.27	3.58	2.61	3.45	2.13	1.86	3.40	2.07	2.59	2.83	2.41	2.57	2.61	1.79	2.74
60	2.40	0.86	3.72	-	2.50	1.58	1.71	2.02	0.32	2.03	1.82	1.29	1.47	0.71	1.52	2.84	2.57	2.21	2.63	2.40	1.17	0.33	1.48	1.05	2.54	2.20	1.59	0.76	1.64	1.41	1.91	2.43	1.44	0.54	2.17	1.80	2.87	0.14	1.82	0.38	1.16	0.88	1.71	1.90	0.33	2.16	1.18	1.35	1.45	1.24	1.18	2.52	1.11
61	0.18	1.64	1.30	2.50	-	1.35	0.80	0.61	2.20	1.50	1.03	1.54	1.17	1.96	1.12	0.72	0.58	0.72	0.85	1.48	1.61	2.54	1.12	1.63	1.43	0.55	1.17	1.74	0.97	1.28	1.08	1.15	1.55	2.32	2.21	0.78	0.81	2.40	0.99	2.33	1.35	2.16	0.85	0.61	2.17	0.88	1.36	1.64	1.13	1.30	1.36	0.91	1.54
62	1.33	1.01	2.64	1.58	1.35	-	0.71	0.75	1.36	1.41	0.60	1.22	0.45	1.38	0.40	1.39	1.15	1.12	1.12	0.96	0.42	1.78	0.93	0.68	1.15	1.00	0.48	0.98	0.46	0.35	0.42	1.18	0.21	1.70	2.38	0.67	1.49	1.46	0.38	1.32	0.70	1.00	0.71	0.93	1.29	0.78	0.80	0.98	0.53	0.61	0.70	1.33	1.02
75	0.74	0.88	2.09	1.71	0.80	0.71	-	0.35	1.42	1.25	0.58	0.96	0.46	1.25	0.41	1.19	0.92	0.76	1.06	1.28	0.85	1.78	0.47	0.85	1.33	0.57	0.51	0.96	0.29	0.53	0.66	1.10	0.87	1.60	1.97	0.20	1.24	1.61	0.47	1.54	0.55	1.38	0.18	2.5	1.39	0.70	0.62	1.04	0.34	0.52	0.50	1.10	0.82
176	0.59	1.21	1.90	2.02	0.61	0.75	0.35	-	1.74	1.28	0.53	1.22	0.61	1.57	0.55	0.85	0.59	0.61	0.72	1.10	1.02	2.12	0.79	1.10	1.14	0.41	0.64	1.27	0.40	0.71	0.51	0.97	0.95	1.94	2.17	0.30	0.97	1.91	0.39	1.81	0.86	1.60	0.41	3.37	1.69	0.52	0.91	1.20	0.61	0.79	0.87	0.88	1.14
212	2.11	0.58	3.42	0.32	2.20	1.36	1.42	1.74	-	1.87	1.59	1.10	1.23	0.52	1.27	1.57	2.29	1.97	2.37	2.17	0.97	0.43	1.16	0.79	2.29	1.90	1.31	0.50	1.38	1.15	1.68	2.14	1.25	0.43	2.05	1.51	2.57	0.30	1.56	0.46	0.88	0.77	1.41	1.60	0.20	1.88	0.93	1.21	1.16	0.98	0.93	2.29	0.79
233	1.35	1.40	2.45	2.03	1.50	1.41	1.25	1.28	1.87	-	0.89	0.86	1.06	1.48	1.13	1.74	1.70	0.80	1.70	2.18	1.39	2.09	1.59	1.70	2.32	1.66	1.61	1.42	1.16	1.48	1.21	2.24	1.45	2.01	1.34	1.41	2.13	1.90	1.36	1.74	1.26	2.05	1.42	1.35	1.73	1.76	1.08	0.72	1.40	1.15	1.12	0.88	1.75
234	0.95	1.09	2.27	1.82	1.03	0.60	0.58	0.53	1.51	0.89	-	0.97	0.37	1.40	0.38	1.15	1.00	0.55	0.98	1.30	0.78	1.95	1.00	1.07	1.44	0.93	0.81	1.11	0.37	0.71	0.34	1.40	0.73	1.84	0.97	0.65	1.43	1.69	0.54	0.69	0.69	0.67	0.77	1.23									
235	1.39	0.63	2.65	1.29	1.54	1.22	0.96	1.22	1.10	0.86	0.97	-	0.80	0.65	0.88	1.98	1.79	1.11	1.88	2.11	1.01	1.29	1.03	1.13	2.22	1.51	1.25	0.69	0.95	1.09	1.26	2.04	1.22	1.17	1.19	1.15	1.26	1.19	1.24	1.11	0.68	1.48	1.08	1.09	0.99	1.63	0.47	0.54	0.98	0.66	0.57	1.41	1.05
236	1.09	0.75	2.45	1.47	1.17	0.45	0.46	0.61	1.23	1.06	0.37	0.80	-	1.09	0.09	1.40	1.17	0.83	1.21	1.31	0.48	1.61	0.73	0.71	1.45	0.95	0.58	0.75	0.23	0.42	0.50	1.36	0.53	1.55	1.36	0.46	1.22	0.41	1.15	0.55	0.70	1.15	0.91	0.43	0.67	0.39	0.27	0.35	1.12	0.88			
237	1.84	0.41	3.12	0.71	1.98	1.35	1.25	1.57	0.52	1.48	1.40	0.65	1.09	-	1.15	2.41	2.16	1.66	2.26	2.26	1.03	0.65	1.07	0.96	2.37	1.79	1.34	0.42	1.25	1.17	1.57	1.17	1.51	1.40	2.48	0.65	1.49	0.68	0.75	1.16	1.29	1.44	0.83	0.74	1.99	0.83							
238	1.06	0.79	2.42	1.52	1.12	0.40	0.41	0.55	1.27	1.13	0.38	0.88	0.09	1.15	-	1.34	1.14	1.23	0.50	1.66	1.70	0.70	1.37	0.87	0.50	0.86	0.16	0.36	0.45	1.27	0.51	1.54	2.04	0.21	0.48	1.47	1.40	1.38	0.44	1.15	0.49	0.65	1.20	0.83	0.49	0.76	0.34	0.32	0.41	1.12	0.87		
241	0.85	2.06	1.52	2.84	0.72	1.39	0.85	2.57	1.74	1.15	1.98	1.40	2.41	1.34	-	0.34	0.98	0.33	1.17	1.76	2.96	1.58	1.89	1.14	0.83	1.35	2.11	2.0	1.48	1.03	1.09	1.59	2.75	1.09	0.57	2.73	1.05	2.61	1.70	2.34	1.22	1.10	2.52	0.86	1.74	1.89	1.42	1.62	1.70	0.93	1.96		
242	0.71	1.79	1.62	2.57	0.58	1.15	0.92	0.59	2.29	1.70	1.00	1.79	1.17	2.16	1.10	0.34	-	0.92	0.31	0.98	1.52	2.68	1.26	1.59	0.93	0.50	1.04	1.84	0.96	1.19	0.86	0.80	1.36	2.50	2.64	0.79	0.45	2.47	0.80	2.36	1.43	2.05	0.91	0.83	2.25	0.54	1.50	1.74	1.37	1.45	1.00	1.64	
246	0.60	1.41	1.80	2.21	0.72	1.12	0.76	0.61	1.97	0.80	0.55	1.11	0.83	1.66	0.98	0.92	-	0.98	1.61	1.30	2.29	1.20	1.51	1.68	0.94	1.17	0.75	1.15	1.27	1.13	1.11	1.06	1.04	1.07	1.04	1.01	1.05	0.88	0.73	1.53													
247	0.95	1.89	1.83	2.63	0.85	1.12	0.60	0.72	2.37	1.70	0.98	1.88	1.21	2.26	1.14	0.33	0.31	0.98	-	0.87	1.51	2.77	1.45	1.32	1.09	0.93	1.31	2.62	2.77	0.94	0.63	2.52	0.82	2.39	1.53	2.07	1.07	1.04	2.32	0.64	1.59	1.74	1.23	1.44	0.98	1.79							
256	1.57	1.86	2.56	2.40	1.48	0.96	1.28	1.10	2.17	2.18	1.30	2.11	1.31	2.26	1.23	1.17	0.98	1.61	0.87	1.24	2.28	0.80	1.23	0.99	-	0.71	1.49	0.72	0.89	0.85	0.67	1.20	2.08	2.42	0.42	0.70	2.12	0.64	2.05	2.05	1.09	1.74	1.50	0.54	1.40	1.04	1.57	0.77	1.06	1.14	1.19	1.19	
258	1.55	0.71	2.93	1.17	1.61	0.42	0.85	1.02	0.97	1.39	0.78	1.01	0.48	1.03	0.50	1.76	1.52	1.30	1.51	1.34	-	1.38	0.92	0.47	1.53	1.28	0.66	0.63	0.64	0.43	0.77	1.51	0.29	1.34	2.20	0.89	1.86	1.05	0.72	0.89	0.52	0.71	0.86	1.09	0.81	1.61	0.78	0.58	0.46	0.50	1.55	0.85	
265	1.43	0.99	3.71	0.33	2.54	1.78	2.12	0.43	2.09	1.95	1.29	1.61	0.65	1.66	2.96	2.68	2.29	2.77	2.60	1.38	-	1.51	1.22	2.72	2.28	1.73	0.																										

