"Yearly cropland mapping over large areas with high resolution satellite image time series"

Waldner, François

Abstract
With human population growth, shifts in diets, biofuel development and climate change, the food supply system is subject to increasing pressures. In this context, timely and dependable information on crop production becomes crucial for market stability and food security. In spite of advances in satellite systems and data processing, there is a disconnect between operational cropland mapping and the state-of-the-art. This thesis seeks to bridge this gap by capitalizing on available land cover maps and by optimizing the satellite inputs. First, priority areas are identified for all countries to strategically allocate future mapping efforts. Second, methods are proposed to enable regular cropland mapping over large areas in the absence of in situ calibration data. The combination of calibration data carefully selected from available land cover maps and spectral-temporal features derived from the satellite image time series yields spatially consistent results with an accuracy that varies...

Document type: Thèse (Dissertation)

Référence bibliographique

Waldner, François. Yearly cropland mapping over large areas with high resolution satellite image time series. Prom. : Defourny, Pierre

Available at: http://hdl.handle.net/2078.1/186331
[Downloaded 2019/03/04 at 02:22:13]
Yearly cropland mapping over large areas with high resolution satellite image time series

François Waldner

Mai 2017

Thèse présentée en vue de l’obtention du grade de docteur en sciences agronomiques et ingénierie biologique
Composition du jury:

**Promoteur**
Pr Pierre Defourny (Université catholique de Louvain, Belgique)

**Président de jury**
Pr Marnik Vanclooster (Université catholique de Louvain, Belgique)

**Lecteurs**
Pr Patrick Bogaert (Université catholique de Louvain, Belgique)
Dr Patrick Meyfroidt (Université catholique de Louvain, Belgique)
Pr Matthew Hansen (University of Maryland, Etats-Unis)
Dr Agnès Bégué (Centre de coopération Internationale en Recherche Agronomique pour le Développement, France)
The ultimate goal of farming is not the growing of crops, but the cultivation and perfection of human beings.

—Masanobu Fukuoka, The One-Straw Revolution
Acknowledgements

Tout d’abord, je tiens à remercier mon promoteur, Pierre Defourny, qui m’a lancé sur les chemins tortueux de l’aventure doctorale. Je le remercie pour sa confiance et pour les nombreuses opportunités qu’il m’a offertes.

Je tiens ensuite à remercier mon jury de thèse pour ses encouragements mais aussi pour leurs questions qui m’ont permis d’approfondir mes recherches. Un très grand merci à Agnès pour son engagement lors des multiples formalités académiques du parcours doctoral ainsi que son enthousiasme lors de nos échanges scientifiques. My deep gratitude goes to Matthew Hansen for accepting to be part of this jury, for waking up at dawn for a long private defense and flying over the Atlantic for the final presentation. His mind-set and comments were very refreshing and much appreciated. Merci également à Patrick Meyfroidt et Patrick Bogaert pour leurs relectures attentives et leurs contributions à l’amélioration du manuscrit de thèse (la première mouture en avait certes fort besoin). Merci aussi à Marnik Vanclooster d’avoir accepté de présider ce jury.

The research carried out in this thesis was supported by three European projects MOCCCASIN (MOonitoring Crops in Continental Climates through Assimilation of Satellite Information), IMAGINES (Implementation of Multi-scale Agricultural Indicators Exploiting Satellites), and SIGMA (Stimulating Innovation for the Global Monitoring of Agriculture). It was a great opportunity to link my activities to a broader context and engage with international researchers. I acknowledge also the support and advice of the partners I worked with: Fred Baret, Marie Weiss, Wenjuan Li, Antonio Di Gregorio, Anne Schucknecht, Javier Gallego, and especially Steffen Fritz and his team (Myroslava Lesiv, Juan-Carlos Lasso-Bayas, Linda See) for their kind and repeated welcome. I also thank my buddy Fabian Löw with whom I share many interests. Je tiens aussi à remercier Greg, mon beauf télédétecteur et PSF-enthousiasant, pour nos nombreuses discussions et particulièrement pour son investissement, son encadrement et ses conseils pour mon quatrième chapitre.

Most of my work would not have been possible without the appropriate data. Therefore, I would like to thank all the people who contributed some data formally or informally. I would like to deeply thank my JECAM partners would collected the field data used in the second chapter, in particular Diego De Abelleysa, Santiago Verón, Miaozhang, Bingfang Wu, Dmitry Plotnikov, Sergey Bartalev, Mykola Lavreniuk, Sergii Sleukun, Nataliia Kussul, Guerriec Le Maire, Stéphane Dupuy, Ian Jarvis. I appreciate our collaboration and I look forward to our next experiments. I acknowledge all the people behind the SPOT Take-5 initiative who provided fantastic time series to anticipate Sentinel-2. My gratitude goes also to all institutions that have open data sharing policies, such as ESA, USGS, etc. A special thank you to those who shared the Land Parcel Identification System of Belgium and to the Department of Agriculture, Forestry and Fisheries of South Africa for sharing the field boundary data. I also acknowledge
the Computational resources provided by the Consortium des Equipements de Calcul Intensif (CECI).

Je remercie mes nombreux collègues en ENGE avec qui j’ai partagé le plus clair de mon temps. Tout d’abord merci à ceux qui m’ont accueilli à mes débuts et qui ont quitté le labo entre temps : le Ska, les Emilies et Astrid, Catherine, Manuel Alarcon Milton Guzman, JPK, François Stevens. Merci à Aline pour ses explications et conseils en traitement d’images radar. Il y a aussi ceux qui étaient là à mon arrivée et qui le sont toujours. Merci à Céline pour nos longues discussions à la cafête, dans l’aquarium, mais aussi en Italie, en Ecosse, en Espagne et j’en passe... et à Inés pour avoir partagé quelques gâteries sucrées sur le balcon lorsque 4h sonnait. Merci aussi aux post-docs Sophie et Xa pour leur partage d’expérience. Dans le même registre, merci au joyeux drille radouxju (fort de 31.311 points sur gis.stackexchange.com au moment de l’écriture de ces remerciements). Un tout grand merci à Guada et Mayo pour nous avoir sorti de notre morne belgitude en apportant une touche d’exotisme au 3e étage du Mendel (et pour avoir balancé la sauce). J’en profite aussi pour faire un clin d’oeil à mes contemporains au labo qui sont parti avant moi (Quentin, Renaud, Catherine, Coco) et au nain sous la montagne. Puis merci à la nouvelle génération d’ENGEs (ou démons ?) qui me fait comprendre que j’ai pris quelques rides. Marie-Julie pour son amitié malgré la lourdeur de mes blagues, Guillaume pour ses lumières engagées et ses conseils niveau siste sous les bureaux, Flo pour sa bonhomie, Cindy pour son sourire, NicoB le roi du Ugly Coyote, Julie et NicoM(ontenegro). Merci aussi aux ENGEs non-télédétecteurs Emmanuel, Pascale, Aurélie, Sarah, Thomas, et Antoine ainsi qu’aux copains PhDHub et de Focus Research. Merci à St Nicolas pour sa générosité et son humour anuel (PS : on attend toujours les canapés).

Ah oui, j’allais oublier mes deux compères du C.385 : Raph et Damien Jacques ex de l’UCL et désormais de l’ULB. Il faudrait un document certainement plus long que cette thèse pour raconter toutes nos frausses. Aussi étrange que cela puisse paraître, nous avons toujours su garder un bon équilibre entre science et... non-science (Mouhahaha). Nous avons été les les GlobCroppers, les rois des Balkans, de challengers with strong skills au NASA hackatton, les trois cowboys, les gagnants du D4D... et bientôt nous serons tous trois docteurs... Enfin! A special thank you to the guru-gis followers and their e-mails of encouragement.

Merci à tous mes amis qui m’ont soutenu et changé les idées durant ces quatre longues années et des chouïas, en particulier à ceux qui ont réussi à remplacer mon mal de tête scientifique par un autre mal de tête.

Un merci tout particulier à mes merveilleuses famille et belle-famille pour leur soutien inconditionnel et leur réconfort. Enfin, merci à ma femme Emilie qui a pris soin de moi pendant toutes ces années doctorales, qui m’a remonté le moral et entouré de son amour. Cette thèse n’aura pas été possible sans cela. Maintenant que celle-ci s’achève, je me réjouis du futur qui nous attend.

Washington D.C., the 16th of April 2017
# Table of Contents

Table of Contents .................................................. v
List of Figures ......................................................... ix
List of Tables .......................................................... xiii
List of Acronyms ......................................................... xv
List of Symbols ........................................................ xvii

## Introduction

Feeding the planet in a changing environment .................. 1
Market intelligence ......................................................... 2
- Sources of commodity market price volatility and its impacts . 3
- The Great Grain Robbery .............................................. 4
- Food security and the response of the international community 4
Agriculture monitoring with Earth Observation ................... 6
- The importance of cropland masks and their current uncertainty 8
- Fifty shades of cropland definition .................................. 10
Scope and objectives ...................................................... 11
Outline of the thesis ..................................................... 14
Review of concepts and methods relevant for cropland mapping research 15
- The spatial resolution of remotely sensed images ............... 15
- Spectral and temporal measurement of vegetation by satellite 17
  - Optical properties of the leaves .................................. 18
  - Soil background properties ........................................ 19
  - Spectral indices ..................................................... 20
  - Texture information ............................................... 20
- Temporal analysis of the crop reflectance ....................... 20
Characterizing the fragmentation of agricultural landscapes .... 22
Agriculture mapping from satellite images time series .......... 23
- Image classification techniques .................................... 23
Accuracy assessment ....................................................... 24
- Accuracy assessment based on the confusion matrix .......... 24
- Pareto boundary analysis .......................................... 26
Synthesis of crop classification performance across landscapes 26
1 Mapping priorities to focus cropland mapping activities: fitness assessment of existing global, regional and national cropland maps 29
   1.1 Introduction .............................................. 30
   1.2 Material ................................................. 32
       1.2.1 Land cover and cropland maps ................. 32
       1.2.2 Global validation data sets .................. 34
       1.2.3 Ancillary data ........................................ 35
   1.3 Method .................................................. 36
       1.3.1 Thematic Consistency Criterion ............... 37
       1.3.2 Timeliness Criterion ............................. 38
       1.3.3 Resolution adequacy criterion ................ 39
       1.3.4 Confidence level criterion ..................... 40
       1.3.5 Criteria aggregation and priority identification .. 40
       1.3.6 Spatial aggregation and assessment .......... 42
   1.4 Results and Discussion ................................. 43
       1.4.1 Multi-criteria analysis ......................... 43
       1.4.2 Spatial aggregation: assessment and comparison . 45
   1.5 Discussion ............................................... 48
   1.6 Conclusion ............................................... 51

2 Automated Annual Cropland Mapping using Knowledge-Based Temporal Features 53
   2.1 Introduction ............................................... 54
   2.2 Materials ................................................. 56
       2.2.1 Remote Sensing Data .............................. 56
       2.2.2 Existing baselines and validation data ........ 58
   2.3 Methodology ............................................... 58
       2.3.1 Cropland Classification Method ................. 58
             Knowledge-based feature extraction ............ 59
             Cleaning imperfect and outdated baselines .... 60
             Classification method ............................... 62
       2.3.2 Accuracy assessment and spatial thematic uncertainty based on high resolution time series ........ 63
       2.3.3 Temporal robustness analysis based on MODIS time series 63
             Year-to-year generalization ...................... 64
             Near real time generalization ................. 64
   2.4 Study sites ............................................... 65
   2.5 Results .................................................. 67
       2.5.1 High resolution temporal features .............. 67
       2.5.2 Cropland classification on high spatial resolution features 70
       2.5.3 Temporal generalization with coarse resolution time series 73
   2.6 Discussion ................................................. 75
   2.7 Conclusions ............................................... 77
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>National-scale cropland mapping based on spectral-temporal features and historical land cover information</td>
<td>79</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>80</td>
</tr>
<tr>
<td>3.2</td>
<td>Study area</td>
<td>82</td>
</tr>
<tr>
<td>3.3</td>
<td>Data</td>
<td>83</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Satellite data pre-processing</td>
<td>83</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Historical land cover map, validation and ancillary data</td>
<td>84</td>
</tr>
<tr>
<td>3.4</td>
<td>Methods</td>
<td>84</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Classification scheme</td>
<td>85</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Extraction of spectral-temporal features</td>
<td>85</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Identifying reliable pixels from the historical land cover map</td>
<td>86</td>
</tr>
<tr>
<td>3.4.4</td>
<td>Stratum-specific classification to handle the spectral diversity</td>
<td>87</td>
</tr>
<tr>
<td>3.5</td>
<td>Methods</td>
<td>89</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Post-classification filtering</td>
<td>88</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Evaluation of the classification accuracy</td>
<td>89</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Explanatory variables of the classification accuracy</td>
<td>89</td>
</tr>
<tr>
<td>3.5.4</td>
<td>Remote sensing features of importance</td>
<td>91</td>
</tr>
<tr>
<td>3.6</td>
<td>Results</td>
<td>91</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Visual assessment</td>
<td>91</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Map accuracy assessment</td>
<td>92</td>
</tr>
<tr>
<td>3.6.3</td>
<td>Drivers of accuracy</td>
<td>94</td>
</tr>
<tr>
<td>3.6.4</td>
<td>Spectral-temporal features of importance</td>
<td>95</td>
</tr>
<tr>
<td>3.7</td>
<td>Conclusions</td>
<td>99</td>
</tr>
<tr>
<td>4</td>
<td>Optimizing the spatial resolution of Sentinel-2 to reduce the burden of data volume and its implications for cropland mapping over large areas</td>
<td>101</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>102</td>
</tr>
<tr>
<td>4.2</td>
<td>Data and study sites</td>
<td>105</td>
</tr>
<tr>
<td>4.3</td>
<td>Methodology</td>
<td>107</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Characterizing the resolution-dependent error: $\Delta E$</td>
<td>107</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Three aggregation approaches</td>
<td>107</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Characterizing landscape fragmentation</td>
<td>111</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Mapping scale requirements while respecting minimal error targets</td>
<td>111</td>
</tr>
<tr>
<td>4.4</td>
<td>Results</td>
<td>113</td>
</tr>
<tr>
<td>4.4.1</td>
<td>The effect of aggregation on $\Delta E$</td>
<td>113</td>
</tr>
<tr>
<td>4.4.2</td>
<td>The spatial distribution of $\Delta E$</td>
<td>113</td>
</tr>
<tr>
<td>4.4.3</td>
<td>The relationship between $\Delta E$ and landscape fragmentation</td>
<td>115</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Data reduction using an optimized resolution</td>
<td>118</td>
</tr>
<tr>
<td>4.5</td>
<td>Discussion</td>
<td>119</td>
</tr>
<tr>
<td>4.6</td>
<td>Conclusion</td>
<td>122</td>
</tr>
</tbody>
</table>
# Table of Content

## 5 Where can pixel counting area estimates meet user-defined accuracy requirements?  
### 5.1 Introduction  
### 5.2 Concepts and methods  
#### 5.2.1 Landscape characterization  
#### 5.2.2 Evaluating the area estimation bias  
#### 5.2.3 Defining the maximum tolerable classifier bias  
#### 5.2.4 Modeling the sensor’s spatial response  
### 5.3 Application to cropland mapping in South Africa  
#### 5.3.1 Study site  
#### 5.3.2 Data  
#### 5.3.3 Calibration and results  
### 5.4 Discussion  
### 5.5 Conclusion  

## Conclusions and perspectives  
### Synthesis  
How to take advantage of diverse ancillary data sources of variable quality to support the continuous update of cropland maps?  
By identifying where to focus cropland mapping efforts in priority.  
By combining the best cropland maps.  
By extracting training data to calibrate new classifiers.  
### How to optimize the remote sensing input to map cropland over large areas?  
By using spectral-temporal features.  
By optimizing the spatial resolution.  
By identifying the applicability domain of area estimation  
### Perspectives and implications for operational cropland mapping  
### Potential improvements and evolutions  
### Implication for operational cropland mapping  
### The bigger picture  