Table 37. VNIROI similarity matrix for BRP/BBG dataset (NL)

cat	20	40	50	60	61	62	175	212	233	234	235	236	237	238	241	242	246	247	256	258	265	294	317	511	515	516	663	664	665	666	672	944	1921	1922	1923	1931	1936	2025	2026	2029	2297	2298	2300	2645	2650	2652	2653	3719	3720	3722	3736	380			
20	-	4.01	1.47	4.23	0.61	4.36	4.99	7.25	8.59	8.60	8.85	10.27	10.19	8.08	8.48	7.37	7.15	5.77	7.63	9.44	12.50	13.12	3.73	9.24	9.78	5.56	6.83	8.35	7.73	9.41	9.34	7.81	13.78	10.50	11.63	8.74	5.57	6.10	10.41	8.24	6.99	7.75	6.57	5.08	8.29	9.21	8.34	11.63	1.29	8.45	3.47	6.45	5.3		
40	4.01	-	5.29	0.59	3.92	1.50	1.26	3.44	4.60	4.87	4.91	6.35	6.20	4.19	4.49	3.75	4.19	2.07	4.01	6.26	8.58	9.19	1.95	5.51	7.12	2.50	3.18	4.36	3.87	5.91	5.79	5.00	9.90	6.70	8.60	4.79	3.25	2.11	4.11	2.75	1.41	4.29	5.64	4.36	7.79	7.22	3.22	4.45	0.54	3.19	1.9				
50	1.47	5.29	-	5.58	1.55	5.67	6.26	8.44	9.83	9.90	10.18	11.46	11.49	9.20	9.78	8.75	8.32	7.14	9.03	10.69	13.44	14.28	4.72	10.43	10.85	6.72	8.08	9.57	9.88	15.15	11.53	12.61	9.98	6.62	7.40	11.76	9.57	8.21	9.07	7.78	6.23	9.54	10.45	9.95	12.95	2.44	9.74	4.77	7.84	6.54					
60	4.23	0.59	5.58	-	4.15	1.22	1.12	3.34	4.42	4.68	4.66	6.29	5.98	4.24	4.27	3.30	3.94	1.62	3.56	5.91	8.28	9.17	2.17	5.33	6.87	2.40	2.94	4.26	3.52	5.66	5.59	4.83	9.57	6.77	8.69	4.63	3.18	1.91	6.19	4.02	2.97	3.76	2.64	1.62	4.14	5.36	4.25	7.56	3.41	4.26	0.87	2.87	2.0		
61	0.61	3.92	1.55	4.15	-	4.15	4.81	7.01	8.45	8.65	8.80	10.21	10.09	7.95	8.37	7.25	6.81	5.66	7.53	9.15	12.39	12.97	3.36	8.97	9.38	5.24	6.59	8.22	7.66	9.12	9.45	7.42	13.70	10.32	11.79	8.57	5.16	6.00	10.32	8.15	6.79	7.55	6.38	4.87	8.14	8.93	8.27	11.65	0.89	8.36	3.39	6.58	5.11		
62	4.36	1.50	5.67	1.22	4.15	-	1.10	3.20	4.59	5.43	5.05	6.81	6.22	4.67	4.52	3.19	3.03	1.74	3.51	5.19	8.34	9.45	1.63	5.01	5.91	1.58	2.54	4.56	3.75	5.12	6.39	3.91	9.68	7.06	9.70	4.68	2.12	2.36	6.35	4.26	2.96	3.41	2.52	1.84	4.32	4.87	4.71	8.08	3.28	4.60	1.55	3.76	1.44		
175	4.99	1.26	6.26	1.12	4.81	1.10	-	2.32	3.67	4.59	4.21	5.78	5.36	3.64	2.79	3.00	1.26	3.10	5.03	7.62	8.43	1.99	4.32	5.92	1.41	1.94	3.56	3.06	4.66	5.60	3.82	9.01	6.00	8.74	3.70	2.38	1.41	5.61	3.44	2.03	2.95	1.61	0.92	3.38	4.41	3.71	7.24	4.00	3.66	1.32	3.29	1.01			
176	7.25	3.44	8.44	3.34	7.01	3.20	2.32	-	1.86	4.08	3.04	4.26	3.65	2.22	2.17	2.40	2.49	2.31	2.66	3.76	5.77	6.38	3.82	2.13	4.81	2.15	1.03	1.90	2.34	2.80	5.17	2.95	7.30	4.08	8.26	1.72	3.05	1.87	4.03	2.20	0.50	1.85	0.71	2.28	1.61	2.60	2.29	3.92	4.13	1.91					
212	8.59	4.60	9.85	4.42	8.45	4.59	3.67	1.86	-	2.83	1.44	2.67	1.82	1.79	0.61	2.46	3.87	3.13	2.46	4.10	4.16	5.00	5.48	2.11	4.21	0.72	1.69	3.04	3.82	4.34	5.64	3.22	7.09	0.52	4.83	2.53	2.33	0.97	1.79	2.21	2.28	3.70	3.21	3.18	2.29	3.76	4.06	4.67	2.83	5.29	1.93	3.09	3.61		
233	8.60	4.87	9.90	4.68	8.65	5.43	4.59	4.08	2.83	-	1.67	2.66	2.81	2.98	2.34	3.58	6.03	3.28	3.40	6.41	5.11	5.93	6.50	4.91	8.22	5.53	4.37	2.62	4.64	5.70	1.10	6.78	6.05	4.62	5.11	3.35	6.60	3.20	3.18	2.29	3.70	4.17	4.06	4.67	2.83	5.29	1.93	3.09	3.61						
234	8.85	4.91	10.18	4.66	8.80	5.05	4.21	3.04	1.44	1.67	-	2.43	1.50	2.53	0.90	2.56	4.83	3.36	2.37	4.86	3.91	5.27	6.17	3.37	6.70	4.76	3.31	1.63	1.44	4.11	2.52	5.48	5.12	3.97	6.38	1.92	5.77	2.85	1.81	0.92	2.78	2.85	3.21	4.37	1.58	3.70	1.26	3.08	8.06	0.84	5.42	3.18	4.44		
235	10.27	6.33	11.46	6.29	10.21	6.81	5.78	4.26	2.67	2.66	2.43	-	2.04	2.52	2.67	4.84	6.51	5.31	4.72	6.53	4.10	3.32	7.55	4.30	8.11	6.29	4.95	2.38	3.74	5.36	3.08	6.89	5.25	2.36	4.90	2.91	7.29	4.48	2.97	3.02	4.15	4.83	6.43	5.58	2.80	5.11	2.11	2.66	9.54	2.42	6.88	4.95	5.77		
236	10.19	6.20	11.49	5.98	10.09	6.23	5.36	3.65	1.82	1.81	1.50	2.04	-	3.05	1.02	1.72	3.55	5.35	4.60	3.36	4.83	2.60	4.05	7.25	3.10	6.64	5.64	4.09	2.16	2.67	3.83	3.37	5.78	3.96	3.39	6.65	2.01	6.55	4.10	4.20	5.45	5.21	3.57	2.10	2.55	6.72	4.68	5.44							
237	8.08	4.19	9.20	4.24	7.95	4.67	3.58	2.22	1.79	2.98	2.53	2.52	3.05	-	2.03	3.69	4.70	3.53	3.77	5.58	5.54	5.00	3.31	6.80	4.05	3.11	1.09	2.85	4.38	3.98	5.08	6.98	2.57	6.17	1.94	5.04	3.19	2.10	3.54	2.48	3.19	1.59	4.17	2.79	1.77	4.74	4.03	4.33							
238	8.48	4.49	9.78	4.27	8.37	4.52	3.64	2.17	0.61	2.34	0.90	2.67	1.72	2.03	-	2.19	4.06	2.93	2.12	4.27	4.14	5.28	5.55	2.87	4.02	2.54	0.98	1.23	3.39	3.32	4.65	5.52	3.63	6.89	1.10	5.00	2.38	2.12	0.43	1.97	2.21	2.44	3.76	3.07	4.17	4.06	4.67	2.83	5.29	1.93	3.09	3.61			
241	7.37	3.75	8.75	3.30	7.25	3.19	2.74	2.40	2.46	3.58	2.56	4.84	3.55	3.69	2.19	-	2.84	1.68	0.37	3.53	5.22	7.35	4.60	3.14	5.09	3.04	1.77	2.87	1.17	3.25	4.48	3.79	5.65	5.65	8.60	2.64	3.86	2.08	3.37	1.81	2.14	1.09	2.23	3.41	2.39	2.80	2.96	5.42	4.45	3.05	3.23				
242	7.15	4.19	8.32	3.94	6.81	3.03	3.00	2.49	2.87	3.67	4.03	6.51	5.35	4.70	4.06	2.84	-	2.98	3.17	2.47	6.81	8.30	3.83	2.94	2.93	1.86	1.75	4.21	6.15	4.21	3.68	2.52	7.06	1.10	8.25	6.29	10.63	6.35	1.87	3.42	5.32	3.92	2.09	2.43	3.43	3.74	2.46	4.77	5.92	3.42	4.45	5.39			
246	5.77	2.07	7.14	1.62	5.66	1.74	1.26	2.31	3.13	3.82	3.36	5.31	4.60	3.53	2.93	1.68	2.98	-	1.96	4.51	6.76	8.08	3.14	4.00	5.80	2.12	1.71	2.04	4.28	4.79	3.98	8.04	8.45	3.33	3.08	1.04	4.69	2.62	1.91	2.21	2.17	1.71	2.01	2.89	3.92	3.21	6.37	4.85	3.02	2.47	2.49				
247	7.63	4.01	9.03	3.56	7.53	3.51	3.10	2.66	2.46	3.40	2.37	4.72	3.36	3.77	2.12	0.37	3.17	1.96	-	3.45	5.09	7.25	4.94	3.25	5.28	3.41	2.10	2.89	1.06	3.37	4.25	4.07	2.67	4.23	2.30	3.11	1.72	2.39	3.13	2.43	2.00	3.24	2.90	5.16	7.11	4.67	3.28	4.00							
256	9.44	6.26	10.69	5.91	5.15	5.19	5.42	4.81	5.67	6.22	6.54	4.05	4.02	3.04	1.86	2.12	3.41	4.21	7.64	8.52	1.99	3.77	4.70	-	1.61	3.93	3.59	3.90	6.59	2.53	9.11	6.11	9.78	8.81	2.41	5.83	3.87	2.16	2.73	1.67	3.63	3.76	4.30	4.15	2.53	0.98									
516	6.83	3.18	8.08	2.94	6.59	2.54	1.94	1.03	2.41	4.37	3.31	4.99	4.05	3.11	2.54	1.77	1.71	2.10	2.64	3.04	7.21	2.74	3.47	4.38	1.61	-	2.63	2.21	2.74	5.43	2.53	7.50	5.06	8.91	2.34	2.49	1.79	4.23	2.40	1.02	1.25	0.84	2.29	2.21	2.47	3.04	6.34	5.74	2.74	3.59	3.92	1.81			
663	8.35	4.36	9.57	4.26	8.22	4.56	3.56	1.90	0.72	2.62	1.63	2.38	2.16	1.09	0.98	2.84	4.71	4.65	4.93	5.33	2.61	6.18	3.93	2.63	-	1.99	3.62	3.64	4.67	6.10	2.89	5.02	4.91	3.06	1.52	3.33	3.73	1.71	2.71	2.26	3.43	1.67	4.21	7.48	8.19	3.76	3.75								
664	7.73	3.87	9.10	3.52	7.66	3.75	3.06	2.34	1.69	2.46	1.44	3.74	2.67	2.85	1.23	1.17	3.68	2.04	1.06	4.09	4.80	6.47	5.03	3.14	3.88	3.59	2.21	1.99	-	3.62	3.41	4.51	6.05	4.82	7.45	2.06	4.57	1.86	2.68	8.80	1.97	3.46	1.64	3.17	1.90	1.75	2.29	3.46	1.37	2.74	2.74	3.41	3.61	2.55	3.58
665	9.41	5.31	9.61	5.12	5.66	9.12	5.12	4.66	2.80	3.04	5.70	4.11	5.36	3.89	3.25	2.52	4.88	3.37	2.43	5.14	4.69	6.38	5.88	1.17	2.78	3.90	2.74	3.62	-	6.62	2.42	6.19	5.02	10.08	2.43	4.31	3.80	4.34	3.14	2.30	3.30	4.90	1.49	6.31	6.10	4.24	2.86								
666	9.34	5.79	10.64	5.95	4.63	5.39	5.60</td																																																

Table 38. NDVI similarity matrix for CDL dataset (CA), continued on following page

cat	1	2	3	4	6	12	21	22	23	24	27	28	29	31	33	36	37	41	42	43	44	46	47	48	49	53	54	55	57	58	59	61	66	67	68	69	71	72	74	75	76	77	92		
1	-	1.74	2.16	0.66	1.56	1.39	3.04	3.63	2.33	2.59	2.38	2.03	3.42	3.22	2.52	2.17	2.06	2.78	2.22	1.13	1.31	0.73	1.71	1.36	2.83	1.49	1.49	2.10	2.95	3.07	1.73	2.71	1.83	1.80	2.39	1.46	1.51	1.83	1.44	1.60	1.68	2.46	1.87		
2	1.74	-	0.91	1.33	2.12	2.03	3.41	2.48	3.53	3.33	3.55	3.12	2.83	3.82	2.93	3.34	4.42	2.39	2.75	1.55	1.40	1.96	0.93	3.11	1.86	1.41	2.14	3.18	4.30	3.03	2.91	2.85	2.42	3.35	2.22	2.25	2.77	2.27	2.61	2.53	3.35	2.66			
3	2.16	0.91	-	1.92	2.60	2.58	4.08	4.64	4.06	3.92	3.94	3.74	3.71	3.33	4.32	3.42	3.69	4.68	2.89	2.34	2.22	1.96	2.55	1.58	3.53	2.54	1.92	2.77	3.77	4.65	3.43	3.39	3.25	2.80	3.64	2.90	2.94	3.48	2.92	3.09	2.93	3.60	3.18		
4	0.66	1.33	1.92	-	1.27	1.14	2.65	3.55	2.48	2.42	2.58	2.06	2.87	3.00	2.24	2.34	3.33	1.87	0.98	0.96	0.64	1.66	0.80	2.56	1.00	1.01	1.81	2.56	2.56	2.22	2.28	2.22	2.04	2.85	1.39	1.41	1.87	1.43	1.84	1.99	2.91	1.79			
6	1.56	2.12	2.60	1.27	-	0.56	2.39	3.24	3.04	2.40	3.00	2.50	2.40	2.69	1.37	3.37	2.91	3.66	1.70	1.30	1.65	1.84	2.25	1.53	1.58	0.99	0.87	1.70	2.07	4.26	2.96	1.94	2.82	2.60	3.54	1.75	1.76	2.36	1.85	2.29	2.56	3.60	1.96		
12	1.39	2.03	2.58	1.14	0.56	-	2.38	3.13	2.85	2.26	2.82	2.30	2.40	2.68	1.41	3.06	2.65	3.52	1.61	1.16	1.34	1.69	1.92	1.46	1.63	0.82	0.85	1.49	2.03	3.96	2.67	1.87	2.47	2.21	3.20	1.37	1.37	2.00	1.44	1.88	2.20	3.27	1.74		
21	3.04	3.41	4.08	2.65	2.39	2.38	-	2.41	2.57	1.20	2.75	2.01	2.30	1.34	1.92	4.18	2.93	4.40	1.65	2.70	2.25	3.01	3.15	2.83	2.35	1.79	2.54	2.59	0.99	4.80	3.87	0.77	3.83	3.75	4.59	2.37	2.26	2.54	2.48	3.09	3.66	4.74	1.94		
22	3.63	4.28	4.64	3.55	3.24	3.13	2.41	-	2.92	1.68	2.31	2.56	3.34	1.28	2.26	4.11	2.99	4.07	2.03	3.38	2.98	3.93	3.56	3.64	2.52	2.95	3.40	3.16	1.96	4.61	4.01	2.36	3.81	3.73	4.45	2.93	2.86	3.20	2.94	3.03	3.64	4.61	1.95		
23	2.33	3.53	4.06	2.48	3.04	2.85	2.57	2.92	-	1.68	0.97	4.11	2.45	3.16	2.40	0.94	2.42	2.62	2.62	2.26	2.44	2.77	2.91	3.65	2.49	3.15	3.15	2.94	2.94	2.66	2.19	2.75	2.57	2.93	3.05	2.13	2.07	1.78	2.08	2.20	2.66	3.25	1.73		
24	2.59	3.33	3.92	2.42	2.40	2.26	1.20	1.68	-	1.62	1.14	2.88	0.97	1.90	3.32	1.97	3.44	1.41	2.39	1.90	2.71	2.69	2.58	2.39	1.80	2.54	2.43	1.35	3.88	3.08	1.33	3.06	3.09	3.77	1.87	1.77	1.96	1.91	2.31	2.94	3.95	1.06			
27	2.38	3.55	3.94	2.54	3.00	2.82	2.75	2.31	0.97	1.62	-	1.10	4.14	2.20	2.91	2.30	1.01	2.29	2.36	2.57	2.26	2.66	2.65	2.98	3.36	2.57	3.11	2.99	2.83	2.65	2.20	2.80	2.40	2.68	2.88	2.12	2.07	2.00	2.05	1.95	2.43	3.07	1.38		
28	2.03	3.12	3.71	2.06	2.50	2.30	2.01	2.56	0.71	1.14	1.10	-	3.55	2.03	2.53	2.46	1.02	2.66	2.00	2.13	1.74	2.13	2.36	2.43	3.02	1.90	2.55	2.30	2.93	2.18	2.09	2.39	2.63	3.01	1.58	1.51	1.35	1.56	1.86	2.40	3.20	1.19			
29	3.42	2.83	3.33	2.87	2.40	2.40	2.30	3.34	2.21	2.88	4.14	3.55	-	2.92	2.00	4.97	4.26	5.59	1.85	2.94	2.48	3.29	3.32	3.29	1.98	2.10	2.06	2.46	1.71	5.85	4.65	1.71	4.32	3.90	5.07	2.92	2.87	3.50	3.03	3.61	3.99	5.16	2.89		
31	3.22	3.82	4.32	2.00	2.69	2.68	1.84	1.34	2.12	2.48	2.45	0.97	2.20	2.03	2.92	-	1.92	4.04	2.75	4.03	1.66	2.92	2.56	3.41	3.20	3.11	2.34	2.20	2.93	2.82	1.31	4.59	3.84	1.57	3.76	2.71	4.48	2.60	2.50	2.82	2.65	3.00	3.59	4.63	1.65
33	2.52	2.93	3.42	2.24	1.37	1.41	1.92	2.26	3.16	1.90	2.91	2.53	2.00	1.92	-	3.78	3.03	3.99	1.18	1.92	2.02	2.80	2.55	2.32	0.53	1.51	1.71	1.62	1.17	4.58	3.47	1.37	3.18	2.94	3.95	1.93	1.90	2.52	2.04	2.46	2.90	4.06	1.70		
36	2.17	3.34	3.69	2.64	3.37	3.06	4.18	4.11	2.40	3.32	2.30	2.46	4.97	4.04	3.78	-	1.58	2.03	2.51	1.98	3.24	4.09	3.16	3.40	2.96	4.12	1.02	0.59	3.95	0.85	1.50	0.74	2.14	2.20	1.90	2.00	1.47	1.22	0.92	2.53					
37	2.06	3.34	3.85	2.34	2.91	2.65	1.93	2.99	0.94	1.97	1.01	1.02	4.26	2.75	3.03	1.58	-	1.96	2.60	2.21	2.10	2.30	2.17	2.86	3.47	2.48	3.03	2.71	3.06	1.94	1.41	2.91	1.70	2.18	2.19	1.68	1.67	1.29	1.59	1.48	1.87	2.40	1.55		
41	2.78	4.42	4.68	3.33	3.66	3.52	4.04	4.07	2.42	3.46	2.29	2.66	5.59	4.03	3.99	2.03	1.96	-	3.95	3.19	3.53	3.27	3.42	4.00	4.33	3.71	3.99	3.42	4.43	1.96	1.89	4.36	4.24	3.04	2.47	3.06	3.10	2.83	2.97	2.59	2.70	2.63	3.02		
42	2.22	2.39	2.89	1.87	1.70	1.61	1.65	2.03	2.62	1.41	2.36	2.00	1.85	1.66	1.18	3.47	2.60	3.95	-	1.83	2.31	2.17	2.15	1.56	1.22	1.54	1.65	1.05	4.27	3.20	1.11	2.96	2.69	3.70	1.65	1.59	2.23	1.74	2.19	2.67	3.82	1.14			
43	1.13	1.75	2.34	0.98	1.30	1.16	1.20	2.70	3.38	2.62	2.39	2.57	2.13	2.94	2.92	1.92	1.28	2.38	2.21	3.19	1.83	-	1.20	1.40	1.08	1.46	2.21	1.33	1.07	1.06	1.47	1.54	2.57	1.59											
44	1.31	1.55	2.22	0.96	1.65	1.34	2.25	2.06	2.98	2.26	1.90	2.26	1.74	2.48	2.56	2.02	1.60	2.30	2.10	3.53	1.31	1.20	-	1.24	1.38	1.02	2.37	1.55	1.50	2.06	3.50	2.31	1.80	2.12	1.86	2.83	0.97	0.94	1.52	1.85	2.93	1.26			
46	0.73	1.40	1.96	1.84	1.69	1.01	3.01	3.93	2.44	2.71	2.66	2.13	3.29	3.41	2.80	2.51	2.31	3.27	2.31	1.40	1.24	1.86	1.02	3.13	1.50	1.51	2.26	3.02	3.37	2.06	2.70	2.20	2.14	2.73	1.66	1.94	2.06	2.80	2.13						
47	1.71	1.96	2.55	1.66	2.25	1.92	1.35	2.56	2.77	2.69	2.65	2.33	3.02	3.30	2.55	2.57	1.98	2.17	2.44	2.21	1.08	1.38	1.86	-	1.99	2.79	1.92	2.06	1.34	1.05	1.16	1.01	1.24	1.46	1.05	1.16	1.01	2.03	1.79						
48	1.36	0.93	1.58	0.80	1.53	1.46	2.63	2.64	2.91	2.58	2.98	2.43	2.39	3.11	2.32	2.84	4.00	1.75	1.46	1.02	1.99	-	2.59	1.11	1.94	2.47	4.15	2.84	2.19	2.76	2.47	3.38	1.78	2.32	1.85	3.45	2.07								
49	2.83	3.11	3.53	2.54	1.58	1.63	2.35	2.52	3.65	2.39	3.36	3.02	1.94	2.53	2.04	0.53	4.09	3.47	4.33	4.56	2.11	2.37	3.13	2.79	2.59	1.88	-	1.88	1.78	1.53	4.92	3.79	1.76	3.43	2.37	2.63	1.24	1.88	2.34	2.14	2.34				
53	1.49	1.86	2.50	1.54	1.00	0.99	0.82	1.79	2.59	2.49	2.08	1.57	1.90	2.10	2.30	1.51	2.07	1.59	1.65	1.01	1.26	1.22	1.24	1.22	1.21	1.20	1.21	1.22	1.23	1.24	1.25	1.26	1.27	1.28	1.29	1.25	1.26	1.27							
54	1.49	1.41	1.92	1.01	0.87	0.85	2.54	3.11	2.45	2.06	2.60	2.93	1.71	2.30	2.04	0.26	3.03	3.40	1.91	1.51	1.05	1.26	1.04	1.89	0.89	0.89	1.24	2.16	4.35	3.01	2.77	2.45	3.03	2.70	2.28	2.50	3.56	2.02							
55	2.10	2.14																																											