## References  

## List of publications
# List of Figures

1. Population and cereal production over time ........................................ 1
2. Wheat price per ton for the 1866-2015 period ................................. 4
3. The three components of crop monitoring (area estimates, yield forecasts, crop status) benefit from the full range of Earth Observation data. ........................................ 7
4. Requirements of operational cropland mapping ............................... 13
5. Schema of the overall thesis approach ............................................... 14
6. Schematic representation of the different factors involved in the image acquisition process ........................................ 16
7. Spectral-temporal signature of crops ................................................. 22
8. Landscape fragmentation of selected agricultural landscapes .......... 23
9. Procedure to derive the Pareto boundary ........................................... 27
1.1 The analytical framework of the assessment is a systematic analysis of four criteria that characterize the fitness of a cropland map ........................................ 37
1.2 Global distribution of field size per country and histograms of the field size distribution for two contrasted countries ........................................ 41
1.3 Distribution of the reference samples used for the validation of the 250-m unified cropland layer ........................................ 42
1.4 Priority map and its update typology ............................................... 44
1.5 National-level agricultural profiles ............................................... 45
1.6 Cropland proportion from the unified cropland layer at 250 m for the year 2014 ........................................ 46
1.7 Agreement map between GLC-Share, the IIASA Cropland and the unified cropland layer ........................................ 48
2.1 Availability of the high resolution images and their temporal distribution over the four demonstration sites ........................................ 57
2.2 Methodology to derive the cropland map ........................................ 59
2.3 Procedure for the knowledge-based feature extraction .................... 61
2.4 Temporal NDVI profiles and related knowledge-based features ........ 61
2.5 Location of the four sites ......................................................... 65
2.6 Zooms of the four study sites .................................................... 66
2.7 Baseline map and false color infra-red knowledge-based features over the Belgian site ........................................ 68
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.8</td>
<td>Scatter plot of red band of the five knowledge-based features.</td>
<td>69</td>
</tr>
<tr>
<td>2.9</td>
<td>Baseline, classification map and uncertainty map for the four sites.</td>
<td>71</td>
</tr>
<tr>
<td>2.10</td>
<td>Distribution of the uncertainty criterion.</td>
<td>73</td>
</tr>
<tr>
<td>2.11</td>
<td>Generalization accuracy of the classification approach.</td>
<td>75</td>
</tr>
<tr>
<td>3.1</td>
<td>South Africa, its main agricultural regions and its geographic breakdown in nine provinces.</td>
<td>83</td>
</tr>
<tr>
<td>3.2</td>
<td>Flowchart of the classification scheme to derive national-scale cropland maps based on outdated land cover information and spectral-temporal features.</td>
<td>86</td>
</tr>
<tr>
<td>3.3</td>
<td>False color composite (minNDVI.red, minNDVI.nir, minNDVI.swir1) of the study area for the years 2013-2015.</td>
<td>87</td>
</tr>
<tr>
<td>3.4</td>
<td>Updated South African Cropland map and the corresponding pixel-level thematic confidence map.</td>
<td>91</td>
</tr>
<tr>
<td>3.5</td>
<td>Selected zooms over four contrasted sites.</td>
<td>93</td>
</tr>
<tr>
<td>3.6</td>
<td>Spatially constrained accuracy assessment for the overall accuracy, and the F-score of the cropland and non-cropland classes.</td>
<td>95</td>
</tr>
<tr>
<td>3.7</td>
<td>Critical Distance diagram.</td>
<td>97</td>
</tr>
<tr>
<td>4.1</td>
<td>Schematic representation of the different factors involved in the image acquisition process.</td>
<td>104</td>
</tr>
<tr>
<td>4.2</td>
<td>Distribution of cropland and the centroids of the very high resolution binary sample units in (a) Belgium and (b) South Africa.</td>
<td>106</td>
</tr>
<tr>
<td>4.3</td>
<td>Pareto boundaries for a high and a low spatial resolution in red and blue, respectively.</td>
<td>108</td>
</tr>
<tr>
<td>4.4</td>
<td>Flowchart describing how the three resampling scenarios were modeled starting from a VHR binary mask.</td>
<td>109</td>
</tr>
<tr>
<td>4.5</td>
<td>Application of the error reduction principles to a multi-scale analysis for a landscape exhibiting a high fragmentation and medium fragmentation.</td>
<td>112</td>
</tr>
<tr>
<td>4.6</td>
<td>Histograms of the mean absolute classification error between i) naive and convolved sub-pixel proportions at 10 m, ii) naive and convolved sub-pixel proportions at 30 m, and iii) naive and upscaled sub-pixel proportions at 30 m for Belgium and South Africa.</td>
<td>114</td>
</tr>
<tr>
<td>4.7</td>
<td>Spatial prediction of $\Delta E$ between 30 and 10 m.</td>
<td>115</td>
</tr>
<tr>
<td>4.8</td>
<td>Spatial prediction of the gain (a, c) and loss (b, d) due to upscaling from 10 to 30 m.</td>
<td>116</td>
</tr>
<tr>
<td>4.9</td>
<td>Relationships between landscape fragmentation and $\Delta E$ under the convolved and the naive resolution as well as the gain and loss due to upsampling for Belgium and South Africa.</td>
<td>117</td>
</tr>
<tr>
<td>4.10</td>
<td>Maps of the optimized resolution for (a) Belgium and (b) South Africa.</td>
<td>118</td>
</tr>
<tr>
<td>5.1</td>
<td>Example of how mixed pixels occur at different spatial resolutions and across several landscape fragmentation.</td>
<td>131</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.2</td>
<td>Example of relationship between landscape fragmentation, spatial resolution, and the resolution bias.</td>
<td>134</td>
</tr>
<tr>
<td>5.3</td>
<td>Distribution of the VHR sample units in South Africa.</td>
<td>136</td>
</tr>
<tr>
<td>5.4</td>
<td>Area estimation bias properties for all sample units plotted across the pixel size and landscape fragmentation dimensions.</td>
<td>137</td>
</tr>
<tr>
<td>5.5</td>
<td>Estimated boundaries in the pixel resolution-landscape fragmentation space used to define the maximum tolerable classifier bias for crop area estimation.</td>
<td>138</td>
</tr>
<tr>
<td>5.6</td>
<td>Spatialization of the maximum tolerable classifier error given an accuracy target of 10% and a spatial resolution of 10 m and 30 m.</td>
<td>139</td>
</tr>
<tr>
<td>10</td>
<td>Proposed global cropland mapping system.</td>
<td>152</td>
</tr>
</tbody>
</table>
List of Tables

1. Diversity of agriculture-related legends .......................... 11
2. Definition of scale-related terms. .............................. 17
3. Illustration of the effect of spatial resolution on six agricultural landscapes of 5x5 km². ................................. 18
4. Illustration of texture information, in this case contrast, extracted at 30 and 100 m across six agricultural landscapes. .................. 21
5. Landscape metrics of four synthetic landscapes. The metrics are computed for the red class which is constant in all landscapes. 24
6. Confusion matrix of four classes with cell entries \((p_{ij})\) expressed in number of pixels/objects. ............................... 25

1.1 Input maps for the analysis. The first and second column detail the extent and references of the considered product. The last column provide the reference year of the product or its time span. 33
1.2 Validation data sets collected, their geometries and percentage of cropland samples. ................................. 35
1.3 Systematic recoding rules for the four cropland criteria used in the multi-criteria analysis .............................. 39
1.4 Linking the observed field size by crowd-sourcing to spatial resolution requirements ............................... 40
1.5 Accuracy assessment of the Unified Cropland Layer. ........ 47

2.1 Source of the baseline data for each site. .................... 58
2.2 Knowledge-based features and their interpretation. .......... 60
2.3 Accuracy indices of the high resolution classifications. .... 70
2.4 Confusion matrices for the MODIS classification on the observed and interpolated reflectances and the associated McNemar’s test. 74

3.1 Potential explanatory variables of accuracy. .................. 90
3.2 Accuracy measures for the classification without filtering and with two different spatial filters. .............................. 94
3.3 Variable importance derived from the MARS models. ........ 96

4.1 Definition of scale-related terms. .............................. 103
4.2 Quantification of how data volume and precessing time can be reduced by using an optimized resolution for cropland mapping for the different countries studied. ............................... 121
5.1 Proportion of the country area and cropland area for which the
accuracy target can be met at 10 or 30 m. . . . . . . . . . . . 138
7 Summary table of the achievements and the main findings of the
thesis. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 144
# List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACI</td>
<td>Annual Cropland Inventory</td>
</tr>
<tr>
<td>AgRISTARS</td>
<td>Agriculture and Resources Inventory Surveys through Aerospace Remote Sensing</td>
</tr>
<tr>
<td>ALOS</td>
<td>Advanced Land Observing Satellite</td>
</tr>
<tr>
<td>AMIS</td>
<td>Agricultural Market Information System</td>
</tr>
<tr>
<td>CAP</td>
<td>Common Agriculture Policy</td>
</tr>
<tr>
<td>CAP</td>
<td>Common Agricultural Policy</td>
</tr>
<tr>
<td>CCI</td>
<td>Climate Change Initiative</td>
</tr>
<tr>
<td>CDL</td>
<td>Cropland Data Layer</td>
</tr>
<tr>
<td>CITARS</td>
<td>Crop Identification Technology Assessment for Remote Sensing</td>
</tr>
<tr>
<td>EO</td>
<td>Earth Observation</td>
</tr>
<tr>
<td>ESA</td>
<td>European Space Agency</td>
</tr>
<tr>
<td>ETM+</td>
<td>Enhanced Thematic Mapper Plus</td>
</tr>
<tr>
<td>FAO</td>
<td>Food and Agriculture Organization</td>
</tr>
<tr>
<td>FEWS NET</td>
<td>Famine Early Warning Systems Network</td>
</tr>
<tr>
<td>FROM</td>
<td>Fine Resolution Observation and Monitoring</td>
</tr>
<tr>
<td>GCV</td>
<td>Generalized Cross-Validation</td>
</tr>
<tr>
<td>GEO</td>
<td>Group on Earth Observations</td>
</tr>
<tr>
<td>GEOBIA</td>
<td>GEographic Object-Based image Analysis</td>
</tr>
<tr>
<td>GEOGLAM</td>
<td>Group on Earth Observations Global Agriculture Monitoring</td>
</tr>
<tr>
<td>GEOSS</td>
<td>Global Earth Observation System of Systems</td>
</tr>
<tr>
<td>GIAM</td>
<td>Global Irrigated Area Map</td>
</tr>
<tr>
<td>GIEWS</td>
<td>Global Information and Early Warning System</td>
</tr>
<tr>
<td>GIFOV</td>
<td>Ground-projected Instantaneous Field Of View</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographic Information System</td>
</tr>
<tr>
<td>GLC2000</td>
<td>Global Land Cover for the year 2000</td>
</tr>
<tr>
<td>GLCN</td>
<td>Global Land Cover Network</td>
</tr>
<tr>
<td>GMFS</td>
<td>Global Monitoring for Food Security</td>
</tr>
<tr>
<td>GMRCA</td>
<td>Global Map of Rainfed Cropland Area</td>
</tr>
<tr>
<td>GOFC-GOLD</td>
<td>Global Observation Forest and Land Cover Dynamics</td>
</tr>
<tr>
<td>GSI</td>
<td>Ground-projected Sampling Interval</td>
</tr>
<tr>
<td>IFOV</td>
<td>Instantaneous Field Of View</td>
</tr>
<tr>
<td>JECAM</td>
<td>Joint Experiment for Crop Assessment and Monitoring</td>
</tr>
<tr>
<td>LACIE</td>
<td>Large Area Crop Inventory Experiment</td>
</tr>
<tr>
<td>LAI</td>
<td>Leaf Area Index</td>
</tr>
<tr>
<td>LCCS</td>
<td>Land Cover Classification System</td>
</tr>
<tr>
<td>LCML</td>
<td>Land Cover Meta Language</td>
</tr>
<tr>
<td>LOESS</td>
<td>Local Polynomial Regression Fitting</td>
</tr>
<tr>
<td>LPIS</td>
<td>Land Parcel Identification System</td>
</tr>
<tr>
<td>MACCS</td>
<td>Multi-sensor Atmospheric Correction and Cloud Screening</td>
</tr>
<tr>
<td>MARS</td>
<td>Multivariate Adaptive Regression Splines</td>
</tr>
<tr>
<td>MERIS</td>
<td>Medium Resolution Imaging Spectrometer</td>
</tr>
<tr>
<td>MODIS</td>
<td>Moderate Resolution Imaging Spectroradiometer</td>
</tr>
<tr>
<td>MPS</td>
<td>Mean Patch Size</td>
</tr>
<tr>
<td>MS</td>
<td>Multi-spectral</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>MTF</td>
<td>Modulation Transfer Function</td>
</tr>
<tr>
<td>NDVI</td>
<td>Normalized Difference Vegetation Index</td>
</tr>
<tr>
<td>NF</td>
<td>Nyquist frequency</td>
</tr>
<tr>
<td>NIR</td>
<td>Near Infrared</td>
</tr>
<tr>
<td>NLC</td>
<td>National Land Cover (of South Africa)</td>
</tr>
<tr>
<td>PALSAR</td>
<td>Phased Arrayed L-band Synthetic Aperture Radar</td>
</tr>
<tr>
<td>PAN</td>
<td>Panchromatic</td>
</tr>
<tr>
<td>PARA</td>
<td>Perimeter Area RAtio</td>
</tr>
<tr>
<td>PSF</td>
<td>Point Spread Function</td>
</tr>
<tr>
<td>RF</td>
<td>Random Forest</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>RSS</td>
<td>Residual Sum of Squares</td>
</tr>
<tr>
<td>SPAM</td>
<td>Spatial Allocation Model</td>
</tr>
<tr>
<td>SPOT</td>
<td>Satellite Pour l’Observation de la Terre</td>
</tr>
<tr>
<td>SRTM</td>
<td>Shuttle Radar Topography Mission</td>
</tr>
<tr>
<td>STEP</td>
<td>System for Terrestrial Ecosystem Parametrization</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>SWIR</td>
<td>Short wave infrared</td>
</tr>
<tr>
<td>TM</td>
<td>Thematic Mapper</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom (of Great Britain)</td>
</tr>
<tr>
<td>UN</td>
<td>United Nations</td>
</tr>
<tr>
<td>US</td>
<td>United States (of America)</td>
</tr>
<tr>
<td>USDA</td>
<td>United States Department of Agriculture</td>
</tr>
<tr>
<td>VHR</td>
<td>Very High Resolution</td>
</tr>
<tr>
<td>VIIRS</td>
<td>Visible Infrared Imaging Radiometer Suite</td>
</tr>
</tbody>
</table>
List of Symbols

$\beta^F_L$ Classification bias
$\epsilon^F_L$ Classifier bias
$\rho^F_L$ Resolution bias
$\tau$ User-defined accuracy target
$\Delta E$ Increase in resolution-dependent error
$\Delta E_G$ Gain of upscaling
$A$ Foreground area
$AD$ Allocation Disagreement
$B$ Spatial Resolution Bias
$C$ Classifier Error
$C$ Committed area
$CC$ Confidence level Criterion
$\Delta E_L$ Loss of upscaling
$E$ Error
$EDI$ Expected Difference of Information
$ER$ Error Reduction
$ERP$ Equivalent Reference Probability
$F$ Landscape fragmentation
$F$-score$_{N}$ F-scores for the non-cropland class
$F$-score$_{C}$ F-scores for the cropland class
$FB$ Presence of Fallow and Bare fields in the legend
$L$ Resolution
$MPM$ Presence of Managed Pasture and Meadows in the legend
$NUC$Normalized Uncertainty Criterion
$\nu$ Pixel size
$\pi$ Sub-pixel proportion
$\Pi_L(x, y)$ Spatial response map for an instrument with a pixel size of $\nu$
$O$ Omitted area
$OA$ Overall Accuracy
$PA$ Producers’ Accuracy
$PI$ Priority Indicator
$QD$ Quantity Disagreement
$RC$ Resolution adequacy Criterion
$ThC$ Thematic Criterion
$TxC$ Timeliness Criterion
$UA$ Users’ Accuracy
$WC$ Presence of Woody Crops in the legend
Introduction

Feeding the planet in a changing environment

Despite a doubling of the total human population, the past half-century has seen marked growth in food production, leading to a dramatic reduction of the world’s people suffering from hunger and malnutrition (Burch et al., 2007; FAOSTAT, 2016). The associated global food demand was partially met by agricultural expansion and intensification (Alexandratos et al., 2012; Erb et al., 2013; Foley et al., 2007; Lambin and Geist, 2008). Nevertheless, more than one in seven people today still do not derive sufficient protein and energy from their diet, and even more suffer from micronutrient malnourishment (Food and Agriculture Organization of the United Nations, 2009). Given the prospects on the human population growth for the current century (Figure 1), the shift in diet to more meat-based and processed food as well as the development of agrobiofuels, the food supply system is expected to be exposed to ever increasing pressures (Godfray et al., 2010; Rounsevell et al., 2005; Searchinger et al., 2008).

Figure 1: World population and cereal crop production growth for the last fifty years and prediction for 2050. As world population grows, crop production needs to keep up. Source: http://farmingfirst.org.

Besides, climate change has begun to hamper agricultural growth. Increases in the frequency and intensity of extreme events such as drought, heavy rainfall,
flood and high maximum temperatures are already occurring and they are expected to accelerate in many regions (Field and Van Aalst, 2014). These changes affect crop production in several regions, with more adverse impacts than positive ones especially in food unsecured countries (Parry, 2007). In Africa for example, food security issues are likely to increase under an uncertain, changing climate (Brown and Funk, 2008; Wheeler and Von Braun, 2013) whereas in Europe, forecasts predict more yield variability (Trnka et al., 2011). The impact that climate change will have on agricultural yields, and the knock-on effect that this will have on food prices, is a critical question (Challinor, 2011). Both direct observations and models have important roles to play in understanding their relationships (Battisti and Naylor, 2009). In Australia for instance, Hochman et al. (2017) modeled water-limited yield potential of wheat and showed it declined by 27% over a 26-year period from 1990 to 2015. The authors attributed this decline to reduced rainfall and to rising temperatures while the positive effect of elevated atmospheric CO$_2$ concentrations prevented a further 4% loss relative to 1990 yields. Yet, this 27% climate-driven decline in water-limited yield is not fully expressed in actual national yields. This is due to an unprecedented rate of technology-driven gains closing the gap between actual and water-limited potential yields.

Cropping systems would thus need to adapt to those climate-driven changes. A variety of options has already been proposed as having the potential to reduce vulnerability of agricultural systems to risks related to climate change. These options include technological developments (new crop varieties and resource management practices), government programs and insurance (subsidy and support programs), as well as farm production practices (Smit and Skinner, 2002). Farm level adaptations relate to planting and harvest dates, crop rotations, selection of crops and crop varieties for cultivation, water consumption for irrigation, use of fertilizers, and tillage practices (Adams et al., 1998).

Market intelligence

In parallel to the growing food demand, people are becoming increasingly dependent on global and regional markets for their supply. In 1961, imports of cereals in global markets averaged at 26 kilograms per person per year whereas by 2009, they had boomed to 46 kilograms per person per year (FAOSTAT, 2016). Globally, 3.5 billion people live in cities (United Nations, 2014) and are thus highly reliant on markets for the provision of food. Estimates suggest that in the rural areas of developing countries around half of the population is smallholder farmers with up to three hectares of cropland (Morton, 2007). Many in the remaining population would rely on markets to some extent. Households that spend most of their income on food are vulnerable to price increases. Percentage of household expenditure on food is 58.7% in India, 60.6% in Indonesia, 74.8% in Kenya and 81.6% in Rwanda (Smith and Subandoro, 2007) whereas the UK average reaches 11.4% (Department for Environment Food and Rural Affairs, 2013). This suggests that food price movements in
global and regional markets have a significant impact on hunger. In 2007/2008 the world experienced an unusually rapid surge in food prices which plunged between 75 and 130 million additional people into malnutrition (Headey, 2011; Food and of the United Nations, 2008).

Sources of commodity market price volatility and its impacts

The stability of food prices is driven by the balance between food supply and demand. Agricultural prices vary because production and consumption are variable themselves. Economists distinguish between predictable and unpredictable variability; the latter being characterized in terms of shocks (Gilbert and Morgan, 2010). Shocks to production and consumption transmit into price variability. Production can vary either because of variations in planted area or because of yield variations, typically owing to weather conditions. Consumption varies because of changes in incomes, changes in prices and shifts in tastes. It is generally supposed that the most important source of price variability in agriculture is weather shocks to agricultural yields. Nevertheless, demand shocks, in particular income shocks (Gilbert, 2010) and policy shocks (Christiaensen, 2008), may also play an important role.