3.86	5.80	1.85	1.89	2.54	3.86	4.41	2.41	2.47	2.58	1.77	2.59	1.52	1.85	2.29	2.50	1.56	1.31	3.52	2.02	1.69	1.82	1.99	2.83	1.31	2.02	1.82	2.91	1.50	1.54	1.68	1.40	3.14	2.24	2.44	3.31	1.38	2.88	1.99	2.15	2.08	1.88	2.68	2.27	
3.58	5.19	2.99	2.55	2.68	3.58	3.84	3.77	3.57	3.90	2.41	3.63	2.50	2.34	2.35	3.59	0.92	1.59	3.85	1.76	2.75	2.81	1.50	3.11	1.66	2.17	2.06	3.16	2.34	2.28	2.29	2.18	4.41	3.39	3.55	3.34	1.22	3.26	2.99	3.08	3.30	1.93	4.07	3.27	
4.21	5.67	3.64	3.25	3.36	4.16	4.39	4.26	3.99	4.33	3.13	4.32	3.93	3.04	4.08	1.33	2.20	4.25	2.28	3.37	3.50	2.00	3.83	2.25	2.90	2.59	3.92	2.86	2.80	2.92	2.79	4.98	3.72	3.91	3.99	1.85	3.54	3.46	3.44	3.72	2.09	4.44	3.96		
3.33	5.23	1.96	1.70	2.12	3.34	3.83	2.79	2.91	3.04	1.56	2.54	1.64	1.64	1.88	2.56	1.11	1.28	3.28	1.50	1.88	1.89	1.44	2.58	0.80	1.64	1.68	2.62	1.63	1.71	1.44	1.40	3.30	2.56	2.76	2.85	1.12	1.93	2.09	2.34	2.32	1.95	3.15	2.29	
2.84	4.64	2.37	1.80	1.81	2.77	3.28	3.29	3.64	3.65	1.70	2.66	2.06	1.56	1.66	2.75	1.42	2.09	2.69	1.05	2.47	2.36	1.48	2.75	0.87	1.68	1.81	2.75	2.09	2.40	1.98	1.93	3.43	3.46	3.77	2.51	1.92	3.68	2.93	3.20	3.11	3.09	3.74	2.67	
2.81	4.71	2.06	1.48	1.62	2.77	3.31	2.99	3.31	3.35	1.35	2.50	1.70	1.29	1.41	2.58	1.39	1.77	2.64	1.07	2.11	2.00	1.46	2.37	0.82	1.34	1.41	2.35	1.71	2.04	1.82	1.51	3.24	3.29	3.57	2.40	1.65	3.54	2.80	3.05	2.98	2.93	3.51	2.36	
1.73	3.67	2.42	1.76	1.30	1.85	2.59	3.56	4.45	4.20	1.90	1.74	2.66	1.96	1.68	2.12	2.94	3.22	2.10	2.30	2.93	2.57	2.31	2.04	2.13	1.86	2.80	2.01	2.91	3.20	1.80	2.71	2.37	3.57	3.96	1.98	1.08	2.85	3.43	2.78	3.15	3.12	3.97	4.16	2.17
3.07	4.81	2.88	2.47	2.38	3.06	3.90	3.49	4.60	4.16	2.69	2.65	2.85	2.88	2.56	1.86	3.64	3.82	1.25	3.28	3.35	3.18	2.73	2.98	2.91	3.02	3.40	2.08	2.58	2.60	3.81	4.39	2.52	3.41	3.21	3.47	3.33	3.67	4.47	3.94	2.80	3.21			
4.06	6.12	1.36	2.01	2.83	4.16	4.92	1.67	2.74	2.26	2.12	1.34	1.94	2.38	2.86	1.08	3.35	2.74	3.45	3.40	1.82	1.76	3.24	2.58	2.65	2.63	2.74	2.25	2.30	1.55	2.23	1.40	1.28	1.73	3.24	2.27	2.70	2.18	1.01	1.15	1.06	2.92	1.99	1.27	
2.56	4.61	1.66	1.30	1.51	2.64	3.47	2.63	3.68	3.32	1.53	1.18	1.93	1.50	1.73	0.96	2.85	2.82	1.90	2.51	2.25	1.97	2.45	1.80	2.11	1.75	2.39	1.94	2.23	2.52	1.24	2.34	1.78	2.74	3.22	1.80	2.49	2.58	2.15	2.29	2.43	3.49	3.18	1.46	
4.08	6.14	1.54	2.02	2.79	4.14	4.94	1.70	2.82	2.29	2.16	1.77	1.79	2.15	2.79	0.75	3.25	2.74	3.00	3.36	1.93	1.97	3.21	2.61	2.72	2.60	2.68	2.86	2.03	2.16	1.61	2.15	1.50	1.61	2.21	3.27	2.60	1.89	1.58	1.26	1.67	2.95	1.91	1.64	
3.42	5.50	0.95	1.35	2.14	3.51	4.27	1.76	2.78	2.40	1.49	0.91	1.56	1.76	2.19	0.85	2.84	2.85	2.89	2.80	1.53	1.35	2.67	2.07	2.12	1.99	2.40	2.20	1.85	1.98	1.01	1.77	1.41	1.73	2.13	2.66	2.24	2.16	1.27	1.45	2.88	2.26	0.82		
1.16	2.52	3.62	2.60	1.62	1.01	1.18	4.77	5.23	2.62	1.57	3.37	3.34	2.36	1.68	3.69	2.27	3.34	2.34	1.52	3.81	3.55	1.60	2.76	2.21	2.05	2.66	3.37	3.66	3.14	4.88	4.81	5.13	1.75	2.79	4.25	4.16	4.37	4.49	4.30	5.37	3.56			
2.49	4.34	2.51	2.03	1.84	2.52	3.35	3.39	4.46	4.07	2.23	1.98	2.65	1.99	2.12	1.59	3.25	3.48	1.51	2.79	3.06	2.81	2.80	2.36	2.58	2.30	2.87	2.53	2.88	3.22	1.96	2.84	2.40	3.47	4.03	1.76	3.09	3.14	2.86	3.16	4.06	3.87	2.40		
2.14	4.02	2.24	1.61	1.19	2.02	2.74	3.43	4.10	3.95	1.65	2.48	2.18	1.24	1.26	2.45	2.16	2.68	1.44	1.53	2.73	2.50	1.90	2.37	1.71	1.54	1.85	2.42	2.28	2.70	2.07	2.18	3.12	3.90	4.32	1.92	2.35	3.62	3.38	3.51	3.62	3.86	3.98	2.60	
5.17	7.30	1.89	2.64	3.68	5.24	5.89	1.21	0.70	0.81	2.57	2.99	1.71	2.77	3.43	2.74	3.40	1.40	1.85	3.82	3.35	3.39	1.66	1.40	2.60	1.87	2.90	2.16	2.22	4.46	2.46	2.83	2.51	2.21	2.50	2.43	1.19	2.42							
4.21	6.36	0.87	1.79	2.80	4.30	5.06	0.84	1.94	1.44	1.84	1.58	1.34	2.09	2.73	1.30	3.18	2.23	3.44	3.35	1.13	1.24	3.26	2.44	2.65	2.43	2.44	2.64	1.60	1.59	1.68	1.48	1.47	1.82	3.45	3.24	2.12	1.56	2.72	1.28	1.18				
5.68	7.76	2.47	3.31	4.26	5.70	6.46	1.69	2.31	1.69	3.32	3.00	2.67	3.45	4.18	2.68	3.46	3.37	4.56	4.42	2.52	2.76	4.45	4.18	3.72	3.92	3.64	3.44	3.21	2.99	2.66	2.31	2.47	3.49	3.64	3.84	2.81	2.64	2.47	3.69	3.11	2.85			
2.07	4.05	2.15	1.29	0.95	2.04	2.76	3.19	3.86	3.73	1.41	2.26	1.88	0.92	0.97	1.49	1.76	2.31	1.44	2.25	1.39	1.93	1.42	1.17	1.64	2.06	1.98	2.30	1.33	1.87	2.93	2.34	2.09	2.71	2.74	3.03	3.17	2.21							
3.31	5.30	1.70	1.42	1.97	3.32	3.86	2.50	2.61	2.74	1.23	2.45	1.21	1.25	1.68	2.51	1.48	0.99	3.04	1.74	1.54	1.58	1.86	2.37	1.33	1.42	1.21	2.18	1.7	1.30	1.50	1.07	3.12	2.88	3.14	2.86	2.94	2.49	2.58	2.75	2.31	2.91	2.21		
2.83	4.84	1.65	1.11	1.54	2.89	3.45	2.62	2.93	3.00	1.06	2.24	1.27	1.07	2.16	1.31	2.3	1.23	2.75	1.50	1.63	1.56	1.91	1.75	1.09	0.98	1.21	1.88	1.28	1.42	1.00	1.04	2.98	2.56	2.05	2.31	2.07	2.10	2.33	1.84					
3.76	5.63	2.06	2.02	2.56	3.79	4.26	2.74	2.72	2.91	1.90	2.69	1.84	2.08	2.33	2.74	1.48	1.37	3.76	2.00	1.90	1.81	2.85	1.85	1.78	1.67	1.65	2.36	3.26	3.28	1.32	1.92	2.17	2.07	2.04	3.03	2.33								
3.64	5.67	1.75	1.60	2.29	3.70	4.22	2.42	2.33	2.18	2.46	1.43	2.75	1.05	0.59	0.35	2.39	1.29	2.04	1.05	2.01	1.79	1.46	2.40	2.08	1.09	1.53	1.00	1.51	1.77	1.27	2.25	2.45	2.25	2.25	2.25	2.25								
3.01	4.75	2.45	1.95	2.04	3.01	3.48	3.50	3.66	3.85	1.89	2.90	2.11	1.82	1.81	2.90	0.70	2.24	2.26	1.24	2.76	2.76	1.77	2.06	2.07	1.77	1.70	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67						
1.37	3.49	2.27	1.36	0.57	1.45	2.22	3.46	4.21	4.06	1.47	1.95	2.29	1.42	0.97	2.23	2.38	2.76	1.77	1.70	1.25	1.21	1.73	1.70	1.25	2.19	1.70	2.49	2.81	1.64	2.27	2.90	1.63	4.00	0.91	2.35	3.33	2.92	3.16	3.73	2.09				
4.59	6.69	1.66	1.24	3.11	4.64	5.26	1.48	1.11	1.36	2.03	2.84	1.19	2.18	2.81	2.82	1.40	1.91	3.26	2.75	2.04	1.47	1.77	2.37	2.04	1.47	2.27	2.50	2.04	2.47	2.76	2.45	2.25	2.25	2.25	2.25	2.25	2.25	2.25	2.25	2.25	2.25			
4.27	6.30	1.96	2.07	2.84	4.31	4.87	2.12	1.77	2.10	1.93	3.11	1.24	1.95	2.49	2.88	1.40	1.77	2.36	2.85	1.40	1.07	3.67	2.85	2.03	2.17	2.71	2.07	1.76	2.04	1.66	2.04	2.64	2.42	2.55	2.50	2.08	2.11							
5.37	7.46	2.33	2.09	2.94	4.35	4.95	1.78	1.56	1.77	1.84	2.84	1.90	2.52	2.66	2.48	1.11	3.65	2.91	3.45	1.05	2.07	2.75	2.43	1.75	2.05	2.87	2.10	1.50	2.04	1.51														