An unstable agricultural commodity market can lead to food price fluctuations. Market intelligence about global crop production ensures that food supply is consistent with demand. If, for example, Australia has a bumper crop of wheat, US farmers can avoid a wheat glut by not planting wheat. Accurate crop estimates thereby translate into dependable food prices by enabling producers to make wise planting decisions and by equipping agricultural commodity traders with the knowledge they need to set realistic and reasonable prices. In addition, market price adjustments or changes in agricultural supply in a given area often result in price adjustments in other areas (Supit and Van der Goot, 2003). But when the food supply is uncertain or when information is asymmetric between individuals, the market becomes speculative and volatility prevails. This is what happened in 2008 when food prices spiked, pushing millions of people into extreme poverty and hunger (Figure 2).

Direct consumption of grains declines as societies become richer (Gilbert and Morgan, 2010). The consequence is that the impact of high and volatile grain prices concentrates on poorer economies and on the poor, rather than the rich within each economy. In general terms, one could argue that in the richer developed economies, energy price volatility is more problematic than food price volatility. Food price volatility therefore has a greater impact on the developing world, where, depending on the region, maize and rice are the most important food staples. Meat consumption, by contrast, rises as consumers become richer, at least at low and moderate income levels, and this translates into a greater indirect dependence on grains as animal feedstocks. The indirect impact of grain prices volatility through meat prices is therefore most acute at middle levels of income.
Figure 2: 150 years of the price of a metric ton of wheat for the State of Kansas, United States. Data from the US Department of Agriculture.

The Great Grain Robbery

After several years with abnormal climatic conditions in the early 1970s, wheat crops in much of the world failed. Simultaneously, very successful wheat crops in the US had led to large US stocks of wheat. In only six weeks, and before the US realized there was a global wheat shortage, shrewd Soviet traders were able to purchase $750 million worth of US wheat at low subsidized prices. By time the US realized that there was a shortage of wheat on the global market, the Soviets had bought 15 million tons of US wheat compared to 300,000 tons in the past. With the US grain supply suddenly low, wheat prices soared from June 1972 to February 1974. Steep price increases meant that many undeveloped nations could not afford to buy grain, and grain-producing nations were forced to pay a premium for the extra fuel and fertilizer needed to meet demand. This episode has come to be known as the Great Grain Robbery. Coincidentally, the first Landsat satellite was launched the same year as the Great Grain Robbery. This event triggered the monitoring of crop growth with remote sensing satellite by the Foreign Agricultural Service of the US Department of Agriculture.

Food security and the response of the international community

Agriculture mapping and monitoring at the national, regional and, global scales is at the center of modern economic, geostrategic and, humanitarian concerns. It has been on the agenda of several international initiatives for decades. Achieving sustainable global food security for all people was a key priority of the World Food Summit that took place from 13 to 17 November
1996 (Food and Agricultural Organization, 1999). This priority was re-affirmed and re-highlighted during the Millennium Summit of the United Nations (UN) in 2000 (United Nations, 2014), which defined the eradication of extreme poverty and hunger as one of the eight Millennium Development Goals. In 2015, which was the Millennium Development Goals deadline, broad progress was reported on various topics, including on the poverty reduction target. In June 2011, the need to enhance food security and address commodity price volatility was mentioned for the first time in the declaration following the meeting of the G20 Ministries of Agriculture in Paris (G20 Meetings of Agriculture Ministers, 2011). This objective was confirmed in the Final Declaration of the Cannes Summit from November 2011 (G20 Cannes Summit, 2011) and re-emphasized in June 2012 in the declaration of the G20 Mexico summit (G20 Leaders Declaration, 2012). The international food supply crises as regularly observed since 2008 have further shown how agriculture production and its markets are globally connected and exhibit both a wide geographical variation and high fluctuations over time (World Bank, 2015).

It is now widely acknowledged that neither the problem of food insecurity nor the impact of increased market volatility are likely to disappear without policy interventions based on sound scientific evidence (Whitcraft et al., 2015a). Consequently, it is critical to develop better agricultural monitoring capabilities able to provide timely information about crop status, crop area and yield forecasts (Brown, 2012). Earth Observation (EO) data can clearly contribute to this objective as a proven source for transparent, timely, accurate and consistent information on the agricultural productivity at global and regional scales (Justice et al., 2007; Soares et al., 2011).

Since 2011, support for satellite agricultural monitoring has become substantial, with formal institutional support, objectives and timeliness. In the context of the Group on Earth Observations (GEO) supporting the sustainable management of the earth’s resources using satellite remote sensing, the Global Earth Observation System of Systems (GEOSS) identified the development of remote sensing agriculture applications as one of its strategic targets. To improve market information and transparency, the G20 launched the Agricultural Market Information System (AMIS) in Rome on 15 September 2011 and the GEO Global Agricultural Monitoring (GEOGLAM) initiative in Geneva on 22–23 September 2011. AMIS focuses on four grain crops which are the major staple food across the globe and are an input into the production of meat products. The GEOGLAM initiative aims to strengthen the international community’s capacity to produce and disseminate relevant, timely and accurate forecasts of agricultural production at national, regional and global scales using EO (Soares et al., 2011). Its output directly supports AMIS. Anticipating the needs of GEOGLAM, the Joint Experiment for Crop Assessment and Monitoring (JECAM) was launched to enable the global agricultural monitoring community to compare results based on disparate sources of data, using various methods, over a variety of local or regional cropping systems.
Agriculture monitoring with Earth Observation

With the advent of optical satellite remote sensing, *i.e.*, the acquisition and analysis of electromagnetic radiation captured by the sensing modality after reflecting off an area of interest on ground (Prasad *et al.*, 2011), the capacity to monitor agriculture globally has tremendously increased. Crop productivity depends on the physical landscape, *e.g.*, soil type, as well as climatic driving variables and agricultural management practices. All these variables highly fluctuate in space and time. They can thus be effectively captured by satellite imagery to provide an objective and consistent assessments across large areas. Agriculture monitoring with Earth Observation data serves three major purposes: crop production assessment, early warning and, agriculture sustainability (Justice *et al.*, 2007).

First, consistent and timely information about crop acreages and their associated yield are required for assessing the agricultural production. On the one hand, crop inventory information can be extracted by classifying satellite image time series to derive thematic maps depicting the spatial patterns of classes of interest. The synoptic view of agricultural landscapes provided by multi-spectral sensors on-board of satellites, and combined with computer-assisted techniques provides a unique opportunity for large area crop surveys (Bauer, 1985). This was demonstrated by the LACIE (MacDonald and Hall, 1980), AgRISTARS (Bauer *et al.*, 1979), and CITARS (Bizzell *et al.*, 1975) programs, which pioneered the operational use of satellite data for crop recognition. On the other hand, yield can be estimated by (i) combining crop model and satellite-derived biophysical variables (such as Leaf Area Index (Clevers, 1997) or the amount of Fraction of Absorbed Photosynthetically Active Radiation (Lobell *et al.*, 2003)) or (ii) by relating the measured signal directly to yield (Mkhabela *et al.*, 2011). At the national level, crop area and yield estimates are required to inform decision on food storage, distribution and exports as well as to assess the losses along the food supply chain (See *et al.*, 2015). Maps of agricultural land use can also provide critical information to support decision makers and agricultural policies, for instance to verify claims by farmers who apply for public subsidies (Blaes *et al.*, 2005), or assisting in precision farming (see Alganci *et al.* (2013b)).

Second, near real time monitoring of crop growing conditions provides an alternative approach to crop production assessment which is faster and easier to implement. In early warning systems, remotely sensed vegetation condition is compared to a long-term average to indicate if the actual condition is favorable or worse compared to the usual situation (Atzberger, 2013). Such analysis is often complemented with expert evaluations, ground surveys, and other spatially explicit information such as meteo data. Early assessment of production reduction could avert disastrous situation and support strategic measures to meet the demand (Doraiswamy *et al.*, 2004). Besides, accurate and timely assessment of a reduction of the crop production, *e.g.*, due to climatic conditions, natural hazards or pest infestation, can be instrumental for countries...
Agriculture monitoring with Earth Observation

where agriculture is the backbone of the economy. At the regional level, such an approach can be used to untangle the impacts of natural or man-made disasters on food production (Funk and Brown, 2009).

Third, remote sensing is well versed to provide spatial indicators of crop practices and their impact of the environment. Because agriculture sustainability is a time- and space-specific concept, its assessment should be closely linked to the context in which the specific farming system takes place (Zhen and Routray, 2003). The current problem for the assessment of farming systems is how to obtain acceptable indicators spatially and temporally and how to apply and integrate these diverse indicators to address whether a practice is sustainable or not from the development, economics, and environment viewpoints. Such information can be used to design sustainable and informed agriculture policies to ensure both production while preserving minimizing the impact of agriculture on the environment.

Figure 3 provides an overview of the links between remote sensing data and the main components of crop monitoring crop area and crop yield estimation. The diagram illustrates the needs in terms of spatial coverage and temporal frequency for each component. For instance, global coverage of coarse resolution geostationary meteorological satellite (5 km - 1 km) are needed to provide hourly monitoring of weather conditions and rainfall data. Daily global coverage of polar orbiting satellites are needed to provide at 1 km/250 m data for cropland mapping (in some area with low cropland fragmentation) and to the monitor vegetation status.

Figure 3: The three components of crop monitoring (area estimates, yield forecasts, crop status) benefit from the full range of Earth Observation data. Simplified from GEOGLAM’s brochure (2010).
The importance of cropland masks and their current uncertainty

Cropland maps, binary maps where all cropped pixels are included regardless of the crop type, are the first building block of most agricultural applications. Their purpose often depends on the scale at which it is used. Their primary application is cropland masking, that is restricting an analysis to a subset of a region’s pixels rather than using all of them. Cropland masking, has been shown to generally improve crop yield forecasting ability (Atzberger, 2013) and crop type identification (Turker and Arikan, 2005). Besides, early warning systems such as FAO Global Information and Early Warning System (GIEWS), and Famine Early Warning Systems Network (FEWS NET) ingest cropland maps to restrict vegetation monitoring analyses to agricultural land.

Detailed and accurate information on the cropland extent also satisfies the needs of the agriculture sustainability community that seeks to better understand the potential for production of intensification or assess the impact of agriculture on the environment. Therefore, it could support the definition an optimum trade-off or even create win-win situations between agricultural expansion and intensification on the one hand, and sustainable ecosystem service provision on the other hand. For example, it could support the impact assessment of land cover change, for instance, to evaluate the economic and environmental impact of converting cropland to forest (Wang et al., 2007). Downstream studies on cropland use intensity such as Estel et al. (2016) would also greatly benefit from annual cropland maps to derive cropland use indicators. Likewise, cropland information would help to unveil cropland abandonment patterns (Löw et al., 2016a; Kuehmerle et al., 2009; Lakes et al., 2009) and their associated impacts on carbon sequestration (Schierhorn et al., 2013).

A last beneficiary of accurate cropland maps is the climate change community. Since agricultural practices contribute to greenhouse gas emissions, better identification crops and their spatial distribution assumes importance from a climate perspective as well (Beach et al., 2008; Peña-Barraquán et al., 2011), for example for carbon sequestration (Watts et al., 2009), ozone pollution through nitric oxide emission (Rolland et al., 2010), or regional climate change due to agriculture practices (Sampaio et al., 2007). Generally, they isolate the cropland component to study its evolution given different climate scenarios (Roumevill et al., 2005; Ramanjutty et al., 2002).

Cropland is characterized by a diverse mosaic of land use and land cover types that dynamically evolve over spatial and temporal scales in response to climate, management practices, and agricultural policies (Epiphanio et al., 2010; Hostert et al., 2011; Kuehmerle et al., 2011; Atzberger and Rembold, 2012). As a result, detailed regional-scale cropping patterns need to be mapped on a repetitive basis (Wardlow et al., 2007; Vieira et al., 2012), typically every year. To date, such maps remain generally unavailable at fine resolution and with frequent updates. In fact, coarse resolution, satellite-derived global
land cover maps are some of the main sources of current cropland information. However, most of these products were originally designed to characterize land cover more in general or to serve the needs of other communities such as the climate change community. As a result, their quality in capturing cropland is highly variable. For instance, GlobCover estimates 20% more cropland than the Moderate-Resolution Imaging Spectroradiometer (MODIS) globally with large local variations (Fritz et al., 2011a). Not only do those products disagree with one another, they also disagree with cropland statistics from the United Nations Food and Agricultural Organization (FAO). For Africa, the FAO estimates the cropland extent to 319 million hectares in 2010 whereas estimates from MODIS and GlobCover are 277 million hectares and 152 million hectares, respectively. In addition, Leroux et al. (2014) found a strong relationship between the users’ accuracy of the MODIS cropland and the fragmentation of the agricultural landscape: accuracy decreases as crop proportion increases. The authors also found out that the accuracy of the MODIS product is quite low (17%-70%) with strong spatial variability. Recently, the first high resolution global land cover map was released (Gong et al., 2013), but the accuracy of the cropland class remained poor (39% producer’s accuracy and 45% user’s accuracy). In fact, Ramankutty et al. (2008) have attempted to quantify this uncertainty at the global scale, estimating that global cropland extent varies between 1.22 and 1.71 billion hectares, i.e., a 40% difference in this estimation. Several reasons explain the poor accuracy of the cropland class, namely: (1) the heterogeneous and dynamic intrinsic nature of the world’s agrosystems; (2) the spatial structure of the landscape (parcel size) and the crop diversity; (3) differences in crop cycles; (4) differences in cropping practices and calendars within the same class; (5) the spectral similarity with other land cover classes; and (6) the cloud coverage (for optical-derived maps). Specific cropland products such as the M3-Cropland layer of agricultural land in 2000 and the cropland probability layer derived from MODIS have insufficient accuracy in various critical areas or are at a resolution too coarse for assessment and planning purposes (Fritz et al., 2013).

At the national or regional scale, the use of imagery at 30 m or the integration of multiple sensors is more common either for multiple land cover classes or for cropland solely. If moving from global to national or regional mapping makes it easier to take the local conditions into account and to tune the algorithm accordingly, the accuracy of the cropland class does not necessarily improve. Some countries with advanced remote sensing programs have established annual national crop type mapping such as the 30-m US Cropland Data Layer (CDL) or the 30-m Canadian Annual Crop Inventory (ACI). Other efforts, e.g., Africover and the Global Land Cover Network (GLCN) program, have released detailed land cover maps at the country level based on visual interpretation of 30-m spatial resolution images. Increasingly, the availability of cloud computing environments and big data tools enable the processing of large amount of data of data as illustrated by Hansen et al. (2013) for forest change mapping. Spatially explicit information on cropland is also available through digitization of field boundaries. For instance, several European countries maintain a land parcel identification system for farmer’s declarations to manage the redistribution
Introduction

of subsidies from the common agricultural policy (CAP). Nevertheless, these datasets remain largely unavailable publicly; and their use is restricted to CAP activities.

Fifty shades of cropland definition

Cropland and land cover maps provide cropland information at different levels of granularity (Table 1). Some of them group all crop-related features in one class while other provide different mosaic classes, or even cropland proportions. Regardless of the number of cropland classes, there is no consensus on the definition of cropland itself. In GLC-Share (Latham et al., 2014), cropland refers to herbaceous crops, (graminoids or forbs, including herbaceous crops used for hay), woody crops, and multiple or layered crops. The latter class covers situations like i) two layers of different crops (usually woody and herbaceous such as oasis) and ii) the presence of one important layer of natural vegetation (mainly trees) that cover one layer of cultivated crops. In the case of GlobeLand30 (Chen et al., 2015a), the definition includes lands used for agriculture, horticulture and gardens, including paddy fields, irrigated and dry farmland, vegetation and fruit gardens. It should be noted that this thesis targeted the cropland definition adopted by JECAM which reads as follows: “Annual cropland from a remote sensing perspective is a piece of land of minimum 0.25-ha (min. width of 30-m) that is sowed/planted and harvestable at least once within the 12 months after the sowing/planting date. The annual cropland produces an herbaceous cover and is sometimes combined with some tree or woody vegetation.”. There are 3 known exceptions to this definition. The first relates toe sugarcane plantation and cassava which are included in the cropland class although they have a longer vegetation cycle and are not planted yearly. Second, taken individually, small plots such as legumes do not meet the minimum size criteria of the cropland definition. However, when considered as a continuous heterogeneous field, they should be included in the cropland. The third case is the greenhouse crops that cannot be monitored (despite some attempts, e.g., Aguilar et al. (2015)) by remote sensing and are thus excluded from the definition.
Table 1: Diversity of agriculture-related classes for the 10 previous global land cover maps. GlobeLand 30 is based on Landsat images (30 m); GlobCover and CCI are based on MERIS images (300 m); MODIS and GLCmno are based on MODIS images (500 m); GLC-2000 is based on SPOT-Vegetation images (1 km); the JRC MARS map hybrid product at 250 m; and IIASA cropland, GLC-Share and Global GFSAD hybrid products at 1-km (adapted from Lambert et al. (2016)).