Table 39. NDWI similarity matrix for CDL dataset (CA), continued on following page

cat	1	2	3	4	5	6	12	21	22	23	24	27	28	29	31	33	36	37	41	42	43	44	46	47	48	49	53	54	55	57	58	59	61	66	67	68	69	71	72	74	75	76	77	92
1	-	1.20	1.76	0.34	1.03	0.69	1.61	2.43	1.90	1.63	1.84	1.50	1.55	1.92	1.29	1.65	1.76	2.98	1.09	1.62	0.95	0.66	1.51	0.47	1.55	0.87	0.49	1.78	1.62	2.69	1.34	1.47	1.70	1.46	2.04	1.16	1.27	1.68	1.28	1.57	1.49	1.99	1.46	
2	1.20	-	1.31	1.36	1.87	1.70	2.45	3.19	2.80	2.52	2.67	2.43	1.93	2.65	2.16	2.32	2.55	3.88	1.84	1.99	1.78	1.83	0.95	0.23	1.78	1.17	2.28	2.47	3.20	2.03	2.28	2.20	1.89	2.37	1.94	2.04	2.44	2.06	2.21	1.99	2.20	2.31		
3	1.76	1.31	-	1.98	1.29	2.59	2.20	3.12	3.49	3.01	3.06	2.95	2.76	2.97	3.30	2.79	2.22	2.57	3.55	2.49	3.09	2.11	1.42	2.75	1.83	2.80	2.53	1.96	3.24	3.10	2.67	1.89	2.99	2.06	1.98	1.91	2.28	2.37	2.48	2.37	2.27	2.04	1.80	3.07
4	0.34	1.36	1.98	-	0.94	0.54	1.31	2.35	1.75	1.41	1.70	1.31	1.33	1.74	1.15	1.60	1.67	3.04	0.93	1.40	0.74	0.77	1.28	0.48	1.53	0.56	0.60	1.56	1.35	2.72	1.35	1.16	1.71	1.42	2.13	0.97	1.03	1.49	1.08	1.50	1.46	2.08	1.21	
6	1.03	1.87	2.59	0.94	-	0.73	1.59	2.43	2.31	1.70	2.22	1.86	1.19	1.84	0.72	2.12	2.23	3.27	1.23	1.29	1.52	1.56	1.35	1.16	1.09	0.87	0.73	1.26	1.49	3.22	1.85	1.36	2.06	1.73	2.51	1.39	1.46	2.01	1.48	1.87	1.77	2.43	1.20	
12	0.69	1.70	2.20	0.54	0.73	-	1.29	2.04	1.73	1.28	1.64	1.29	1.33	1.55	0.72	1.41	1.54	2.80	0.80	1.53	0.87	0.95	1.25	0.92	1.07	0.68	0.73	1.42	1.16	2.54	1.16	1.10	1.44	1.15	1.92	0.73	0.82	1.32	0.82	1.21	1.17	1.87	1.06	
21	1.61	2.45	3.12	1.31	1.59	1.29	1.94	1.27	0.63	1.23	0.88	1.28	1.02	1.32	1.79	1.57	3.16	0.98	1.37	1.08	1.78	1.21	1.56	1.86	0.88	1.75	1.28	0.46	2.94	1.87	0.32	2.13	1.89	2.70	2.15	1.15	1.09	1.43	1.13	1.71	1.89	2.71	0.83	
22	2.43	3.19	3.49	2.35	2.43	2.04	1.94	-	1.64	1.34	1.31	1.65	2.42	1.11	1.77	1.59	1.43	2.08	1.54	2.81	2.08	2.36	2.25	2.55	1.65	2.27	2.50	2.21	1.58	2.12	1.81	1.99	1.72	1.88	2.12	1.70	1.89	1.93	1.71	1.46	1.71	2.22	1.72	
23	1.90	2.80	3.01	1.75	2.31	1.73	1.27	1.44	-	0.94	0.42	0.49	2.35	1.44	1.96	1.32	0.78	2.21	1.44	2.51	1.19	1.72	2.18	2.03	2.22	1.69	2.23	2.36	1.31	2.08	1.54	1.48	1.89	1.98	2.31	1.39	1.43	1.25	1.36	1.54	1.86	2.45	1.55	
24	1.63	2.52	3.06	1.41	1.70	1.28	0.63	1.34	0.94	-	0.75	0.63	1.54	0.62	1.24	1.43	1.15	2.64	0.77	1.77	1.10	1.70	1.42	1.67	1.60	1.15	1.78	1.46	0.47	2.50	1.58	0.77	1.79	1.66	2.34	1.01	1.08	1.32	1.01	1.37	1.61	2.39	0.69	
27	1.84	2.67	2.95	1.70	2.22	1.64	1.31	1.34	0.42	0.75	-	0.54	2.17	1.13	1.79	1.15	0.61	2.21	1.20	2.40	1.18	1.67	1.94	1.90	2.06	1.65	2.13	2.18	1.16	1.97	1.42	1.40	1.67	1.75	2.11	1.23	1.33	1.21	1.31	1.63	2.23	1.40		
28	1.50	2.43	2.76	1.31	1.86	1.29	0.88	1.65	0.49	0.63	0.54	-	1.90	1.22	1.55	1.18	0.80	2.43	1.04	2.06	0.77	1.40	1.73	1.61	1.89	1.22	1.30	1.93	0.93	2.19	1.32	1.03	1.68	1.66	2.17	0.98	1.00	1.00	0.96	1.30	1.58	2.26	1.15	
29	1.55	1.93	2.97	1.33	1.19	1.33	1.28	2.42	2.35	1.54	2.17	1.90	-	1.46	1.16	2.32	2.27	3.86	1.12	1.51	1.64	1.96	0.48	1.30	1.67	1.01	1.31	0.44	1.23	3.52	2.24	1.04	2.33	1.90	2.86	1.50	1.53	2.15	2.05	1.99	2.74	0.98		
31	1.92	2.65	3.30	1.74	1.84	1.55	1.02	1.11	1.44	0.62	1.13	1.22	1.46	-	1.26	1.69	1.51	2.85	0.86	1.81	1.56	2.04	1.25	1.90	1.52	1.47	1.94	1.28	0.72	2.70	1.85	1.07	1.90	1.77	2.45	1.28	1.42	1.72	1.31	1.51	1.71	2.46	0.79	
33	1.29	2.16	2.79	1.15	0.72	0.72	1.32	1.77	1.96	1.24	1.79	1.55	1.16	1.26	-	1.73	1.82	2.90	0.84	1.47	1.42	1.61	1.18	1.41	0.63	0.98	1.11	1.04	1.03	2.83	1.58	1.12	1.69	1.41	2.21	1.05	1.17	1.69	1.11	1.42	1.40	2.16	0.83	
36	1.65	2.32	2.22	1.60	2.12	1.41	1.79	1.59	1.32	1.43	1.15	0.68	1.23	1.22	1.64	1.73	-	0.57	2.11	1.32	2.63	1.10	1.21	1.99	1.83	1.77	1.80	1.92	2.33	1.59	1.76	1.63	0.86	1.05	0.87	1.01	0.77	0.86	0.45	0.74	1.17	1.83		
37	1.76	2.55	2.57	1.67	2.23	1.54	1.57	1.43	0.78	1.15	0.66	0.80	2.37	1.32	1.82	1.82	0.57	-	1.95	1.32	2.63	1.08	1.29	1.99	1.92	1.77	2.08	2.37	1.44	1.40	1.89	1.89	1.64	1.16	1.37	1.55	1.04	1.58	1.71	1.70				
41	2.98	3.86	3.55	3.04	3.27	2.80	2.16	2.08	2.21	2.64	2.21	2.43	3.86	2.85	2.90	2.11	1.95	-	2.80	4.12	2.77	2.71	3.74	3.34	2.64	2.23	3.23	3.77	2.96	1.74	2.16	2.30	2.73	2.27	2.67	2.82	2.55	2.67	2.30	2.51	2.55	3.07		
42	1.09	1.84	2.49	0.93	1.23	0.80	0.98	1.54	1.44	0.77	1.20	1.04	1.12	0.86	0.84	1.32	1.32	2.80	-	1.45	0.93	1.25	0.96	1.08	1.18	0.86	1.13	1.15	0.78	1.45	1.20	1.98	0.71	0.71	1.38	0.81	1.12	1.18	1.94	0.70	0.70			
43	1.62	1.99	3.09	1.40	1.29	1.41	1.53	1.81	2.81	2.51	1.77	2.40	2.06	0.51	0.81	1.47	2.63	2.63	4.12	1.45	-	1.78	2.10	0.94	1.32	2.04	1.41	1.48	3.84	2.52	1.18	2.38	1.86	2.41	2.37	3.10	1.19	1.19	1.19	1.19	1.19	1.19		
44	0.95	1.78	2.11	0.74	1.52	0.87	1.08	2.08	1.19	1.10	1.18	0.77	1.64	1.42	1.10	1.08	2.77	0.93	1.78	-	0.78	1.42	1.03	1.78	0.89	1.29	1.80	1.14	2.24	1.04	1.04	1.44	1.26	1.90	0.65	0.64	0.88	1.69	1.15	1.26	1.91	1.27		
46	0.66	1.33	1.42	0.77	1.56	0.95	1.78	2.36	1.72	1.70	1.67	1.40	1.96	2.04	1.61	1.21	1.39	2.71	1.25	2.10	0.78	-	1.77	0.89	1.78	1.27	1.07	2.16	1.76	2.14	0.89	1.68	1.30	1.18	1.57	1.03	1.14	1.33	1.13	1.25	1.16	1.76		
47	1.51	1.83	2.75	1.28	1.35	1.25	1.21	2.25	2.18	1.42	1.97	1.73	0.48	1.35	1.18	1.99	2.07	3.11	0.96	0.90	1.42	1.77	-	1.25	1.65	1.05	1.66	2.00	1.96	2.00	1.27	1.71	1.66	2.03	1.03	1.27	1.27	1.27	1.27	1.27	1.27	1.27		
48	0.47	0.95	1.83	0.48	1.66	0.92	1.56	2.55	2.03	1.67	1.94	1.61	1.30	1.90	1.41	1.83	1.93	3.34	1.08	1.34	1.03	0.85	1.25	-	1.75	0.85	1.61	1.56	1.62	2.92	1.59	1.41	1.86	1.54	2.15	1.44	1.44	1.44	1.44	1.44	1.44			
49	1.55	2.33	2.80	1.53	1.09	1.07	1.86	1.65	2.22	1.60	2.00	1.89	1.67	1.53	0.63	1.77	1.94	2.64	1.16	1.24	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07					
53	0.87	1.78	2.53	0.56	0.87	0.68	0.88	2.27	1.69	1.15	1.65	1.25	1.22	1.01	1.47	0.98	1.77	1.72	1.77	1.72	1.05	1.65	1.75	-	1.53	0.92	1.19	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09
54	0.49	1.17	1.96	0.60	0.73	0.73	1.73	2.50	2.23	1.78	2.13	1.81	1.31	1.94	1.11	1.92	1.22	2.00	1.22	1.06	2.3																							

Table 40. CI similarity matrix for CDL dataset (CA), continued on following page

cat	1	2	3	4	6	12	21	22	23	24	27	28	29	31	33	36	37	41	42	43	44	46	47	48	49	53	54	55	57	58	59	61	66	67	68	69	71	72	74	75	76	77	92		
1	-	0.62	0.84	0.45	0.31	0.29	0.46	0.58	1.00	0.46	0.86	0.69	0.95	1.31	0.72	0.42	0.81	0.60	0.36	1.06	0.41	0.70	1.07	0.33	0.37	0.48	0.24	0.80	0.38	1.11	0.37	0.30	0.59	0.57	0.72	0.39	0.50	0.56	0.43	0.43	0.51	0.63	0.83		
2	0.62	-	1.39	0.24	0.77	0.74	0.75	1.00	1.45	0.87	1.19	1.13	0.35	1.88	0.61	0.71	1.28	0.68	0.80	0.62	0.48	0.99	0.63	0.31	0.27	0.56	0.83	0.73	1.52	0.76	0.52	0.91	0.42	0.84	0.71	0.43	0.94	0.72	0.56	0.59	0.89	0.66			
3	0.84	1.39	-	1.24	0.97	0.93	1.16	1.07	1.97	1.23	1.06	1.31	1.10	1.70	1.16	1.36	1.15	1.09	1.28	1.02	1.80	1.23	1.18	1.79	1.07	1.19	1.31	1.05	1.48	1.08	1.37	1.05	1.12	1.20	1.38	1.45	1.12	1.30	1.12	1.14	1.22	1.30	1.26	1.54	
4	0.45	0.28	1.24	-	0.53	0.56	0.52	0.87	1.20	0.64	0.93	0.90	0.61	1.65	0.58	0.49	1.05	0.45	0.61	0.79	0.32	0.88	0.99	0.46	0.23	0.14	0.37	0.80	0.49	0.28	0.63	0.31	0.75	0.32	0.68	0.53	0.25	0.76	0.51	0.38	0.44	0.70	0.47		
6	0.31	0.77	0.97	0.53	-	0.30	0.31	0.59	0.77	0.27	0.63	0.47	1.09	1.16	0.77	0.21	0.62	0.48	0.27	1.13	0.40	0.54	1.04	0.36	0.48	0.55	0.27	0.79	0.22	0.89	0.35	0.29	0.49	0.59	0.55	0.34	0.52	0.46	0.32	0.39	0.46	0.73			
12	0.29	0.74	0.93	0.56	0.30	-	0.32	0.68	0.79	0.39	0.73	0.49	1.04	1.16	0.94	0.31	0.63	0.53	0.30	1.18	0.36	0.47	0.91	0.19	0.51	0.54	0.23	0.62	0.30	0.98	0.33	0.33	0.53	0.67	0.66	0.29	0.59	0.36	0.44	0.55	0.57	0.85			
21	0.46	0.75	1.16	0.52	0.31	0.32	0.70	0.72	0.24	0.50	0.40	1.03	1.21	0.91	0.16	0.56	0.45	0.28	1.17	0.33	0.37	0.88	0.24	0.54	0.49	0.26	0.68	0.18	0.81	0.40	0.26	0.40	0.53	0.41	0.18	0.44	0.28	0.13	0.27	0.36	0.48	0.67			
22	0.58	1.00	0.97	0.87	0.59	0.68	0.70	-	1.06	0.52	0.93	0.77	1.31	1.25	0.85	0.68	0.81	1.04	0.45	1.28	0.79	0.80	1.32	0.69	0.77	0.88	0.64	1.10	0.65	1.00	0.46	0.64	0.45	0.76	0.71	0.58	0.76	0.68	0.61	0.64	0.61	0.49	1.17		
23	1.00	1.45	1.23	1.20	0.77	0.79	1.25	1.07	1.06	-	0.68	0.47	0.34	1.73	0.62	1.51	0.77	0.28	0.94	0.78	1.84	1.01	0.41	0.46	1.31	0.87	1.21	1.20	0.92	1.13	0.74	0.44	0.91	0.95	0.83	1.22	0.92	0.81	1.14	0.62	0.78	0.97	1.03	0.92	1.21
24	0.46	0.87	1.06	0.64	0.27	0.39	0.24	0.52	0.68	-	0.47	0.35	1.18	1.09	0.90	0.27	0.46	0.63	0.18	1.26	0.49	0.38	1.05	0.37	0.63	0.64	0.36	0.84	0.20	0.69	0.36	0.36	0.28	0.59	0.43	0.24	0.53	0.30	0.19	0.34	0.81				
27	0.86	1.19	1.31	0.93	0.63	0.73	0.50	0.93	0.47	0.47	-	0.38	1.48	1.03	1.24	0.58	0.40	0.75	0.64	1.60	0.82	0.46	1.25	0.72	1.00	0.94	0.74	1.10	0.52	0.40	0.81	0.54	0.53	0.69	0.73	0.63	0.86								
28	0.69	1.13	1.10	0.90	0.47	0.49	0.40	0.77	0.34	0.35	0.38	-	1.43	0.82	1.21	0.45	0.18	0.72	0.44	1.53	0.69	0.20	1.08	0.55	0.89	0.89	0.58	0.42	0.52	0.57	0.63	0.51	0.90	0.63	0.48	0.82	0.30	0.46	0.70	0.64	0.99				
29	0.95	0.35	1.70	0.61	1.09	1.04	1.03	1.31	1.73	1.18	1.48	1.43	-	2.19	0.84	1.00	1.58	0.92	1.10	0.54	0.74	1.35	1.01	0.91	0.62	0.56	0.87	0.94	1.04	1.82	1.05	0.83	1.19	0.70	1.08	1.00	0.72	1.22	1.02	0.86	0.87	1.19	0.86		
31	1.21	1.88	1.16	1.65	1.16	1.21	1.25	0.62	1.09	1.03	0.82	2.19	-	1.87	1.23	0.70	1.45	1.14	2.25	1.46	0.99	1.74	1.29	1.61	1.66	1.33	1.52	1.20	1.82	1.22	1.29	1.20	1.68	1.40	1.24	1.24	1.42	1.48	1.22	1.75					
33	0.72	0.61	1.36	0.58	0.77	0.94	0.91	0.85	1.51	0.90	1.24	1.21	0.84	1.87	-	0.81	1.34	0.89	0.85	0.66	0.75	1.24	1.33	0.89	0.52	0.63	0.75	0.71	0.57	0.87	0.87	0.60	1.12	0.85	0.71	0.68	0.74	0.80							
36	0.42	0.71	1.15	0.49	0.21	0.31	0.16	0.68	0.77	0.27	0.58	0.45	1.00	1.23	0.81	-	0.62	0.40	0.28	1.08	0.28	0.87	0.28	0.46	0.23	0.65	0.19	0.87	0.36	0.23	0.44	0.51	0.40	0.24	0.44	0.38	0.22	0.29	0.36	0.45	0.64				
37	0.81	1.28	1.09	1.05	0.62	0.63	0.56	0.58	0.81	0.28	0.46	0.40	0.18	1.58	0.70	1.34	0.62	-	0.88	0.56	1.69	0.86	0.39	0.24	0.82	0.70	1.05	0.74	0.56	0.78	0.59	1.03	0.72	0.61	0.94	0.72	1.13	0.72	1.13						
41	0.60	0.68	1.28	0.45	0.48	0.53	0.45	1.04	0.94	0.63	0.75	0.72	0.92	0.45	0.14	0.89	0.40	0.88	-	0.66	1.06	0.40	0.73	0.91	0.51	0.54	0.46	0.46	0.73	0.45	1.11	0.71	0.50	0.64	0.70	0.58	0.59	0.54	0.63	0.78	0.43				
42	0.36	0.80	1.02	0.61	0.27	0.30	0.28	0.45	0.78	0.18	0.64	0.44	1.10	1.14	0.85	0.28	0.56	0.66	-	1.17	0.42	0.43	0.97	0.29	0.54	0.59	0.28	0.74	0.29	0.84	0.19	0.31	0.26	0.57	0.46	0.17	0.52	0.29	0.23	0.33	0.38	0.41	0.87		
43	1.06	0.62	1.80	0.79	1.13	1.18	1.17	1.28	1.84	1.26	1.60	1.53	0.54	0.25	0.66	1.08	1.69	1.06	1.17	-	0.88	1.49	1.10	1.09	0.70	0.75	0.25	1.00	1.06	1.19	1.00	0.95	1.25	0.79	1.08	1.13	0.87	1.36	1.15	0.98	0.94	1.16	1.00		
44	0.41	0.48	1.23	0.32	0.40	0.46	0.36	0.33	0.79	1.01	0.49	0.82	0.69	0.74	1.46	0.75	0.28	0.86	0.40	0.42	0.88	-	0.64	0.73	0.24	0.28	0.24	0.20	0.52	0.34	1.13	0.41	0.19	0.58	0.42	0.52	0.32	0.36	0.52	0.36	0.29	0.37	0.63	0.62	
46	0.70	1.08	1.18	0.88	0.54	0.47	0.37	0.80	0.46	0.38	0.46	0.20	1.35	0.93	1.24	0.46	0.30	0.73	0.43	1.49	0.64	0.96	0.48	0.87	0.57	0.77	0.43	0.61	0.54	0.60	0.46	0.86	0.59	0.41	0.79	0.17	0.42	0.59	0.66	0.67	1.00				
47	1.07	0.99	1.79	0.99	1.03	0.91	0.94	0.88	1.32	1.31	1.05	1.25	1.08	1.01	1.74	1.33	0.87	1.24	0.91	0.97	1.10	0.73	0.96	0.52	0.88	0.32	0.97	1.50	0.89	0.90	1.01	0.88	0.68	1.00	0.91	0.95	0.91	0.94	1.16	1.13					
48	0.33	0.63	1.07	0.46	0.36	0.19	0.24	0.69	0.87	0.37	0.72	0.55	0.91	1.29	0.89	0.28	0.69	0.51	0.29	1.09	0.24	0.48	0.83	-	0.43	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.62	0.76								
49	0.37	0.31	1.19	0.23	0.48	0.51	0.54	0.77	1.21	0.63	1.00	0.89	0.62	1.61	0.52	0.46	1.08	0.54	0.28	0.28	0.87	0.91	0.43	-	0.21	0.32	0.71	0.50	1.30	0.50	0.29	0.70	0.37	0.49	0.35	0.73	0.52	0.39	0.43	0.67	0.65				
53	0.48	0.27	1.31	0.14	0.55	0.54	0.49	0.88	1.20	0.64	0.94	0.89	0.56	1.66	0.63	0.64	1.04	0.46	0.59	0.75	0.24	0.84	0.42	0.21	-	0.35	0.35	0.70	0.48	1.28	1.09	0.59	0.51	0.69	0.51	0.51									
54	0.24	0.56	1.05	0.37	0.27	0.23	0.26	0.64	0.92	0.36	0.74	0.59	0.87	1.33	0.75	0.23	0.74	0.24	0.60	0.20	0.57	0.15	0.32	0.35	0.23	0.63	0.23	1.03	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31						
55	0.80	1.43	1.48	0.80	0.79	0.62	0.68	1.10	1.13	0																																			

Table 41. MSAV12 similarity matrix for CDL dataset (CA), continued on following page