<table>
<thead>
<tr>
<th>Product</th>
<th>Cropland</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlobeLand30</td>
<td>Cultivated land</td>
</tr>
<tr>
<td>GFSAD</td>
<td>Cropland, irrigation major</td>
</tr>
<tr>
<td></td>
<td>Cropland, irrigation minor</td>
</tr>
<tr>
<td></td>
<td>Cropland, rainfed</td>
</tr>
<tr>
<td></td>
<td>Cropland rainfed minor fragments</td>
</tr>
<tr>
<td></td>
<td>Cropland rainfed very minor fragments</td>
</tr>
<tr>
<td>GLCmno</td>
<td>Cropland: herbaceous crop</td>
</tr>
<tr>
<td></td>
<td>Cropland: herbaceous crop</td>
</tr>
<tr>
<td></td>
<td>Cropland/other vegetation mosaic</td>
</tr>
<tr>
<td></td>
<td>Paddy field: graminoid crops/non graminoid crop</td>
</tr>
<tr>
<td>GlobCover</td>
<td>Rainfed cropland</td>
</tr>
<tr>
<td></td>
<td>Mosaic cropland (50%-70%)/vegetation (20%-50%)</td>
</tr>
<tr>
<td></td>
<td>Mosaic vegetation (50%-70%)/cropland (20%-50%)</td>
</tr>
<tr>
<td></td>
<td>Cultivated and managed areas</td>
</tr>
<tr>
<td></td>
<td>Post-flooding or irrigated croplands</td>
</tr>
<tr>
<td>GLC-Share</td>
<td>Cropland</td>
</tr>
<tr>
<td>MODIS</td>
<td>Cropland</td>
</tr>
<tr>
<td></td>
<td>Mosaic cropland/natural vegetation</td>
</tr>
<tr>
<td>IIASA</td>
<td>&gt;25% of probability of crop</td>
</tr>
<tr>
<td>CCI Land Cover</td>
<td>Cropland rainfed</td>
</tr>
<tr>
<td></td>
<td>Cropland irrigated or post flooding</td>
</tr>
<tr>
<td></td>
<td>Mosaic cropland (&gt;50%)/natural vegetation</td>
</tr>
<tr>
<td></td>
<td>Mosaic natural vegetation (&gt;50%)/ cropland</td>
</tr>
<tr>
<td>GLC-2000</td>
<td>Cultivated and managed areas</td>
</tr>
<tr>
<td></td>
<td>Mosaic cropland/shrubland or grass cover</td>
</tr>
<tr>
<td></td>
<td>Mosaic cropland/tree cover/natural vegetation</td>
</tr>
<tr>
<td>JRC-MARS cropland</td>
<td>see GlobCover</td>
</tr>
</tbody>
</table>

Scope and objectives

Despite remarkable technological advances in Earth observation systems and major breakthroughs in classification methods, data mining and solutions to handle big data, the use of remote sensing for operational regional to global scale crop mapping and monitoring has not evolved significantly since the early days of LACIE and AgRISTARS. Operational cropland mapping requires methods capable of delivering accurate cropland maps over large areas on a regular basis. Indeed, the dynamic nature of the cropping systems with fallows, crop rotations, and multiple seasons does not allow one-shot map productions. While the ability to derive cropland maps has long been recognized, the requirements for transitioning the research to operations remain challenging to overcome for most. Historically, two main bottlenecks have constrained their operationalization: the scarcity of remote sensing data and the availability of ground truth data. In addition to those technical constraints, those activities are limited by resources (financial or practical). Crop identification relies gen-
Introduction

erally on supervised classification methods that require within-season *in situ* data or human interpretation of spectral signatures, making the classification process resource-intensive, time-consuming, and difficult to repeat over space and time. Three exceptions are the ongoing annual crop mapping activities in the US, Canada and India that have advanced operational crop mapping programs, some of them being largely supported by insurance company information.

With the availability of Sentinel-2A imagery and the recent launch of its twin satellite Sentinel-2B, the remote sensing community is entering a new era of 5-day revisit frequency and 10 m systematic global optical imagery available free of charge. Combined with Landsat-8 data (30 m; 16-day revisit cycle) or other sensors, the revisit frequency can be further increased and benefit applications such as agriculture monitoring that require regular observations (see Figure 4). In a near future, the remote sensing datasets will inevitably become more complex due to ever-finer pixel sizes, increased acquisition frequencies, and enhanced radiometric resolutions. The unprecedented volume of satellite data is a game changer for agriculture mapping because it is removing the deep-rooted constraint of satellite data availability. Yet, a new bottleneck of a different nature is simultaneously emerging: timeliness. The data volume is becoming challenging to store, process and deliver in due time. In parallel, a wealth of datasets has become available over the course of the last decades, including land cover maps, field boundaries, validation datasets, etc. Such existing ancillary data could be of great value to support and inform future cropland mapping strategies.

The necessity to handle complex remote sensing data sets with limited calibration data drives the motivation of this thesis (Figure 4). To answer the challenges resulting from frequent large-area cropland mapping at high resolution, it aims at optimizing the use of remote sensing data and existing ancillary data towards bridging the gap of automated and operational cropland map production. What is meant by large area is that the extent of the geographic coverage ranges from regional to national and global scales. Having clarified the meaning of large area, the overarching objective of the thesis can be formulated as follows:

*To develop methods for yearly cropland mapping with high resolution satellite image time series towards a regularly updated global map and improved area estimates.*

As this objective can be achieved by leveraging ancillary data or by optimizing the use of remote sensing data (Figure 4), it is more specifically structured by attempting to answer the two following research questions.

1. **How to take advantage of multiple ancillary data sources of variable quality to support the continuous update of cropland maps?**
   This first question seeks to capitalize on the wealth of data sets that have
been produced during the last decades thanks to the takeoff of remote sensing technologies and Geographic Information Systems (GIS). Regardless of their initial purpose, these data provide opportunities to better inform methodological and strategic choices on where and how to focus future cropland mapping efforts. One could therefore capitalize on these data to i) highlight priority areas for cropland mapping, ii) improve the current depiction of the global cropland extent and iii) calibrate new classifiers in the absence of field data or to alleviate field data collection.

2. How to optimize the remote sensing data inputs in terms of spatial resolution and spectral-temporal features?
Satellite image time series have become increasingly detailed due to ever-finer spatial, temporal, radiometric and spectral resolutions. The traditional trade-off between spatial resolution and temporal frequency has been progressively removed thanks to technical progress and to the launch of satellite constellations such as PlanetScope or RapidEye*. This evolution was mostly driven by applications needing improved spatial details and a high-frequency revisit cycle such as agriculture. However, complexity grew simultaneously as the datasets became richer. Two sources of complexity are the increase temporal and spatial resolutions. The motivation is thus two-fold: i) find the appropriate spatial resolution for specific applications and ii) identify features that guarantee a high class separability for classification. This will in turn have consequences on the accuracy of the methods as well as on the timeliness of the product delivery to the users (reduced volume to be processed).

![Figure 4: Requirements of operational cropland mapping, its related constraints for implementation and the strategies developed in this thesis to alleviate them.](https://www.planet.com/)
Outline of the thesis

The core of this thesis (chapters 1 to 5) is based on scientific papers that are at different levels of progression towards publication in international peer-reviewed journals. Chapters 1, 2 and 5 have already been published while chapters 3 is accepted pending minor revisions and chapter 4 has recently been submitted (see page 183 for details). Each chapter is briefly introduced below by stating its research objectives and by outlining its links with the other chapters and with the research questions as stated before. At the beginning of each chapter, a short collection of bullet points highlights the core findings. Figure 5 illustrates the aforementioned research questions and optimization strategies are addressed in this thesis.

![Figure 5: Overall structure of the thesis highlighting the relationshipS between chapters and research questions.](image)

**Chapter 1** links directly to the first research question by proposing a multi-criteria approach to inventory the availability of spatial cropland information to identify priority areas for cropland mapping. It identifies where to focus the mapping efforts in priority so that their added value is maximized. In addition, it capitalizes on the multi-criteria analysis to generate a new hybrid global cropland map, hereafter referred to as the Unified Cropland Layer, by combining the best available maps.

**Chapter 2** introduces a new methodology for automated cropland mapping when *in situ* data are not available. Learning from the local scale to cope with the diversity of the world’s agrosystems, it seeks to propose a generic method that can deliver sufficiently accurate results over large areas. The rationale is to identify reliable pixels in existing maps to serve as calibration data in the absence of ground truth data. The latter could typically be identified by the multi-criteria analysis presented in Chapter 1. It also introduces a set of new spectral-temporal features that are based on the expected temporal trajectory of crop development to reduce within-class variability. Temporal generalization potential of the method is also investigated.
Chapter 3 elaborates on the method proposed in Chapter 2. New methodological developments are introduced to cope with large-scale specificities such as natural vegetation gradients and with heterogeneous crop management practices. Its potential is demonstrated over South Africa where a wall-to-wall validation data set is available. This validation dataset allows to gain insight on the drivers of the accuracy.

Chapter 4 is an attempt to optimize the spatial resolution of satellite data in order to reduce the burden of data volume. Indeed, the unprecedented volume of high resolution data that are becoming available poses challenges to users in terms of problem complexity, computational resources and processing time, beckoning the increasingly relevant question: at which resolution should this imagery be processed? Based on very high resolution subsets of field boundaries, the reduction of resolution-dependent error is assessed for three scenarios: with and without explicitly modeling the sensors' point spread functions, and upscaling. A framework to identify an “optimized” resolution is also introduced.

Chapter 5 seeks to identify the maximum tolerable classifier bias for pixel counting is competitive for area estimation. To that aim, it models the relationship between the resolution bias, the spatial resolution and the landscape fragmentation. Given a user-defined accuracy target on the area estimate, this model allows to anticipate either the spatial resolution that should be used in a given landscape or the applicability domain of pixel counting for a certain resolution.

The main findings and future perspectives are summarized and discussed in the conclusion of this document. But before getting down to the nitty gritty, the next section lays down the physical background, methods and concepts on which this thesis builds.

Review of concepts and methods relevant for crop-land mapping research

The purpose of this section is to review the literature and summarize some key concepts relevant to the research objectives and frequently mobilized throughout the thesis. The first subsection explains the concepts behind scale and spatial resolution in remote sensing. The second subsection describes some aspects on satellite image classification for crop mapping. It also details the features available for class discrimination and the methods to assess the accuracy of a thematic map. Finally, the third subsection introduces landscape fragmentation and its quantification.

The spatial resolution of remotely sensed images

In remote sensing imagery, scale can be defined as the number and size of the spatial sampling units used to partition a geographic area (Lam and Quattrochi,
Introduction

However, a distinction must be made between the physical sampling of the observations by the instrument and the spacing of the grid in which the data is provided. Here, we refer to the on-ground distance between the centers of two observations as the Ground-projected Sampling Interval (GSI; see Figure 6 and Table 2 for definitions of the main scale-related terms).

Figure 6: Schematic representation of the different factors involved in the image acquisition process: (a) the footprint of the observation is different that the footprint of grid cell it is stored in, (b) the blurring is characterized by the Point Spread Function and (c) the corresponding Modulation Transfer Function.

The spacing between the ground projection of two pixels is referred to as pixel size (and denoted $\nu$). These can differ even on a single scan line on whiskbroom instruments such as the Moderate Resolution Imaging Spectroradiometer (MODIS), where increasing viewing angle combined with Earth’s curvature lead to larger GSI at the edge of the swath than at nadir, while the pixel size of the delivered image remains the same. Another common misconception is that the shape of the observation footprint is the same as the rectangular ground projection of the pixel (Cracknell, 1998). Instead, a substantial portion of the measured radiance originates from surrounding areas (Forster and Best, 1994; Townshend, 1981). At every ground sampling interval, a detector measures the incoming radiance within its instantaneous field of view (IFOV) during a specific time interval. The IFOV is an angular measure and its ground projection is known as the GIFOV. The width of the GIFOV does not exactly match the GSI because of several factors such as the optics of the instrument, the electronics of the detector, and the image motion (Markham and Barker, 1986; Schowengerdt, 2006; Kavzoglu, 2004). Thus the image of the scene viewed by the sensor is not a completely faithful reproduction of the real ground features. Small details are blurred relative to larger features and this blurring effect can be characterized by the net sensor Point Spread Function (PSF). Alternatively, the PSF can be expressed by the Modulation Transfer Function (MTF) which is its equivalent in...
the frequency domain. Several studies investigated the impact of the PSF/MTF on land cover classification (Huang et al., 2002b), sub-pixel landscape feature detection (Radoux et al., 2016), sub-pixel class proportion estimation (Huang et al., 2002b; Townshend et al., 2000).

### Table 2: Definition of scale-related terms.

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Acronym</th>
<th>Definition</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground-projected Sample Interval</td>
<td>GSI</td>
<td>The spacing between two observations on the ground</td>
<td>Schowengerdt (2007)</td>
</tr>
<tr>
<td>Pixel size</td>
<td>$\nu$</td>
<td>The spacing between the ground projection of two pixels.</td>
<td></td>
</tr>
<tr>
<td>Field of View (Ground)</td>
<td>FOV</td>
<td>The angular extent of data acquisition cross-track</td>
<td>Schowengerdt (2007)</td>
</tr>
<tr>
<td>Instantaneous Field Of View</td>
<td>(G)IFOV</td>
<td>The geometric projection of a single detector on the ground</td>
<td>Duveiller and Defourny (2010)</td>
</tr>
<tr>
<td>PSF</td>
<td></td>
<td>The image of a point source intensity as a function of position (x, y).</td>
<td>Cracknell (1998)</td>
</tr>
<tr>
<td>Modulation Transfer Function</td>
<td>MTF</td>
<td>describes how the true contrast between high contrast bars is progressively reduced in the image as their width, Fourier transform of the PSF.</td>
<td>Markham (1985)</td>
</tr>
<tr>
<td>Nyquist frequency</td>
<td>NF</td>
<td>The Nyquist frequency is defined as the highest sinusoidal frequency that can be represented by a sampled signal and is equal to one half the sampling rate of the system.</td>
<td>Kohm (2004)</td>
</tr>
<tr>
<td>Upscaling</td>
<td></td>
<td>The process of taking information at smaller scales to derive processes at larger scales</td>
<td>Hay et al. (1997)</td>
</tr>
</tbody>
</table>

As the two last chapters of the thesis deal with spatial resolution and changes of spatial resolution, Table 3 illustrates how a decrease in spatial resolution reduces the ability to distinguish image objects for six agricultural landscapes around the world.

**Spectral and temporal measurement of vegetation by satellite**

Single-date crop recognition is based on the interaction of solar radiation with the crop elements. Five physical factors condition the spectral reflectance of crops: 1) leaf optical properties, 2) soil (background) reflectance, 3) canopy geometry — in particular leaf area index and leaf angle distribution, 4) solar illumination and view angles, and 5) atmospheric transmittance (Bauer, 1985). In this introduction, the emphasis is on the two first factors as deeper knowledge about these two in particular may facilitate the understanding of the approaches developed in the next chapters. The reader is referred to the review of Bauer (1985) for further information on the other factors.
**Table 3:** Illustration of the effect of spatial resolution on six agricultural landscapes of 5x5 km². The images are false color composite (near infrared, red, green) based on Sentinel-2 data.

<table>
<thead>
<tr>
<th></th>
<th>10 m</th>
<th>30 m</th>
<th>100 m</th>
<th>250 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td>Argentina</td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
<tr>
<td>Brazil</td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
<tr>
<td>Russia</td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
<td><img src="image16.png" alt="Image" /></td>
</tr>
<tr>
<td>South Africa</td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
</tr>
<tr>
<td>Australia</td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
</tr>
</tbody>
</table>

**Optical properties of the leaves**

At the vegetative stage, leaves are the primary scattering elements. Plant leaves reflect, absorb, and transmit incident solar radiation in a way that is characterized uniquely by pigmented cells containing water (Gates *et al.*, 2020).
In the range of 0.4 to 2.5 µm, three regions associated with different reflectance, transmittance, and absorption can be identified: the visible, the Near InfraRed (NIR), and the Short Wave InfraRed (SWIR) portions of the spectrum.

The reflectance of leaves is relatively low in the visible portion of the spectrum: about 2-3% of the incident radiation or about half of the total reflectance is reflected—mostly in a specular way—from the surface of leaves (Woolley, 1971). The low reflectance and transmittance results from the absorption by leaf pigments. Chlorophyll absorbs most of the incident energy in the blue and red wavelength bands centered at around 0.45 and 0.67 µm and generally masks the presence of other pigments (carotenes, xanthophylls, anthocyanins) except during senescence. Bands in the blue region of the spectrum are sensitive to carotenoid pigments (Tucker, 1977; Blackburn, 1998), loss of chlorophyll, browning, ripening, senescence, and soil background effects (Thenkabail et al., 1999). However, the use of blue range remains questionable due to atmospheric effects (Thenkabail et al., 2002). The green band is mostly sensitive to pigment content (Nichol et al., 2000). Absorption in the red band varies significantly due to changes in biomass, leaf area index, soil background, cultivars, canopy structure, nitrogen, moisture and stresses (Elvidge and Chen, 1995; Carter, 1994; Blackburn, 1998). It is also the spectral band with the chlorophyll absorption maxima and it offers the strongest contrast between crop and soil (Thenkabail et al., 1999).

In the NIR region, there is a significant increase in reflectance. Leaves typically reflect 40-50% and absorb less than 5% of the incident energy at these wavelengths. As the internal structure is specific to species, differences in reflectance between crops are generally more marked in the near-infrared than in the visible wavelengths. The NIR is also sensitive to the total chlorophyll.

In the SWIR region (1.3-2.6 µm), the reflectance of green vegetation is dominated by strong water absorption bands which are located around 1.4, 1.9 and 2.7 µm. Nonetheless the leaf water content has also a strong influence on these absorption bands. In fact, the leaf reflectance in this part of the spectrum is inversely related to the total amount of water present in the leaf (Tucker, 1980).

**Soil background properties**

Soil reflectance across the spectrum generally increases with increasing wavelength. Its amplitude is the resultant of the contribution of iron oxide, organic matter, moisture, texture, and surface roughness. It is a major factor influencing the reflectance of vegetation canopies (Colwell, 1974) and canopy reflectance varies with the soil reflectance, amount of soil exposed, and the optical characteristics of vegetation sensor. Five general reflectance curves were identified by Stoner and Baumgardner (1981) based on the absorption of ferric iron bands at 0.7 and 0.9 µm, as well as organic matter content and texture.
Spectral indices

In the field of remote sensing applications, scientists have developed vegetation indices (VI) for qualitatively and quantitatively evaluating vegetative covers using spectral measurements. VIs are spectral transformations of two or more bands designed to enhance the contribution of vegetation properties and allow reliable spatial and temporal inter-comparisons of terrestrial photosynthetic activity and canopy structural variations (Huete et al., 2002). Besides, they mitigate atmospheric and soil effects. As a simple transformation of spectral bands, they are computed directly without any bias or assumptions regarding land cover class, soil type, or climatic conditions. The Normalized Difference Vegetation Index (NDVI; Tucker (1979)) is certainly the most popular. The NDVI relies on the contrast between absorption of red light by leaf pigments and strong scattering of near-infrared radiation by foliage. Consequently, changes in NDVI are linked to changes in various vegetation biophysical properties including leaf area, green biomass and the fraction of absorbed photosynthetically active radiation.

Texture information

In addition to the reflectance that is measured, local variations of the reflectance or texture can also be very informative from a classification point of view. Texture is an important element for pattern recognition and interpretation in the human visual system. In a digital image, it represents the visual impression of smoothness or coarseness produced by the uniformity or variability of image color or tone (Myint, 2001). Texture processing algorithms are usually divided into three major categories: structural, spectral and statistical (Chica-Olmo and Abarca-Hernandez, 2000). However, descriptive terms for texture patterns presented in remote sensing imagery are yet to be established. To illustrate the additional descriptive information of texture, a basic texture descriptor, i.e., the contrast within a 7x7 pixel moving window, was derived from the NDVI of the sites presented in Table 3 at 30 and 250 m. Dark areas correspond to areas with low texture.

Temporal analysis of the crop reflectance

Time is an additional criterion to consider when looking at the spectral separability of crops as the variation of the spectral reflectance during the season. Reflectance, transmittance, and absorption of light by the leaves depend on the concentration of pigments and water, along with the internal cell structure of each species, which in turn depends on the development stage, senescence and stress. Figure 7 presents the typical reflectance trajectory of an herbaceous crop pixel over time in the red an NIR channels. At the beginning of the season, the reflectance is that of bare soil as has just been sown. As leaves grow and mature, their visible reflectance decreases whereas the near-infrared reflectance increases. This is related to the greater number of intercellular air spaces in the mesophyll of mature leaves. Senescence produces the opposite effect of maturation: visible reflectance increases due to the loss of
Table 4: Illustration of texture information, in this case contrast, extracted at 30 and 100 m across six agricultural landscapes.

<table>
<thead>
<tr>
<th>Country</th>
<th>NDVI 30 m</th>
<th>Texture 30 m</th>
<th>NDVI 100 m</th>
<th>Texture 100 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td>Argentina</td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
<tr>
<td>Brazil</td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
<tr>
<td>Russia</td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
<td><img src="image16.png" alt="Image" /></td>
</tr>
<tr>
<td>South Africa</td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
</tr>
<tr>
<td>Australia</td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
</tr>
</tbody>
</table>

chlorophyll and infrared reflectance decreases. Therefore, it becomes evident that the timing of image acquisition will significantly affect the class separability.

It has been shown that a denser temporal sampling of the growing season, e.g., acquisition at sowing and harvesting dates of different crops, usually boosts classification accuracy compared to mono-temporal data sets (Conrad et al.,
**Introduction**

Figure 7: a) Distribution of reflectance values in a remote sensing image in the red and near-infrared regions of the electromagnetic spectrum are found in the gray shaded area. The greater the amount of photosynthetically active vegetation present, the greater the near-infrared reflectance and the lower the red reflectance. This condition moves a pixel’s spectral location in a perpendicular direction away from the soil line. b) The migration of a single vegetated agricultural pixel in red and near-infrared multi-spectral space during a growing season is shown. After the crop emerges, it departs from the soil line, eventually reaching complete canopy closure. After harvesting, the pixel will be found on the soil line, but perhaps in a drier soil condition. Adapted from Jensen (2009).

2011; Wardlow et al., 2007; Wardlow and Egbert, 2010; Murakami et al., 2001; Serra and Pons, 2008; Simonneaux et al., 2008b; Brown et al., 2013). Besides, there is also evidence that employing time series of multi-spectral, or multi-source images (optical and radar) bare great potential in crop mapping (Blaes et al., 2005; McNairn et al., 2009b; van Oort et al., 2004).

**Characterizing the fragmentation of agricultural landscapes**

Forman (2014) defined landscape fragmentation as “the breaking up of a habitat, ecosystem, or land-use type into smaller parcels, which is not directly suitable to quantify fragmentation.” Fragmentation is an umbrella term that comprises many different patch aspects, such as the number and typical shape, the inter-patch distance, pattern, connectivity, and patch configuration. In a categorical map, fragmentation can be described as the spatial heterogeneity, or the spatial composition and arrangement of foreground objects in an image. In addition, fragmentation accounts for the number of objects and the distance between them, hence addressing foreground and background characteristics at the same time. To illustrate this concept, Figure 8 presents the landscapes introduced in Table 3 ranked by increasing fragmentation.

Due to its holistic nature, the description of fragmentation may be rather complex when accounting for and especially when trying to summarize its individual components. Therefore many quantitative measures for pattern (McGarigal...
et al., 2002; Soille and Vogt, 2009; Wickham et al., 2010) and connectivity (Saura et al., 2011) have been proposed. Let us consider four synthetic cases with two classes of identical areas but distinct spatial patterns (Table 5). Let us characterize the fragmentation of the red class within these landscapes. In addition to the total area (A) and perimeter (P), let us also consider three fragmentation indices:

the mean patch size (MPS): provides a measure of central tendency of the mean patch area across the entire landscape.

the perimeter-area ratio (PARA): simple ratio of patch perimeter to area, in which patch shape is confounded with patch size; holding shape constant, an increase in patch size will cause a decrease in the perimeter-area ratio.

the Matheron Index (MI): similar to core-perimeter indices it is computed as follows $MI = \frac{P_{class}}{\sqrt{A_{class}} \times \sqrt{A_{total}}}$.

The Matheron Index deserves some mention because it has intensively been used by the remote sensing community (Mayaux and Lambin, 1995, 1997; Eva and Lambin, 1998; Vanclay et al., 1999; Imbernon and Branthomme, 2001; Mas et al., 2004; Vintrou et al., 2012a; Leroux et al., 2014; Lambert et al., 2016) and will be used frequently in the thesis. Despite having the same area, one can now compare different landscapes by quantitative measures describing their specific patterns and characteristics.

Agriculture mapping from satellite images time series

Image classification techniques

The principal method to identify and map crops is image classification with one of the many available classifier concepts (Mather and Tso, 2016). In remote sensing, the problem of multi-dimensional classification is complex and depends on many parameters. More specifically, it requires: (1) to determine which
Table 5: Landscape metrics of four synthetic landscapes. The metrics are computed for the red class which is constant in all landscapes.

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>P</td>
<td>20</td>
<td>24</td>
<td>32</td>
<td>22</td>
</tr>
<tr>
<td>MPS</td>
<td>32</td>
<td>16</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>PARA</td>
<td>0.66</td>
<td>0.75</td>
<td>1</td>
<td>0.69</td>
</tr>
<tr>
<td>MI</td>
<td>0.44</td>
<td>0.53</td>
<td>0.71</td>
<td>0.49</td>
</tr>
</tbody>
</table>

classifier algorithms can handle the amount and the variability of data; (2) to evaluate the stability of classifier parameters; (3) to select the best feature set used as input data to find the good trade-off between classification accuracy and computational time; and (4) to establish the classifier accuracy over large areas (Pelletier et al., 2016).

The literature on crop mapping is replete with classification approaches across scales such as (a) maximum likelihood (Abou El-Magd and Tanton, 2003), (b) nearest neighbors (Samaniego and Schulz, 2009), (c) logistic regression (Peña et al., 2014), (d) expert-based decision rules (Conrad et al., 2010), (e) spectral angle mapper (Alganci et al., 2013b), (f) tasseled-cap transformation (Crist and Cicone, 1984), (g) decision tree algorithms (Wardlow et al., 2007; Pittman et al., 2010), (h) random forest (Duro et al., 2012; Long et al., 2013), (i) neural network methods (Liu et al., 2005b), (j) support vector machine (Alganci et al., 2013b), (k) spectral unmixing techniques (Yang et al., 2007a), (l) and biologically inspired algorithm (Omkar et al., 2008) that have been applied either pixel-based or object-based (Duro et al., 2012; Long et al., 2013). While most of these supervised methods generally outperform the unsupervised ones (Sharma et al., 2013), they rely extensively on within season in situ data or on human interpretation of spectral signatures, making the classification process resource-intensive, time-consuming, and difficult to repeat over space and time (Zhong et al., 2014). The first research question seeks to propose an alternative to this issue.

Accuracy assessment

Accuracy assessment based on the confusion matrix

Accuracy assessment quantifies data quality so that map users may evaluate the utility of a thematic map for their intended applications. There are three basic components in the validation of remote sensing classifications (Stehman and Czaplewski, 1998): 1) the sampling design used to select the reference sample;
Accuracy assessment based on the confusion matrix

2) the response design used to obtain the reference land cover classification for each sampling unit; and 3) the estimation and analysis procedures. Briefly, an accuracy assessment starts with the definition of the target population, i.e., the area to be mapped defined by individual units (pixels or polygons). A sample of these units is selected from this population for accuracy assessment. The reference or “true” classification is obtained for each sampling unit based on visual interpretation, ground visits, or a combination of the two. The response design characterizes the process of collecting this reference information. Then, the classification is compared to the reference data, and the extent to which these two classifications agree is defined as map accuracy. In the following, the emphasis is on the estimation of the accuracy. The reader is referred to Stehman and Czaplewski (1998); Olofsson et al. (2012) for more details on the sampling schemes.

The common way of deriving the map accuracy is the analysis of the confusion matrix. A confusion matrix is a square matrix that compares a classified map to reference data. Diagonal values represent the agreement between the reference and the classification while non-diagonal values represent the errors (Table 6 presents a confusion matrix for four classes).

Table 6: Confusion matrix of four classes with cell entries \( p_{ij} \) expressed in number of pixels/objects.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Map</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class 1</td>
<td>( p_{11} )</td>
<td>( p_{12} )</td>
<td>( p_{13} )</td>
<td>( p_{14} )</td>
<td>( p_1 )</td>
</tr>
<tr>
<td>Class 2</td>
<td>( p_{21} )</td>
<td>( p_{22} )</td>
<td>( p_{23} )</td>
<td>( p_{24} )</td>
<td>( p_2 )</td>
</tr>
<tr>
<td>Class 3</td>
<td>( p_{31} )</td>
<td>( p_{32} )</td>
<td>( p_{33} )</td>
<td>( p_{34} )</td>
<td>( p_3 )</td>
</tr>
<tr>
<td>Class 4</td>
<td>( p_{41} )</td>
<td>( p_{42} )</td>
<td>( p_{43} )</td>
<td>( p_{44} )</td>
<td>( p_4 )</td>
</tr>
<tr>
<td>Total</td>
<td>( p_{1} )</td>
<td>( p_{2} )</td>
<td>( p_{3} )</td>
<td>( p_{4} )</td>
<td>1</td>
</tr>
</tbody>
</table>

Accuracy parameters derived from a confusion matrix with \( q \) classes include the overall accuracy (probability of a randomly selected pixel to be classified accurately):

\[
OA = \sum_{j=1}^{q} p_{jj} \tag{1}
\]

the users’ accuracy of class \( i \) (the proportion of pixels mapped as class \( i \) that has reference class \( i \)):

\[
UA_i = \frac{p_{ii}}{p_i} \tag{2}
\]

or its complementary measure, the commission error \( (CE_i = 1 - UA_i) \), the producers’ accuracy of class \( j \) (the proportion of pixels of reference class \( j \) that has class \( j \)):

\[
PA_j = \frac{p_{jj}}{p_j} \tag{3}
\]

or its complementary measure, the omission error \( (OE_j = 1 - PA_j) \). Another class-specific accuracy indicator, the F-score, has recently gained attention in
the remote sensing community. The F-score is an synthesis accuracy indicator to be interpreted as the harmonic mean of the producers' accuracy and the users' accuracy:

\[
F_{\text{score}} = \frac{2 \times U_{A_i} \times P_{A_i}}{U_{A_i} + P_{A_i}}
\]

(4)

Pareto boundary analysis

Confusion matrices do not consider contextual influence of mixed pixels on the product accuracy (Boschetti et al., 2004). Besides, when validating a coarse resolution product with a high resolution reference map, the assumption of equal spatial resolution between the reference and the product is violated. The Pareto boundary method is an alternative to deal with these shortcomings. The number of low resolution pixels covering multiple classes is closely linked to the ground features (reference data) and is a function of their shape, size and spatial patterns (Eva and Lambin, 1998; Mayaux and Lambin, 1995). The difference in spatial resolution between high and low resolution data is referred to as the low-resolution bias (Boschetti et al., 2004). The resolution bias sets down the omission and commission errors as conflicting objectives. Effectively, residual error after classification cannot be avoided. Any attempt to reduce the commission errors will inevitably lead to an increase of the omission errors and conversely.

Therefore in the OE/CE bi-dimensional space, a region of unreachable accuracy limited by the Pareto boundary separates the errors due to the spatial resolution and the method. The Pareto boundary determines the maximum user and producer's accuracy values that could be attained jointly and represents such a lower limit as a boundary. To generate the Pareto boundary, the high resolution binary reference map is degraded to the low resolution pixel size (Figure 9). Each new pixel value corresponds to the percentage of high resolution pixels of the class of interest. A set of low-resolution product is obtained by thresholding the low resolution reference map. For each threshold defining the percentage for which a pixel is considered as vegetation, the pair of efficient error rates OE/CE is computed. The line joining all these points defines the Pareto boundary of a specific high resolution reference to a defined low-resolution pixel size. The distance between the product and the boundary indicates the performance of the method. The area under the efficient solution curve indicates the accuracy of the detection algorithm.

Synthesis of crop classification performance across landscapes

Based on the literature review, it occurs that the performance of a classifier in an agrosystem is mainly dictated:

- the choice of the classifier and its calibration parameters;
- the quantity and quality of the calibration data;
Figure 9: The procedure for generating a discrete set of points belonging to the Pareto Boundary, starting from the high-resolution map, for a desired low spatial resolution. (a) A low-resolution grid is overlaid on the high-resolution reference map. (b) The percentage of class $\omega_1$ is computed for each cell. (c) A set of low-resolution products is generated by thresholding the percentage of class $\omega_1$; threshold $t$ varies in the interval $(0, 1]$. These maps are efficient solutions according to Pareto’s criterion. The map generated with $t = 100\%$ will have no commission (no mixed pixels included) but large omission, and the map with $t = 1\%$ will have no omission (all the mixed pixels are included) but large commission; in general, the higher $t$, the lower the commission and the higher the omission. (d) The confusion matrix is produced for each one of these maps. Omission error and commission error are derived and plotted in the omission error/commission error space. Source: Boschetti et al. (2004)

- the joint effect of spatial resolution and landscape fragmentation (number of mixed pixels);
- the density of the time series at hand.

Crop mapping and land cover mapping in general suffer from a trade-off between spatial and temporal frequency. Inherent limitations of remote sensing systems cannot provide simultaneously observations frequently over large areas and at a fine resolution. Complex landscapes typical of smallholder farming
systems can only be accurately described with high or very high spatial resolution while intensive agrosystems are generally more simple and easy to classify. For instance, Wardlow and Egbert (2008) showed that MODIS 250-m time-series could provide accurate (94%) cropland maps in areas with an average field size of 32 ha or larger and Waldner et al. (2016) concluded that MODIS data had the ability to resolve fields down to 20 ha. In the fragmented landscapes of Mali, Vintrou et al. (2012a) reported that a crop patch needs to be eight times larger than 25 ha in order to be detected by MODIS. For complex landscapes, methods could benefit from the addition of high or very high spatial resolution imagery (Vaudour et al., 2015) or from any other satellite-derived environmental information, such as elevation data (Sesnie et al., 2008). This is well illustrated by Delrue et al. (2013) who evaluated the potential of discriminating crops in an area with small-scale farming in central Ethiopia. A classification of high resolution (30 m) images, yielding good results for commercial farming, could not deal with mixed pixels due to the small parcels. To effectively address the challenges associated with mapping tropical smallholder agricultural systems at a regional scale, Lebourgeois et al. (2017) combined 10-m data with 0.5-m data and ancillary data. Similar difficulties were reported by Inglada et al. (2015) who benchmarked the classification approaches across a wide range of landscapes. For sites of intensive farming, the overall accuracy of classifications of main crop types were always higher than 80%, whereas for sites characterized by smallholder agriculture, the overall accuracies were around 50%.
Chapter 1

Mapping priorities to focus cropland mapping activities: fitness assessment of existing global, regional and national cropland maps

Highlights

• A multi-criteria analysis was designed at the country level to identify priority areas for cropland mapping.

• Priority regions such as Western Africa, Ethiopia and Madagascar and South East Asia were identified for the remote sensing community to focus its efforts on.

• A unified cropland layer at 250 m for the year 2014 was produced combining the fittest national-scale products.

• The accuracy of the unified cropland layer ranged from 82% to 94% depending on the validation dataset.

• A large share of the discrepancies with other hybrid products can be attributed to differences in the legend definition.

Abstract. Timely and accurate information on the global cropland extent is critical for applications in the field of food security, agricultural monitoring, water management, land-use change modeling and earth system modeling. On the one hand, it gives detailed location information on where to analyze satellite image time series to assess crop condition. On the other hand, it isolates the agriculture component to focus food security monitoring on cropped areas and assess the potential impacts of climate change on agriculture. The cropland class is often

poorly captured in global land cover products due to its dynamic nature and the large variety of agro-systems. The overall objective was to evaluate the current availability of cropland data sets in order to propose a strategic planning and effort distribution for future cropland mapping activities and therefore maximize their impact. Following a very comprehensive identification and collection of national to global land cover maps, a multi-criteria analysis was designed at the country level to identify the priority areas for cropland mapping. As a result, the analysis highlighted priority regions such as Western Africa, Ethiopia and Madagascar and South East Asia for the remote sensing community to focus its efforts on. A unified cropland layer at 250-m for the year 2014 was produced combining the fittest products. It was assessed using global validation data sets and yields an overall accuracy ranging from 82 to 94%. Masking cropland areas with a global forest map reduced the commission errors from 46% to 26%. Compared to the GLC-Share and the IIASA-IFPRI cropland maps, significant spatial disagreements were found which might be attributed to discrepancies in the cropland definition. This advocates for a shared definition of cropland as well as a global validation data sets relevant for the agriculture class in order to systematically assess existing and future cropland maps.

1.1 Introduction

Mapping the global cropland extent is of paramount importance for food security. Indeed, accurate and reliable information on-cropland and on the location of major crop types is required to make future policy, investment, and logistical decisions (Fritz et al., 2012), as well as production monitoring (Justice et al., 2007). Timely cropland information directly feed early warning systems such as GIEWS, GMFS, and FEWS NET (Hannerz and Lotsch, 2008; Vancutsem et al., 2012). In addition, other disciplines benefit from this information, e.g. environmental climate change (Lobell et al., 2006). In agriculture monitoring as well as climate modeling, cropland maps mask out non-cropland areas to analyze satellite image time series analysis for crop condition monitoring and to investigate how cropland responds to different climatic projections.