cat	1	2	3	4	6	12	21	22	23	24	27	28	29	31	33	36	37	41	42	43	44	46	47	48	49	53	54	55	57	58	59	61	66	67	68	69	71	72	74	75	76	77	92		
1	-	1.86	2.16	0.70	1.48	1.23	3.24	3.68	2.11	2.57	2.09	1.88	3.58	3.28	2.58	1.91	1.82	2.36	2.34	1.01	1.29	0.75	1.56	1.45	3.02	1.41	1.59	1.97	3.18	2.75	1.55	2.73	1.76	1.76	2.29	1.26	1.32	1.63	1.29	1.48	1.60	2.35	1.59		
2	1.86	-	0.85	1.35	2.03	1.99	3.56	4.49	3.46	3.34	2.42	3.10	2.77	3.95	2.95	3.24	3.29	4.15	2.52	1.81	1.58	1.49	1.97	0.98	3.20	1.84	1.40	2.05	3.32	4.12	3.01	2.95	2.95	3.42	3.20	2.33	2.81	2.35	2.70	2.65	3.40	2.66			
3	2.16	0.85	-	1.79	2.34	2.40	4.05	4.44	3.83	3.75	3.67	3.53	3.06	4.20	3.22	3.49	3.63	4.28	2.79	2.25	1.95	2.37	1.50	3.40	2.35	1.76	2.47	3.72	4.35	3.30	3.46	3.23	2.83	3.61	2.78	2.84	3.34	2.84	3.03	2.93	3.58	2.99			
4	0.70	1.35	1.79	-	1.17	1.03	2.86	3.56	2.33	2.41	2.35	2.98	3.00	3.10	2.30	2.46	2.21	3.00	1.98	0.96	0.96	0.63	1.60	1.84	2.73	0.93	1.06	1.75	2.77	3.33	2.13	2.32	2.25	2.12	2.83	1.41	1.40	1.86	1.42	1.87	2.02	2.87	1.68		
6	1.48	2.03	2.34	1.17	-	0.68	2.60	3.20	2.80	2.35	2.73	2.35	2.61	2.80	1.48	3.12	2.74	3.24	1.76	1.33	1.64	1.69	2.20	1.50	1.82	0.97	1.75	2.30	3.98	2.79	2.04	2.84	2.70	3.47	1.87	1.84	2.39	1.88	2.38	2.61	3.51	1.99			
12	1.23	1.99	2.40	1.03	0.68	-	2.64	3.13	2.55	2.22	2.47	2.10	2.71	2.84	1.60	2.73	2.39	3.03	1.74	1.12	1.27	1.53	1.81	1.48	1.97	0.81	0.97	1.53	2.35	3.58	2.40	2.01	2.40	2.23	3.04	1.40	1.38	1.93	1.39	1.87	2.16	3.10	1.65		
21	3.24	3.56	4.05	2.86	2.60	2.64	-	2.40	2.69	1.26	2.87	2.21	2.70	1.55	2.09	4.30	3.22	4.41	1.78	2.92	2.54	3.19	3.44	2.79	2.53	2.63	2.70	2.93	1.04	4.98	4.06	0.80	4.16	4.15	4.85	2.89	2.68	2.99	3.55	4.00	4.97	2.47	2.47		
22	3.68	4.29	4.44	3.56	3.20	3.13	2.40	-	3.04	1.80	2.59	2.72	3.35	1.26	2.10	4.25	3.29	4.11	1.86	3.44	3.06	4.00	3.68	3.65	2.34	2.94	3.35	3.22	1.91	4.85	4.15	2.27	4.11	4.05	4.70	3.24	3.10	3.48	3.14	3.46	3.94	4.83	2.43		
23	2.11	3.46	3.83	2.33	2.80	2.55	2.69	3.04	-	1.67	0.81	0.61	4.35	4.23	3.07	2.25	0.98	2.24	2.56	2.31	2.13	2.32	2.57	2.77	3.67	2.27	3.03	2.94	3.06	2.65	2.06	2.59	2.43	2.76	2.92	1.89	1.75	1.58	1.78	2.04	2.47	3.08	1.40		
24	2.57	3.34	3.75	2.41	2.35	2.22	1.26	1.80	1.67	-	1.66	1.20	3.12	1.09	1.90	3.31	2.14	3.36	1.38	2.35	1.97	2.72	2.75	2.56	2.46	1.75	2.52	2.49	1.47	3.93	3.11	1.17	3.22	3.29	3.88	2.11	1.90	2.17	1.99	2.58	3.09	4.01	1.42		
27	2.09	3.42	3.67	2.35	2.73	2.47	2.87	2.59	0.81	1.66	-	0.93	2.43	2.24	2.81	0.91	2.05	2.29	2.22	2.06	2.48	2.38	2.08	3.37	2.91	2.95	2.71	1.98	2.54	1.97	2.65	2.22	2.48	2.70	1.80	1.70	1.69	1.70	1.77	2.22	2.86	1.05			
28	1.88	3.10	3.53	1.98	2.35	2.10	2.21	2.72	0.61	1.20	0.93	-	3.79	2.11	2.54	2.34	1.96	2.46	2.03	1.92	2.07	2.26	2.36	3.13	1.76	2.56	2.46	2.51	2.89	2.10	2.02	2.37	2.61	2.96	1.51	1.34	1.34	1.40	1.86	3.11	1.02				
29	3.58	2.77	3.06	3.00	2.61	2.71	2.70	3.35	4.35	3.12	4.23	3.79	-	3.34	2.20	5.01	4.48	5.48	2.02	3.24	3.71	3.42	3.58	2.46	2.08	2.09	2.83	1.98	5.90	4.78	2.17	4.65	4.29	5.29	3.49	3.40	4.00	4.06	4.33	5.34	3.24				
31	3.28	3.95	4.20	2.10	2.80	2.84	2.84	1.55	2.68	2.43	1.09	2.24	2.11	2.34	-	2.02	4.00	2.88	3.85	1.70	3.02	2.74	3.51	3.43	3.24	2.46	2.45	3.06	3.01	1.55	4.60	3.85	1.69	3.94	2.96	4.57	2.93	2.75	3.06	2.83	3.31	3.79	4.69	2.12	
33	2.58	2.95	3.32	2.30	1.48	1.60	2.09	2.10	3.07	1.90	2.81	2.54	2.20	2.02	-	3.73	3.08	3.77	1.16	2.11	2.14	2.84	2.69	2.40	0.61	1.63	1.75	1.85	1.34	4.53	3.49	1.51	3.41	3.25	4.09	2.32	2.23	2.80	2.28	2.80	3.17	4.16	2.06		
36	1.91	3.24	3.49	2.65	3.12	2.71	4.30	4.25	2.25	3.31	2.94	2.08	2.34	5.01	4.00	3.73	-	1.36	1.81	3.47	2.06	2.42	2.37	1.73	1.73	1.11	2.12	2.23	2.77	1.04	1.70	1.05	0.88	0.85	1.99	1.71	1.51	1.70	1.05	0.88	1.99				
37	1.82	3.29	3.63	2.21	2.74	2.39	3.22	3.29	0.96	2.14	0.91	1.08	1.48	2.88	3.08	1.36	-	1.74	2.70	1.92	2.01	2.19	1.78	3.63	2.34	2.98	2.52	3.37	1.82	1.23	2.91	1.53	1.95	2.01	1.37	1.34	1.03	1.32	1.26	1.63	2.18	1.17			
41	2.36	4.15	4.28	3.00	3.24	3.03	4.41	4.11	2.24	3.36	2.05	2.46	2.58	3.77	3.01	2.74	-	3.22	2.95	3.01	3.70	4.23	3.33	3.70	3.49	4.46	1.80	1.65	4.12	2.08	2.59	2.20	2.52	2.59	2.39	2.53	2.15	2.29	2.33	2.45					
42	2.34	2.52	2.79	1.99	1.76	1.74	1.78	1.86	2.56	2.24	1.38	2.29	2.09	2.02	1.16	3.47	2.70	3.75	-	2.02	1.52	2.45	2.41	1.87	1.59	1.35	1.63	1.87	1.18	4.28	3.27	1.17	3.23	3.05	3.89	2.09	1.95	2.53	2.02	2.58	2.97	3.98	1.60		
43	1.01	1.81	2.25	0.96	1.33	1.12	2.92	3.44	2.31	2.35	2.22	1.92	3.24	3.02	2.11	2.06	1.92	2.74	2.02	-	1.17	1.30	1.01	1.55	2.25	1.50	1.07	1.77	2.77	2.93	1.78	2.33	1.66	2.40	0.87	0.94	1.41	0.98	1.40	1.52	2.44	1.39			