Space borne Earth Observation provides opportunities for global cropland monitoring in a spatially explicit, economic, efficient, and objective fashion (Yu et al., 2013b). Over the last forty years, numerous initiatives aimed at deriving cropland from satellite images either specifically (cropland vs. non-cropland) or as a land cover class. A large diversity of mapping strategies ranging from the local to the global scale and associated with various degrees of accuracy can be found in the literature. Cropland is often depicted according to a land cover typology that focuses mainly on the natural vegetation types and is often included in mosaic or mixed classes making them difficult to use for agricultural applications (neither as agricultural mask, nor as a source of information for cropped area). This is typical of global land cover products, such as GLC2000 (Bartholomé and Belward, 2005), GlobCover 2005/2009 (Arino et al., 2008), MODIS Land Cover (Friedl et al., 2002), which are not specifically
targeting the agriculture component of the landscape. This remains valid for the most recent European Space Agency (ESA) Climate Change Initiative (CCI) Land Cover products (Defourny et al., 2012a). Recently, the first high resolution global land cover map was released (Gong et al., 2013) but the accuracy of the cropland class remains poor (39% of producer’s accuracy and 45% of user’s accuracy). Several reasons explain the poor accuracy of the cropland class namely: a) the heterogeneous and dynamic intrinsic nature of the world’s agrosystems, b) the spatial structure of the landscape (parcel size) and the crop diversity, c) differences in crop cycles, d) differences in cropping practices and calendars within the same class, e) the spectral similarity with other land cover classes and f) the cloud coverage (for optical-derived maps). Moreover, when analyzing the consistency between products, Fritz et al. (2011a) highlighted that most recent global maps tend to underestimate cropland compared to the official statistics and also disagree with one another. In fact, Ramankutty et al. (2008) have attempted to quantify this uncertainty at the global scale, estimating that global cropland extent varies between 1.22 and 1.71 billion hectares, i.e., differing by more than 40%. Several specific cropland maps were produced at the global or at the continental scale. Pittman et al. (2010) produced the map of global cropland extent at 250-m spatial resolution using multi-year MODIS and thermal data. Two other global maps specifically dedicated to croplands were produced with an emphasis on water management: the global map of rainfed cropland areas (GMRCA; Biradar et al. (2009) and the global irrigated area map (GIAM; Thenkabail et al. (2009). However, their coarse spatial resolution (10 km) does not meet the needs for operational applications and suffer from large uncertainties (Vancutsem et al., 2012) –especially in complex farming systems in Africa.

At the national or regional scale, the use of imagery at 30-m or the integration of multi-sensor images is more common either for the land cover or just for the cropland. If the reduced spatial extent makes it easier to take the local conditions into account and to tune accordingly, the accuracy of the cropland class does not improve necessarily. Besides national land use/land cover programs, some countries have established dedicated annual national crop type mapping based on satellite remote sensing data such as the 30-m US Cropland Data Layer or the 30-m Canadian Annual Crop Inventory. As the production of those maps rely on extensive training data (Boryan et al., 2011), they are still limited to countries with advanced remote sensing programs. Other efforts such as Africover and the GLCN program have realized detailed land cover maps at the country level based on visual interpretation of 30-m spatial resolution images rather than automatic classification. Their update is therefore less frequent. Last, several European countries maintain a Land Parcel Identification System for farmer’s declarations to manage the redistribution of subsidies from the Common Agricultural Policy (CAP). Nevertheless these data sets remain largely unavailable publicly and their use is restricted to CAP activities.
According to Herold et al. (2006), available maps are only marginally validated and when they are, a key concern is that their quality is judged as insufficient for operational applications (Giri et al., 2005; Foody, 2002). Given their large discrepancies in terms of accuracy and spatial agreement, a new trend has emerged: combining existing global and national land cover maps to produce a hybrid map (Fritz et al., 2015) with an increased accuracy. Since the concept has been declined in various approaches either combining the best products (Vancutsem et al., 2012; Latham et al., 2014) or fusing them (Xu et al., 2014). Even though the spatial resolution has increased in the 30-m resolution FROM-Cropland (Yu et al., 2013a), large discrepancies remain between the estimated cropland area and the statistics, e.g., Australia, Africa, Indonesia and Eurasia.

Despite the availability of multiple land cover maps, it is not readily apparent which is most useful for specific applications or how to combine them to provide an improved data set (Herold et al., 2008). The overall objective is to evaluate the current situation to allow strategic planning and mapping effort distribution to maximize the impact of new mapping activities. Prioritizing is key to support to rationalize land cover mapping efforts and focus on countries with actual data gaps. This research proposes to initially assess the best publicly available cropland information coming from global, regional or national land cover maps. Bearing in mind the potential and range of quality of existing land cover maps, this research capitalizes on previous works to harmonize them. It proposes an analytical framework to quantitatively evaluate these maps using four criteria: 1) the thematic information relevant for the cropland definition, 2) the timeliness, 3) the spatial resolution and 4) the confidence level. Based on this initial analysis, two main outputs could be derived: the identification of priority areas for cropland mapping and a unified cropland layer at 250-m that combines the fittest products with regards to the criteria selected.

1.2 Material

1.2.1 Land cover and cropland maps

The identification and collection of national, regional and global land cover maps is a long term enterprise due to the variety of sources and producers involved as well as different data distribution policies. The elaboration of an exhaustive inventory and spatial database is a continuous effort following product releases, updates or changes in policies on data access. Global, regional and national data sets were identified by means of systematic review during working sessions with key individual experts, literature review and web-based search. While collecting them, it was necessary to distinguish the existing data sets from the publicly freely available data (Table 1.1), the former having a distribution policy that prevents its use or having issues to access to the actual georeferenced data set. Therefore, the general rule was that data sets that were not publicly available were considered as non-existent and thus discarded. In the event of products
delivered at multiple epochs, *e.g.*, USDA Cropland Data Layer), the map most contemporaneous with the year 2014 was selected.

**Table 1.1:** Input maps for the analysis. The first and second column detail the extent and references of the considered product. The last column provide the reference year of the product or its time span.

<table>
<thead>
<tr>
<th>Extent</th>
<th>Product Name - Reference</th>
<th>Epoch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>FROM-GLC (Gong <em>et al.</em>, 2013)</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>GlobCover 2009 (Arino <em>et al.</em>, 2008)</td>
<td>2009</td>
</tr>
<tr>
<td></td>
<td>ESA LandCover CCI (Defourny <em>et al.</em>, 2012a)</td>
<td>2008-2012</td>
</tr>
<tr>
<td></td>
<td>MOD12Q1 NASA</td>
<td>2005</td>
</tr>
<tr>
<td></td>
<td>FAO GLC-Share (Latham <em>et al.</em>, 2014)</td>
<td>1990-2012</td>
</tr>
<tr>
<td></td>
<td>IISA-IFPRI Cropland (Fritz <em>et al.</em>, 2015)</td>
<td>1990-2012</td>
</tr>
<tr>
<td></td>
<td>GLC2000 (Bartholomé and Belward, 2005)</td>
<td>1999-2000</td>
</tr>
<tr>
<td></td>
<td>IGBP (Eidenshink and Faundeen, 1994)</td>
<td>1992-1993</td>
</tr>
<tr>
<td></td>
<td>GLCNMO (Tateishi <em>et al.</em>, 2014)</td>
<td>2007-2009</td>
</tr>
<tr>
<td>Regional</td>
<td>Corine Land Cover EEA</td>
<td>2006</td>
</tr>
<tr>
<td></td>
<td>SADC land cover database - CSIR</td>
<td>2002</td>
</tr>
<tr>
<td></td>
<td>JRC Cropland Mask (Vancutsem <em>et al.</em>, 2012)</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>North American Environmental Atlas CEC</td>
<td>2005</td>
</tr>
<tr>
<td></td>
<td>SERENA LAC (Blanco <em>et al.</em>, 2013)</td>
<td>2008</td>
</tr>
<tr>
<td></td>
<td>SEA CRISP (Miettinen <em>et al.</em>, 2012)</td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td>Land Cover CentralAsia (Klein <em>et al.</em>, 2012)</td>
<td>2009</td>
</tr>
<tr>
<td></td>
<td>Africover FAO</td>
<td>1999-2001</td>
</tr>
<tr>
<td></td>
<td>Land Parcel Identification System</td>
<td>2012-2014</td>
</tr>
<tr>
<td></td>
<td>USGS</td>
<td>2000-2001</td>
</tr>
<tr>
<td></td>
<td>SOPAC</td>
<td>1999-2010</td>
</tr>
<tr>
<td></td>
<td>Land Cover Scheme II SERVIR</td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td>GlobeLand30 NGCC (Chen <em>et al.</em>, 2015a)</td>
<td>2009-2011</td>
</tr>
<tr>
<td></td>
<td>JAXA HR LU-LCMap (Takahashi <em>et al.</em>, 2011)</td>
<td>2006-2011</td>
</tr>
<tr>
<td></td>
<td>ACCA (Thenkabail and Wu, 2012)</td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td>Corine Database of Burkina Faso</td>
<td>2000</td>
</tr>
<tr>
<td></td>
<td>Annual Crop Inventory - AAFC</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>Cropland Data Layer USDA</td>
<td>2013</td>
</tr>
</tbody>
</table>
Australia | Digital Land Cover Database GA-Australia (Lymburner and Australia, 2011) | 2011
---|---|---
Cambodia | JICA Land Cover of Cambodia | 2002
New Zealand | Land Cover Data Basev4 Ministry for the Environment | 2004
South Africa | National Land Cover CSIR | 2000-2001
South Africa | National Land Cover SANBI | 2009
Canada | National Resources of Canada | 2005
Uruguay | Land Cover Uruguay UNA-ONU | 2010
Mexico | Land Cover of Mexico CONABIO | 1999
Argentina | Cobertura y uso del suelo - INTA | 2006
Ecuador | Uso del Suelo departamento de Inf. Ambiental | 2001
Thailand | Royal Forest Department of Thailand | 2000
Chile | Chile Corporacion Nacional Forestal | 1999
India | Land Use Land Cover of India NRSC (Sreenivas et al., 2014) | 2012
Gambia | (Holecz et al., 2013) | 2013
Ukraine | Land Cover Ukraine (Lavreniuk et al., 2015) | 2010
Russia | TerraNorte Arable Lands of Russia (Bar-talev et al., 2011) | 2014

### 1.2.2 Global validation data sets

Global reference data sets are mandatory to evaluate the accuracy of land cover maps. Recently, diverse initiatives proposed to share validation data sets (Tab. 1.2). The Global Observation Forest and Land Cover Dynamics (GOFC-GOLD) (Herold et al., 2006) Land Cover Project Office centralizes and provides validation data sets such as the consolidated GLC2000 data set, the consolidated GlobCover 2005 data set, the Visible Infrared Imaging Radiometer Suite (VIIRS) data set, and the System for Terrestrial Ecosystem Parametrization (STEP) data set.

The consolidated GLC2000 data set consist of 1253 samples (70% of the initial data set randomly selected). The sample’s class was converted to an aggregated generalized legend and reinterpreted with Landsat or Google Earth (Bartholomé and Belward, 2005).

The GlobCover 2005 validation data set was built relying on a network of experts familiar with image interpretation and land cover over large areas (Defourny et al., 2009, 2012b). The validation samples were interpreted in Google Earth and with NDVI profiles (annual and average profiles) to illustrate the seasonal dynamics. For a given sample, the expert saw not only the sample point but also a box that coincided with the so-called observational unit corresponding to 5x5 MERIS pixels (225 ha). The experts could describe up to 3 land cover types for each observational unit and provide their level of confidence. The consolidated version results from a reinterpretation of 500 samples randomly selected and re-interpreted. Only samples with high confidence (186) were kept.

The Visible Infrared Imaging Radiometer Suite (VIIRS) Surface Type validation database relies on a a stratified random sample of five hundred 5x5-km
blocks distributed globally (Olofsson et al., 2012; Stehman et al., 2012). The strata were defined by the intersection of a modified Koppen climate classification with a human population density. The sample allocation and distribution within each stratum targeted heterogeneous and complex land cover types more difficult to map. Finally, the samples were interpreted with very high resolution imagery.

In the STEP database (Friedl et al., 2002), each sample is a polygon of about 4-km² is considered a stable example of a specific land cover type. Samples are drawn and labeled in Google Earth. Because the samples distribution does not follow a probability sampling scheme, this data set is suitable for training but not for validation (GOFC-GOLD, 2011).

Recognizing the importance of validated product and their inter-comparison, Zhao et al. (2014) have produced an independent set of well-distributed validation samples. They built a global validation point data set based on interpreting Landsat Thematic Mapper (TM) and Enhanced TM Plus (ETM+) images for a total of 38,664 sample units pre-determined with an equal-area stratified sampling scheme. This was supplemented by MODIS enhanced vegetation index time series data and other high-resolution imagery on Google Earth. Recently, Fritz et al. (2009) proposed a tool—the GeoWiki—to collect volunteered geographic information on land cover from crowd-sourcing. The GeoWiki Project capitalizes on a global network of volunteers who wish to help to improve the quality of global land cover maps. The volunteers are asked to review hot-spot maps of global land cover disagreement and determine, based on what they actually see in Google Earth and their local knowledge, if the land cover maps are correct or incorrect (Fritz et al., 2012). From those data sets, four have been used for two purposes: 1) one for the assessment of the individual maps (Zhao et al., 2014) and 2) three for the assessment of the unified cropland map (GeoWiki, GlobCover 2005 and VIIRS).

Table 1.2: Validation data sets collected, their geometries and percentage of cropland samples.

<table>
<thead>
<tr>
<th>Validation set</th>
<th>Geometry</th>
<th>Sample Size</th>
<th>Cropland [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlobCover 2005</td>
<td>Polygon (225-ha)</td>
<td>186</td>
<td>9</td>
</tr>
<tr>
<td>VIIRS</td>
<td>Polygon (5x5-km)</td>
<td>3664</td>
<td>27</td>
</tr>
<tr>
<td>STEP</td>
<td>Polygon (4x4-km)</td>
<td>1780</td>
<td>26</td>
</tr>
<tr>
<td>GLC-2000</td>
<td>Point</td>
<td>1253</td>
<td>9</td>
</tr>
<tr>
<td>Zhao et al.</td>
<td>Point</td>
<td>38 664</td>
<td>7</td>
</tr>
<tr>
<td>GeoWiki</td>
<td>Polygon (1x1-km)</td>
<td>12 833</td>
<td>29</td>
</tr>
</tbody>
</table>

1.2.3 Ancillary data

Country-wise cropland percentages were extracted from the FAOSTAT database. The FAOSTAT database records an inventory of land resources per country on a yearly basis. The database is constituted of official reports collected from over 200 countries. FAO defines the arable land as: “the land under temporary agricultural crops (areas with multiple cropping are counted only once), temporary
meadows for mowing or pasture, land under market and kitchen gardens and land temporarily fallow (less than five years). The abandoned land resulting from shifting cultivation is not included in this category.

Field size information was also made available (Fritz et al., 2015). This data set was collected through a GeoWiki crowd-sourcing campaign in which users were asked to label the field size using high-resolution imagery, where examples were provided to guide the user. Four categories were proposed: very small, small, medium and large. Note that the result is provided in arbitrary units (very small = 10, small = 20, medium = 30 and large = 40). A validation exercise of the global field size map revealed satisfactory agreement with control data, particularly given the relatively modest size of the field size data set used to create the map.

1.3 Method

To play its role as mask, a cropland map should comply to different criteria such as adequate class definition, accuracy, timeliness and adequate spatial resolution with regards to the area of interest. Areas where the current maps do not satisfy these criteria are considered as priority areas for cropland mapping. Therefore, the assessment of the cropland products must consider these different criteria. To handle these different dimensions, this study proposed an approach based on a multi-criteria analysis. Four criteria have been selected to evaluate the need for an updated cropland at the national level: the adequacy of the current legend, the adequacy of the spatial resolution, the timeliness and the confidence level with respects to the validation data sets (Figure 1.1). The rationale behind using a multi-criteria analysis was to combine the conflicting objectives described by different data sources into a single index form for multiple criteria evaluation (Setegn et al., 2009) in order to support decision making and priority analysis (Mourão et al., 2014). Similarly to other studies, e.g. (Iojă et al., 2014)), the general outline of the methodology included the following steps:

- Constructing a data base containing all the spatial information;
- Transforming the chosen criteria into scores;
- Determining the weight for each criterion;
- Aggregating the data weights, obtaining the scores for each data set and selecting the best score in overlapping regions.

At each step, scores were attributed to every product on a per-country basis following a default mathematical rule and were then reinterpreted by experts to ensure more robustness. The four criteria were finally aggregated into a priority indicator following a summation model with equal weights and again reinterpreted by the experts. For example, all maps covering country $i$ were analyzed with respect to the four criteria. Those were then aggregated into a priority index, one per map. The fittest map is the one that minimizes the
1.3. Method

Figure 1.1: The analytical framework of the assessment is a systematic analysis of four criteria that characterize the fitness of a cropland map. The four criteria are the thematic consistency, the timeliness, the resolution adequacy and the confidence level. From this analysis, two outputs were derived: 1) a priority map for global cropland mapping and 2) a unified cropland layer combining the fittest products.

priority index. Areas with a high priority index characterize priority areas for cropland mapping, whereas areas with low scores correspond to an accurate and precise current mapping. Besides the multi-criteria evaluation of the maps it was also possible to derive the fittest global cropland map at 250 m thanks to the output of from the multi-criteria analysis.

1.3.1 Thematic Consistency Criterion

As there were no common agreement on the cropland definition, products often provided their own definition of cropland and might not be compatible with one another, making inter-comparison and legend harmonization problematic (Herold et al., 2006; McCallum et al., 2006). The Food and Agriculture Organization (FAO) has been undertaking efforts to establish international standards for land cover since late 90’s. Its experience in class definition was synthesized in the Land Cover Classification System (LCCS) (Di Gregorio and Jansen, 2000). LCCS concept has recently evolved in the Land Cover Meta Language (LCML). LCML is an object oriented classification system where each land cover features is characterized by a series of elements that can be further detailed by a set of attributes. The class meaning is not anymore related to a simple class name but to a more exhaustive and modern model populated by the elements and attributes characterizing the land cover features.

Recently, a cropland definition tailored for cropland monitoring and conform to LCML has been adopted by the Joint Experiment of Crop Assessment and Monitoring (JECAM) network. The general definition-cropland (including area affected by crop failure) reads as follows: “The annual cropland from a remote sensing perspective is a piece of land of minimum 0.25 ha (min. width of 30 m) that is sowed/planted and harvestable at least once within
the 12 months after the sowing/planting date. The annual cropland produces an herbaceous cover and is sometimes combined with some tree or woody vegetation.”. There are three known exceptions to this definition. The first concerns the sugarcane plantation and cassava crop which are included in the cropland class although they have a longer vegetation cycle and are not yearly planted. Second, taken individually, small plots such as legumes do not meet the minimum size criteria of the cropland definition. However, when considered as a continuous heterogeneous field, they should be included in the cropland. The third case is the greenhouse crops that cannot be monitored by remote sensing and are thus excluded from the definition. This definition discards perennial crops and fallow as they are less important to monitor from a food security point of view. In addition, multi-annual crop are less sensitive to climatic conditions and thus less interesting for monitoring using remote sensing for food security. If the JECAM definition seems the most appropriate for cropland masking, a more pragmatic definition was adopted because the major part of the data sets available did not consider the annual herbaceous cropland. The JECAM definition was thus modified as follows: “the cropland is a specific area occupied by an herbaceous crop under permanent or fallow cultivation period (including active shifting cultivation fields).”. Note that this definition is also LCML compatible. To evaluate the thematic distance between a map and the desired legend, presence/absence of given components of the legends were assessed on the basis of a set of binary criteria. The components were the following:

1. Absence of woody crops [WC];
2. Presence of fallow and bare fields [FB];
3. Absence of managed pasture and meadows [MPM];

If for a given product a criterion is met, this product scores 1 and 0 conversely. The final thematic criterion ($ThC$) is the sum of the scores:

$$ThC = WC + FB + MPM$$

The scores were then reclassified according to Table 1.3a.

1.3.2 Timeliness Criterion

As reported by numerous studies, the world’s croplands are very dynamic not only because of crop rotations and practices but also because of land cover changes such as land conversion from agriculture to urban (del Mar López et al., 2001), from forest to agriculture (Morton et al., 2006), or even agricultural abandonment (Benayas et al., 2007; Baumann et al., 2011). To address both these annual and multi-annual changes, the timeliness criterion ($TiC$) characterizes the number of years elapsed since the reference year of a map. The timeliness criterion is computed as the difference between the year of interest (2014) and the epoch of the product:

$$TiC = 2014 - t_p$$
Table 1.3: Systematic recoding rules for the four cropland criteria used in the multi-criteria analysis. Each criterion is recoded in a four level score ranging from one to four.

(a) Rules for the thematic criterion

<table>
<thead>
<tr>
<th>Thematic criterion</th>
<th>Code</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Good Thematic Agreement</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Moderate Thematic agreement</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>Low Thematic agreement</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>No thematic agreement</td>
<td>1</td>
</tr>
</tbody>
</table>

(b) Rules for the timeliness criterion

<table>
<thead>
<tr>
<th>Timeliness criterion</th>
<th>Code</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>Up to date</td>
<td>4</td>
</tr>
<tr>
<td>2-5</td>
<td>Recent</td>
<td>3</td>
</tr>
<tr>
<td>5-10</td>
<td>Old</td>
<td>2</td>
</tr>
<tr>
<td>10+</td>
<td>Out of date</td>
<td>1</td>
</tr>
</tbody>
</table>

(c) Rules for the resolution adequacy criterion

<table>
<thead>
<tr>
<th>Resolution Adequacy Criterion</th>
<th>Code</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;0</td>
<td>Completely adequate</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>Adequate</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Inadequate</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Completely Inadequate</td>
<td>1</td>
</tr>
</tbody>
</table>

(d) Rules for the confidence level criterion

<table>
<thead>
<tr>
<th>Confidence Level Criterion</th>
<th>Code</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>80-100%</td>
<td>High Confidence Level</td>
<td>4</td>
</tr>
<tr>
<td>70-80%</td>
<td>Good Confidence Level</td>
<td>3</td>
</tr>
<tr>
<td>60-70%</td>
<td>Low Confidence</td>
<td>2</td>
</tr>
<tr>
<td>0-60%</td>
<td>Very Low Confidence Level</td>
<td>1</td>
</tr>
</tbody>
</table>

where \( t_p \) is the epoch of map \( p \) as defined in Table 1.1. The differences are then reclassified into four groups for which a score is assigned (Table 1.3b).

1.3.3 Resolution adequacy criterion

The spatial resolution required for accurate cropland mapping is a function of the field size, the landscape fragmentation and to some extent the crop diversity. Indeed, areas with small parcels but low crop diversity tend to behave similarly to a large field. Several methods have been developed to evaluate the resolution required for crop mapping in a specific area (Löw and Duveiller, 2014; Leroux et al., 2014; Vintrou et al., 2012a). However, they all require (very) high spatial resolution-cropland/crop type map as input to diagnose the required resolution, which makes those methods unsuitable for an a priori resolution definition especially as they are not available globally. In this study, the field size solely was taken as proxy to derive the spatial resolution requirements for cropland mapping. Capitalizing on the GeoWiki tool (Fritz et al., 2012), volunteers were asked to label the size of the fields based on a very high resolution image and a
Chapter 1. Priority mapping

reference 1x1-km square box. Observations were then interpolated within the cropland areas using an simple inverse distance approach. The four observed labels (large, medium, small and very small) can then be related to the GEOGLAM requirements for cropland mapping (Table 1.4).

Table 1.4: Linking the observed field size by crowd-sourcing to spatial resolution requirements

<table>
<thead>
<tr>
<th>GeoWiki Field Size</th>
<th>GEOGLAM field size [ha]</th>
<th>GEOGLAM resolution requirements [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
<td>&gt;15</td>
<td>100-500</td>
</tr>
<tr>
<td>Medium</td>
<td>&gt;1.5</td>
<td>20-100</td>
</tr>
<tr>
<td>Small</td>
<td>&gt;0.15</td>
<td>5-20</td>
</tr>
<tr>
<td>Very Small</td>
<td>&lt;0.15</td>
<td>&lt;5</td>
</tr>
</tbody>
</table>

For each country, a histogram of the field sizes was computed (Figure 1.2a). To define the spatial resolution required for cropland mapping, the assumption was that the target resolution, and thus field size, should be adequate for at least 75% of the largest fields (Figure 1.2b and Figure 1.2c). The extracted resolution was then related to the GEOGLAM requirements (as shown in Tab. 1.4). The difference in number of classes between the actual class and the requirement class gave the resolution adequacy criterion score $RC$ (Table 1.3c).

1.3.4 Confidence level criterion

Prior to the confidence level assessment, it was necessary to harmonize the legend of both the maps and the validation data sets. The legend of each data set was thus translated into the binary legend that corresponded best to the proposed legend. The confidence level criterion ($CC$) was assessed by means of confusion matrices from which overall accuracy indicators were derived and then reclassified in four categories (Table 1.3d). The reference data set produced by Zhao et al. (2014) was utilized to extract the national-level confusion matrices as i) it was the most populated—and thus relevant at the national scale—and ii) its legend definition is thematically close to the one proposed in this study.

1.3.5 Criteria aggregation and priority identification

For each product, the scores were reinterpreted by experts to ensure that the mathematical score attribution matched their experience as user and/or visual analysis. Weighted linear combination of the criteria is the most common approach to aggregate the different dimensions and different approaches exist to estimate their relative weight (Saaty, 1990; Mendoza and Prabhu, 2000). Equal weights were assigned to each criterion as they were considered equally important. Therefore, the priority indicator ($PI$) for country $i$ is computed as:

$$PI_i = \min_j [16 - (ThC_{ij} + TiC_{ij} + RC_{ij} + CC_{ij})]$$

(1.3)

where $j$ is the $j$th product and 16 is the maximum aggregated score (four criteria of score levels each). The priority indicator was then reclassified into
1.3. Method

Figure 1.2: Distribution of field size per country and histograms of the field size distribution for two contrasted countries. The density corresponds to relative frequencies (probabilities) of pixel occurrence. Note that the field size is provided in arbitrary units (very small = 10, small = 20, medium = 30 and large = 40).

three classes: no priority (0-5), low priority (6-7) and high priority (8-11). One could argue that summing equally-weighted criteria would not allow to capture situations with one prohibitive criterion. Therefore, the typology of the priority was introduced: if a criterion scored less than 2, it was reported as requiring an update. The use of the error typology can highlight dilution of a low criterion. For example, the confidence level of j map could be very low but all three other criteria at their maximum level, resulting in a priority index of 3. The analysis the error typology allow both to have a synoptic view of the dominant type of update required and to report situations where a criterion with low score is diluted in the priority index. It should be underlined here that the error typology indicates only the dominant error type; yet it does not necessarily imply that the other typology are necessarily satisfactory.
1.3.6 Spatial aggregation and assessment

For each country, the fittest product corresponded to the one with the lowest priority index. As a result of the analysis, one could produce a cropland for the year 2014 by joining these fittest products in a predefined grid of 250-m cells. This unified cropland layer was assessed by means of confusion matrices derived from the GLC2000, GlobCover and GeoWiki reference data sets (Figure 1.3), from which accuracy indicators such as the overall accuracy, producer and user accuracy. It is assumed that an accuracy assessment with multiple validation data sets would give the most realistic estimation of the unified cropland layer as none of them match its legend perfectly.

Figure 1.3: Distribution of the reference samples used for the validation of the 250-m unified cropland layer. Three global validation data sets are available (VIIRS, GlobCover and GeoWiki) and account together more than sixteen thousand validation samples distributed over the globe.

With the recent releases of high resolution global forest cover products (Hansen et al., 2013; Shimada et al., 2014), masking out forested areas from agricultural lands is now possible. This could be particularly valuable in areas where products have a coarse resolution and lack timeliness. In particular, the Advanced Land Observing Satellite (ALOS) Phased Arrayed L-band Synthetic Aperture Radar (PALSAR) forest maps provides a global forest maps at 25-m four for epochs. These maps were validated with three types of the ground truth data at an accuracy of 90%. The 2010 ALOS-PALSAR forest map was resampled using the grid of the Unified Cropland Layer and pixels with a majority of forest cover were reclassified as non-cropland. The impact of this mask was assessed with the GeoWiki validation data set.

Furthermore, the unified cropland layer was also resampled to 1-km and compared to GLC-Share (Latham et al., 2014) and the IIASA-IFPRI Cropland (Fritz et al., 2015). Each product legend was converted to a binary cropland/non-cropland legend: pixels were considered as cropland if the cropland proportion
was not null. Spatial disagreement areas and corresponding the percentages of disagreement between products were computed at the continental level.

1.4 Results and Discussion

1.4.1 Multi-criteria analysis

The multi-criteria approach allowed assessing quantitatively four characteristics of a cropland map and summarized it into a single spatialized priority index (Figure 1.4a) as well as its associated the priority typology (Figure 1.4b). From a general point of view, the high resolution land cover mapping initiatives in Northern America—the Cropland Data Layer and the Annual Crop Inventory in the US and Canada—Europe—the Land Parcel Information System—and in Africa (Africover) are well captured by the priority index. The current medium spatial resolution maps seems to fit large field areas such as Russia and Central Asia. The dominant typologies of updates are timeliness (Brazil, Chile, South Africa), resolution (West Africa) or both (Madagascar, Myanmar, Pakistan). To further interpret the result, the priority index can be broken down into three classes: no priority (0-5), low priority (6-7) and high priority (8-11). One can observe that high priority areas are mainly associated with resolution adequacy and/or timeliness improvements. However, these dominant error typologies are tightly tied with other typologies such as the confidence level.

To support the prioritization, the priority index and the associated typology can be related to the proportion of cropland at the national scale (Figure 1.5a). A way of prioritization would be to ensure that both major agricultural commodity producers—whose production is likely to influence the market prices—and food insecure countries are well mapped (Figure 1.5b). These are the countries monitored by the Agriculture Market Information System (AMIS) and by the Food Early Warning System network.

Mexico and Central American countries would benefit from a thematic update and a temporal update respectively. Because of the high percentage of cropland in the national landscape, one should also consider in Cuba and Dominican Republic. South America does not appear as a critical area for update. Colombia and Peru scored the highest priority index associated to a temporal and resolution update respectively. However, one might consider Brazil—an AMIS country—for the large cropped areas and its major role as soybean producer.

Europe appears as a moderate priority area (orange shades) which might be explained by the fact that Corine Land Cover 2006 was selected and sanctioned for its timeliness, its ambiguous legend and the mismatch of the spatial resolution in certain countries. Greece appears in the typology of the priority map as it was not included in Corine 2006. In the course of 2015, the Corine Land Cover map for 2012 and covering thirty nine European countries should be released. With its increased spatial resolution (10 m), this new Corine Land
Chapter 1. Priority mapping

Figure 1.4: Priority map and its update typology. Areas with high priority index (reddish shades) characterize priority areas for cropland mapping, whereas areas with low scores correspond to an accurate and precise current mapping (greenish shades). West Africa, Ethiopia and South East Asia (Indonesia) clearly appear as priority areas for cropland mapping.

Cover map is expected to strongly improve the cropland delineation over Europe.

There are two main areas for improvements in Africa: Western and Southern Africa. In particular, countries such as Nigeria, Benin, Togo, Ghana and Sahelian Countries in the West and Mozambique, Madagascar and Zimbabwe in the South would really benefit from an update as cropland constitutes a large part of the landscape. In addition, the type of updates required varies from an improvement in the resolution and in the timeliness or both (Fig. 1.4). In the Greater Horn of Africa, Ethiopia—a country at risk—should also be considered as a priority. It is worth noting that the results for Africa corroborate those obtained by other studies that underlined the need for an update in Uganda, Ivory Coast and Nigeria (Vancutsem et al., 2012) and Burkina Faso (Fritz et al., 2012). Leroux et al. (2014) highlighted that the MODIS land cover over Ethiopia had limited accuracy. However, an update does not necessarily mean making a new map but instead releasing an existing one freely and openly to the public, e.g., Ethiopia, Mali, Algeria.
1.4. Results and Discussion

(a) Percentage of cropland (FAO)

(b) FEWS and AMIS countries distribution

Figure 1.5: National-level agricultural profiles. The percentage of cropland and the belonging to the major agricultural commodity producers (AMIS countries) and/or to the countries at risk from a food security point of view (FEWS) further support the effort prioritization. As an example, Sahelian countries are characterized by a high priority index and considered as countries at risk. New cropland mapping efforts appear critical to update the cropland information in this area.

In the Middle East, the priority index of Saudi Arabia and Yemen yields high values, especially due to the mismatch between the resolution of the current product and the suitable resolution. However, those countries have a particularly low proportion of cropland. In Asia, Pakistan, Myanmar and Vietnam appear as high priority and are associated with lack of spatial resolution and/or timeliness. In addition, those countries are major commodity producers and have therefore a large importance on the agriculture markets.

Finally, the large rice growing areas of South East Asia, mostly in Indonesia, occurred as priority since the high cloud coverage in these areas limit the efficiency of optically-derived land cover maps.

1.4.2 Spatial aggregation: assessment and comparison

The best performing products, i.e., those having the lowest priority index, were extracted country-wise. They were resampled at 250-m and combined to form
Chapter 1. Priority mapping

the Unified Cropland Layer (Fig. 1.6).

Figure 1.6: Cropland proportion from the Unified Cropland Layer at 250 m for the year 2014. The Unified Cropland Layer combines the best performing products with regards to the multi-criteria analysis. It should be noted that cropland proportions are likely to be overestimated in areas where the spatial resolution of the original product is not higher than 250-m.

The unified cropland layer was validated using three different and independent reference data sets: the consolidated GlobCover 2005 (Table 1.5a), the VIIRS data set (Table 1.5b) and the GeoWiki data set (Table 1.5c). The overall accuracy figures vary between 82% and 95%. However, it should be noted that 95% accuracy might be optimistic as the consolidated GlobCover data set would mostly focus in places with obvious interpretation. While the producer and user accuracies are stable for the non-cropland class, large differences appeared for the cropland class. Besides, as mentioned in (Tsendbazar *et al.*, 2014), attention has to be paid to the definition of cropland and pasture (Klein Goldewijk *et al.*, 2007; Ramankutty *et al.*, 2008) when validating maps with global land cover reference data sets as they might introduce some bias. In this case, the assessment might wrongly penalize fallow areas and permanent crops that were excluded from the unified cropland layer as a result of differences in the legend definition. The best accuracy figures for the cropland class were obtained with the most up-to-date data set (VIIRS): 72% of producer accuracy and 87% of user accuracy. The additional masking of the forested area displays a positive impact on the overall accuracy (from 82.2% to 84.5%) and appears even more beneficial for the user’s accuracy of the cropland class (+20%), *i.e.*, less commission errors (Table 1.5d). This gain is counterbalanced by a diminution of the user’s accuracy for the non-cropland class and of the producer’s accuracy of the cropland class.

The comparison with two similar products (GLC-Share and the IIASA Cropland) revealed large spatial differences ~45% of agreement of all three products globally (Fig. 1.7). It is worth noting that GLC-Share takes the IIASA Cropland directly as input leading inevitably to inbred agreement. Besides, the IIASA map is actually consistent with FAO statistics and whereas the unified
Table 1.5: Accuracy assessment of the Unified Cropland Layer. The overall accuracy figures vary between 82% and 95%. However, it should be noted that 95% might be optimistic as the consolidated GlobCover data set would mostly focus in places with obvious interpretation. Masking cropland areas with a forest map reduces the commission errors from 46% to 26%.