44	1.29	1.58	2.07	0.96	1.64	1.27	2.54	3.05	2.36	2.13	1.97	2.06	1.70	2.71	2.74	2.14	2.42	2.01	3.22	1.52	1.17	-	1.28	1.32	1.08	2.58	1.94	1.34	1.47	2.37	3.28	2.20	1.92	2.16	1.95	2.79	2.79	1.11	1.04	1.59	1.06	1.61	1.89	2.86	1.23
46	0.75	1.49	1.95	0.63	1.69	1.53	3.19	4.00	2.32	2.72	2.46	2.07	3.42	3.51	2.84	2.37	2.20	2.95	2.45	1.30	1.28	1.23	1.80	1.04	3.28	1.41	1.53	2.18	3.23	3.17	2.00	2.71	2.22	2.19	2.72	1.17	1.61	1.89	1.62	1.98	2.07	1.97			
47	1.56	2.37	2.77	1.30	2.20	1.81	3.44	3.68	2.57	2.75	2.26	2.38	2.65	3.58	3.43	2.69	2.31	2.57	1.93	1.70	3.01	2.41	1.01	2.02	3.06	1.91	1.92	0.93	1.08	1.39	1.05	1.30	1.02	1.94	1.54	1.54	1.94	1.54	1.94	1.54	1.94				
48	1.45	0.98	1.50	0.84	1.50	1.48	2.79	3.65	2.77	2.56	2.46	2.34	2.40	3.11	2.78	3.70	1.87	1.55	2.04	1.90	2.02	-	2.75	1.11	0.90	1.99	2.64	3.97	2.80	2.23	2.86	2.65	3.42	1.93	1.40	2.46	2.60	3.47	2.12	1.87					
49	3.02	3.20	3.40	2.73	1.82	1.97	2.53	2.46	3.67	2.44	3.37	3.13	2.04	2.46	0.61	4.18	3.66	4.23	1.59	2.55	2.88	3.06	2.75	-	2.10	2.01	2.15	2.65	5.00	3.97	2.06	2.32	2.40	3.47	2.33	2.33	3.55	4.53	2.59	2.59					
51	1.41	2.84	2.35	0.93	0.71	2.03	2.47	2.27	2.31	2.17	2.36	2.16	2.45	2.45	1.63	2.92	2.32	2.64	2.03	1.91	1.11	1.31	1.11	1.27	1.27	1.27	1.27	1.27	1.27	1.27	1.27	1.27	1.27	1.27	1.27	1.27	1.27	1.27	1.27	1.27	1.27				
55	1.97	2.05	2.47	1.75	1.75	1.53	2.93	3.22	3.29	2.94	2.49	2.71	2.6	2.04	3.11	2.76	2.04	2.16	1.69	1.53	2.11	1.23	1.27	1.27	2.52	1.27																			

Table 42. VNIROI similarity matrix for CDL dataset (CA), continued on following page

8.6 GRASS Modules

Various GRASS modules were written in Python to perform the calculated need to complete this research. Some of the more important scripts are summarized in this section while other, mainly used for plot generation, are not included.

8.6.1 probav_prep.py

Description

This script prepares the raw PROBA-V data for classification attempts. It extracts raw data from achieves while applied a standard labeling format. Labeled reflectance raster images are then imported into GRASS using GDAL. The script then generates spectral index maps and runs the GRASS *r.hants* module with desired options.

Inputs

PROBA-V reflectance ZIP files (accompanied by XML metadata)

Outputs

- Labeled GEOTIF and XML files organized by location.
- Raw reflectance maps imported to GRASS
- Raw index maps
- Reconstructed HANTS index maps and harmonic components

8.6.2 unsup_classifier.py and error_kappa.py

Description

These two scripts were used to generate unsupervised classification results. The first script uses the *i.pca* module to perform principle component analysis on harmonic component maps before using the *i.cluster* and *i.maxlik* modules for unsupervised classification. The second script first uses statistical information from *r.univar* for data-assisted labeling before performing an accuracy assessment with the *r.kappa* module.

Inputs

Harmonic component index maps from HANTS

Outputs

- Unlabeled and labeled classification maps
- Accuracy assessment values

8.6.3 sig_classifier.py

Description

This script performs the RMSE classification as described in section 3.6.2 as well as code needed for the attempts to improve accuracy. This involves traversing through the one-year time series multiple times for intermediate calculations and generation final results.

Inputs

Times series HANTS index maps

Outputs

- RMSE maps for each class/index
- Labeled classification maps (up to top 5 results)
- Accuracy assessment values (up to top 5 results)

8.6.4 similarity_matrix.py

Description

These two scripts were used to generate the similarity matrices listed in section 8.5. This is done by generating RMS error between mean index curves for each class.

Inputs

Times series HANTS index maps

Outputs

Similarity matrix for selected index and study area in CSV format

8.6.5 calc_class_var.py

Description

This script was used to generate data in appendix section 8.4 which highlights in-class variation. This module largely relies on univariate statistical data generated by the *r.univar* module.

Inputs

Time series HANTS NDVI maps and 1st order phase maps

Outputs

- Monthly standard deviation by class (mean and st. dev)
- Peak vegetation by class (mean and st. dev)