(a) Confusion matrix obtained with the GlobCover 2005 data set

<table>
<thead>
<tr>
<th></th>
<th>Non-cropland</th>
<th>Cropland</th>
<th>User’s Acc. [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-cropland</td>
<td>158</td>
<td>9</td>
<td>95.2</td>
</tr>
<tr>
<td>Cropland</td>
<td>2</td>
<td>16</td>
<td>87.5</td>
</tr>
<tr>
<td>Producer’s Acc.</td>
<td>98.8</td>
<td>63.6</td>
<td>Overall Accuracy [%]: 94.5</td>
</tr>
</tbody>
</table>

(b) Confusion matrix obtained with the VIIRS data set

<table>
<thead>
<tr>
<th></th>
<th>Non-cropland</th>
<th>Cropland</th>
<th>User’s Acc. [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-cropland</td>
<td>631</td>
<td>63</td>
<td>89.4</td>
</tr>
<tr>
<td>Cropland</td>
<td>251</td>
<td>985</td>
<td>87.5</td>
</tr>
<tr>
<td>Producer’s Acc.</td>
<td>85.9</td>
<td>72.2</td>
<td>Overall Accuracy [%]: 82.3</td>
</tr>
</tbody>
</table>

(c) Confusion matrix obtained with the GeoWiki data set

<table>
<thead>
<tr>
<th></th>
<th>Non-cropland</th>
<th>Cropland</th>
<th>User’s Acc. [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-cropland</td>
<td>8490</td>
<td>384</td>
<td>95.6</td>
</tr>
<tr>
<td>Cropland</td>
<td>1698</td>
<td>2055</td>
<td>54.7</td>
</tr>
<tr>
<td>Producer’s Acc.</td>
<td>83.3</td>
<td>84.3</td>
<td>Overall Accuracy [%]: 82.2</td>
</tr>
</tbody>
</table>

(d) Confusion matrix obtained for the Unified Cropland Layer masked by the ALOS PALSAR Forest mask with the GeoWiki data set

<table>
<thead>
<tr>
<th></th>
<th>Non-cropland</th>
<th>Cropland</th>
<th>User’s Acc. [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-cropland</td>
<td>8085</td>
<td>999</td>
<td>89.0</td>
</tr>
<tr>
<td>Cropland</td>
<td>986</td>
<td>2763</td>
<td>73.7</td>
</tr>
<tr>
<td>Producer’s Acc.</td>
<td>89.1</td>
<td>73.4</td>
<td>Overall Accuracy [%]: 84.5</td>
</tr>
</tbody>
</table>

cropland layer is not. Large cropland disagreements are evident in South-East Asia, North America and South America as well as some African countries. The interpretation of the disagreement map is not straightforward as it integrates classification errors when different sources are used but also demonstrates thematic differences or semantic discrepancies, e.g., permanent crops, pastures, fallows). In addition, areas of agreement do not imply automatically accurate mapping as the three products may use the same source as input. In West Africa, inter-product disagreement appears as low as only global land cover maps are available. Disagreements in confusion errors due to the pastures and grasslands contaminates the cropland in the South of Belgium, the Netherlands and Northern France. In Spain, the same occurred with olive groves, fruit trees and vineyards or pastures as well as in the USA. One of the largest areas of disagreement is certainly South East Asia where the cropland extent appears more restrictive in the Unified Cropland Layer –33% of agreement for the continent. The lack of agreement underlines the poor accuracy of the cropland class in the current product and advocates for the development of new methods and products best suited for this particular land cover class. However, the spatial disagreement analysis supports the prioritization for new mapping
efforts are necessary to clearly identify cropland in disagreement areas. Previous priorities are confirmed: South-East Asia (Indonesia and Vietnam), South America (Brazil and Colombia) and African countries (Western and Southern countries, Ethiopia and Madagascar).

![Figure 1.7: Agreement map between GLC-Share, the IIASA Cropland and the unified cropland layer. Large cropland disagreements are evident in South-East Asia, North America and South America as well as some African countries. A significant part of the disagreements might be due to the different legends chosen. Previous priorities are confirmed: South-East Asia (Indonesia and Vietnam), South America (Brazil and Colombia) and African countries, e.g., Western and Southern countries and Madagascar.](image)

### Figure 1.7

(a) Agreement distribution

(b) Agreement and disagreement proportions at the continental level

#### 1.5 Discussion

To improve the quality and reliability of global cropland maps as base information for crop condition monitoring and Earth system modeling, four critical issues must be addressed: identifying priority areas for cropland mapping, adopting a shared cropland definition, open data access policies and sharing the maps as well as setting up validation mechanisms to systematically validation new cropland maps.

This paper tackled in particular the first issue by means of a multi-criteria analysis that was carried out globally with publicly available data sets. Existing land cover maps not available at the time of the writing were considered as not existing. The analysis might thus be biased for a country according to the prod-
1.5. Discussion

uct availability. An underlying hypothesis was that countries were considered as homogeneous ensembles with regards to the field size. This assumption seemed reasonable for most countries but might not hold true for countries presenting a large variety of field sizes and crop diversity. The quality of the confidence level criterion depends on the density of reference data set and its distribution within each country. Certainly, some drawbacks are i) the limited number of experts involved in the validation data collection and ii) that points are very sensitive to geolocation errors. The combination of an expert-based approach with the multi-criteria analysis allowed identifying a set of countries for strategic effort planning such as: Mexico, Brazil and several Central American countries, Ethiopia, Western and Southern African countries, Myanmar and Indonesia. Overall, this diagnosis is confirmed (for instance Mozambique, Indonesia or Ethiopia) by the attempt of See et al. (2015) to map priority countries on the basis of an index that combines the level of spatial disagreement of global cropland maps with regards to FAO Statistics and the Global Hunger Index. However, it should be noted that for several priority countries such as Ethiopia and Algeria—the update needed does not necessarily imply making a new map but rather releasing an existing map openly to the public. It should be noted that the multi-criteria approach remains valid to create national level and local level cropland maps for other purposes.

This study would not have been possible without formal or informal data sharing and open data policies. This study showed that by combining the best-performing cropland maps, the cropland class reached a level of accuracy of 82-95% which outperformed the accuracy of single global cropland products. However, these figures might be revised if using a validation data set more adapted to the legend and better suited for cropland validation. These accuracy figures are close to those obtained by other multi-product global cropland maps IIASA (82.4%) (Fritz et al., 2015). In the moderate/long term, the priority map will evolve with future map releases, e.g., Corine Land Cover 2012, changes in data policy and with new collection from different national and international institutions. As a direct result, the accuracy of the unified cropland layer is expected to increase. The exhaustive identification of land cover products is a tedious task. To facilitate the identification and use of land cover products by the community, this study recommends to register systematically Earth Observation resources into the GEOSS Portal which is a main entry point to Earth Observation data linking world-wide community of practice in nine societal benefit area among which agriculture. With the new and upcoming high resolution optical satellites such as Landsat-8 and Sentinel-2, the number of high resolution land cover products is expected to increase and their accuracy to improve. However, cloud cover impedes optical satellite remote sensing instruments from obtaining clear views of the Earth’s surface. In some regions, cloud cover impedes passive satellites from imaging the Earth’s surface at key moments of the development cycle for crop recognition and monitoring. Whitcraft et al. (2015a) studied in depth how cloud cover impacts the probability of securing reasonably clear views of croplands using passive optical Earth observations as the agricultural growing season progresses. They highlighted that in many important agricul-
tural areas where the cloud cover is so persistent and pervasive that less than half of their 8-day composites would be even 70% cloud-free. This suggests that in those areas synthetic aperture radar could play a major role in improving the accuracy of land cover maps, especially with the recent systematic acquisition of the Sentinel-1 mission. The potential of radar for cropland mapping has already been demonstrated in several configurations, separately (Blaes et al., 2005; McNairn et al., 2009b,a, 2014) or combined with optical data (Forkuor et al., 2014; Kussul et al., 2013). Radar could be particularly valuable to improve the cropland information in two priority areas, namely South East Asia and around the Guinea Gulf (Nigeria). One could also capitalize on the synergies between medium spatial resolution- high temporal frequency sensors and high spatial resolution-low temporal frequency sensors (Gao et al., 2006; Bisquert et al., 2015). But these future products have yet to be released with open data policies. As stated by See et al. (2015), crop and land use maps produced by multi-lateral organization such as the United Nations and the World Bank should be widely shared, for opening up data can lead to increased innovation and entrepreneurship along with substantial financial gains. The situation seems to move towards more open policies through concerted efforts of the United Nations, the Group on Earth Observations and national governments.

Without legend harmonization, multi-products will remain by definition inconsistent. This is not only valid for the cropland but for all land cover classes. The cross-comparison with other equivalent products underlined the limitations of working with different legends. The FAO Land Cover Meta Language provides a robust theoretical framework for legend definition and should certainly pay a key role in class definition harmonization. The cropland definition used for the unified cropland layer is a pragmatic one, still constrained by the current and diverse definitions. The recommendation is to move towards the JECAM definition of the cropland which has already been adopted by several projects and research. This shared definition will facilitates across-site comparisons in benchmarking activities, one of the key objective of JECAM. A direct implication of this definition is that the cropland becomes dynamic as the class of same field might change from one year to another according to the fallow cycle or if the field is harvested. This implies that efforts should be also directed to develop automated cropland classification methods to be applied every year. For a timely crop status monitoring, it means that those methods need to deliver dynamic cropland masks as the mask of one specific season cannot be re-used from year to year.

Finally, the collection of a global validation data set relevant for cropland mapping at the global scale would be a major contribution to the Agriculture Community in order to assess systematically the future cropland maps. Such a mechanism should be coordinated and set up within Group on Earth Observations. This will require greater involvement of experts on the ground or by means of image-based visual assessment and the collection of a larger quantity of \textit{in situ} data. The GeoWiki could be a valuable tool to collect, to cross-validate and to update the status of global common validation data sets. It would be
also critical that instructions given to the interpreters is provided to the users along with the data itself. Confidence levels associated with the label would also be valuable to inform users on the expected robustness of a specific validation sample.

1.6 Conclusion

For both food security monitoring and climate modeling, a good cropland mask is critical to isolate the agriculture component of the landscape to analyze the crop conditions or simulate the response of cropland to different climatic simulations. A plethora of initiatives mapped cropland at different scales but due to its dynamic intrinsic nature and the wide variety of agro-systems, its accuracy is often limited—especially in general global land cover products. The overall objective of this study was to evaluate the current situation in terms of data set availability to allow strategic planning and mapping effort distribution and therefore to maximize the impact of new mapping activities. After an exhaustive identification and collection of land cover maps, a multi-criteria analysis was designed at the country level to highlights those priority areas for cropland mapping. Three critical priority areas were identified: African countries mainly in West Africa, South East Asia (Indonesia) and South America (Brazil). Other countries such as Ethiopia, Madagascar, Mozambique and Pakistan should also strongly be considered. Moreover, building on the priority analysis, a Unified Cropland Layer was also produced by combining the fittest products. The accuracy of the Unified Cropland Layer was assessed with available global validation datasets and yielded an overall accuracy ranging from 84 to 95%, outperforming most global cropland maps. Besides, masking cropland areas with a forest map reduced the commission errors from 46% to 26%. Compared to other multi-product maps, strong spatial disagreements were found and might be attributed to the differences in the legend definition (mainly due to the inclusion of grassland and perennial crops in the cropland class). This should encourage the community to adopt a shared definition of the cropland as well as to collect a global validation data set relevant for the agriculture class to systematically assess future cropland maps.
Chapter 2

Automated Annual Cropland Mapping using Knowledge-Based Temporal Features*

**Highlights**

- A simple and comprehensive methodology is introduced to map cropland based on high spatial and temporal frequency satellite data.
- Reliable pixels are extracted from existing maps to calibrate new classifiers.
- The method was applied in four contrasted landscapes and reached >80% accuracy.
- The accuracy of the existing maps used for calibration affects more the classification confidence than its outcome.
- The knowledge-based spectral-temporal features are stable over time when extracted in a rolling windows which increases the generalization potential of the method.

**Abstract.** Timely, accurate and cost-effective cropland mapping is a prerequisite for reliable crop condition monitoring. This article presented a simple and comprehensive methodology capable to meet the requirements of operational cropland mapping by proposing 1) five knowledge-based temporal features that remain stable over time, 2) a cleaning method that discards misleading pixels from a baseline land cover map and 3) a classifier that delivers high accuracy cropland maps (>80%). This was demonstrated over four contrasted agrosystems in Argentina, Belgium, China and Ukraine. It was found that the quality and accuracy of the baseline impact more the certainty of the classification rather

---

than the classification output itself. In addition, it was shown that interpolation of the knowledge-based features increases the stability of the classifier allowing for its re-use from year to year without recalibration. Hence, the method shows potential for application at larger scale as well as for delivering cropland map in near real time.

2.1 Introduction

Given the prospects on the human population growth, the shift to more meat-based diets and processed food consumption as well as the development of agro-biofuels, the food supply system is subject to ever increasing pressures (Godfray et al., 2010; Rounsevell et al., 2005; Searchinger et al., 2008). Besides, climate change is already hampering agricultural growth. Increases in the frequency and intensity of extreme events such as drought, heavy rainfall, flooding and high maximum temperatures are occurring and expected to accelerate in many regions (Field and Van Aalst, 2014). These changes affect crop production in several regions of the world, with more negative effects, especially in food unsecured developing countries (Parry, 2007). In Africa for example, food security issues are likely to increase under an uncertain, changing climate (Brown and Funk, 2008; Wheeler and Von Braun, 2013) whereas in Europe, yields are expected to become more variable (Trnka et al., 2011). The impact that climate change will have on agricultural yields, and the knock-on effect that this will have on food prices, is a critical question (Challinor, 2011). Both direct observations and models have important roles to play in understanding their relationships (Battisti and Naylor, 2009). Cropping systems would need to adapt to those changes and a variety of options has already been proposed as having the potential to reduce vulnerability of agricultural systems to risks related to climate change. These options include technological developments (new crop varieties and resource management), government programs and insurance (subsidy and support programs) as well as farm production practices (Smit and Skinner, 2002). Farm level adaptations relate to planting and harvest dates, crop rotations, selection of crops and crop varieties for cultivation, water consumption for irrigation, use of fertilizers, and tillage practices (Adams et al., 1998). For example, Teixeira et al. (2013) suggested that adapting crop types and sowing dates at the local level may partially mitigate heat stress.

In this context, timely and dependable information on agricultural production appears crucial for agricultural market stability and would particularly benefit agencies working towards increasing food security (Justice et al., 2007). Mapping the cropland extent is a prerequisite necessary for both near real time crop monitoring and environmental climate change simulations (Waldner et al., 2015c): cropland maps serve to i) isolate the agricultural component to assess the crop growth conditions and to ii) investigate how the croplands respond to different climatic conditions.
2.1. Introduction

The literature on cropland mapping is replete with classification approaches across scales such as (a) maximum likelihood (Abou El-Magd and Tanton, 2003), (b) nearest neighbors (Samaniego and Schulz, 2009), (c) logistic regression (Peña et al., 2014), (d) expert-based decision rules (Conrad et al., 2010), (e) spectral angle mapper (Alganci et al., 2013b), (f) tasseled cap transformation (Crist and Cicone, 1984), (g) decision tree algorithms (Wardlow et al., 2007; Pittman et al., 2010), (h) random forest (Duro et al., 2012; Long et al., 2013), (i) neural network methods (Liu et al., 2005b), (j) support vector machine (Alganci et al., 2013b), (k) spectral unmixing techniques (Yang et al., 2007a), (l) and biologically inspired algorithm (Omkar et al., 2008) that have been applied either pixel-based or object-based (Duro et al., 2012; Long et al., 2013). However, most of these supervised methods rely extensively on in situ data or on human interpretation of spectral signatures, making the classification process resource-intensive, time-consuming, and difficult to repeat over space and time (Zhong et al., 2014). These supervised methods generally outperform the unsupervised ones (Sharma et al., 2013) but their dependence on within-season training data appears as a major drawback for operational monitoring. These calibration sets are not available early in the season as the crops are not yet sown or have not yet emerged. In the field of crop mapping only few efforts targeted generalization (Zhong et al., 2014), i.e., the development of method applicable from year-to-year without being retrained. This idea of applying an automated algorithm to multiple years without further recalibration ties up to the idea of classifier extension proposed by Botkin et al. (1984) and developed for natural vegetation (Woodcock et al., 2001; Baraldi, 2011; Hestir et al., 2012) and recently for agriculture (Zhong et al., 2014). Nevertheless, methods have been developed to circumvent the recalibration issues or to foster the re-utilization of calibration data for agricultural applications such as time series kriging (Masse, 2013), dynamic time warping (Petitjean et al., 2012) and Markov logic models (Osman et al., 2015). Another shortcoming of field data driven algorithms is their inability to perform accurately if the season being mapped falls outside the range of those observed so far. In marginal years (delay in growth, flooding, ...) –those that matter most from a food security monitoring point of view and that are likely to increase because of global warming– their performances are expected to be limited: temporal profiles of unusual years need to be observed first to be included in the training. To cope with the above mentioned constrains, several initiatives relied on existing land cover information called existing baseline from which they extract the labeling or training data (Arino et al., 2007; Bontemps et al., 2012; Radoux et al., 2014).

Several studies investigated in depth which pertinent features to select from satellite images and vegetation index time series for crop discrimination because i) large numbers of feature might result in a loss of accuracy and ii) the larger number of features, the longer computing times (Cánovas-García and Alonso-Sarria, 2015). Some remained at the reflectance level (Akbari et al., 2006), others exploited principal components (Potgieter et al., 2007), harmonic components (Mingwei et al., 2008) or biophysical variables (Waldner et al., 2015d). Arvor et al. (2011) extracted pixel-level temporal descriptors of vegetation index
time series which in fact relates to the crop phenology. The texture information has also been investigated. It was found important to distinguish between permanent crops (Peña-Barragán et al., 2011). However, such textural features are scale-dependent (Emerson et al., 1999) and site-dependent; therefore, their interpretation varies from one agrosystem to another and their computation is time-consuming. Besides methods have been devised to suppress the within-class temporal variability (Conrad et al., 2011) which affect negatively the classification accuracy.

From an agriculture monitoring viewpoint, it is important to have a timely and accurate cropland mask over large areas including those where field data are unreliable or lacking for various reasons such as conflicts or accessibility. The overarching objective of this chapter is thus to propose a cropland classification algorithm that complies with the requirements of operational agriculture monitoring. Therefore, this research proposes a fully automated classification method that relies on the knowledge of the expected cropland temporal trajectories to determine temporal features to be used in the classification. Hence, these features have a straightforward interpretation that are consistent throughout the globe even if subjected to local variations. The classification method itself is a two-step method in which training data is extracted from existing land cover information, hence alleviating the burden of within season field data collection. The baseline land cover information is first cleaned of potential errors and is secondly used to train state-of-the-art classifiers. Relying on a baseline rather than on in situ data is of great value for both routinely mapping large areas and for regular temporal updates. The core of this chapter is composed of two parts: first the performance and accuracy of such a method is assessed with high spatial resolution imagery in four contrasted agro-systems; second, the generalization capabilities are investigated both from a year-to-year and along the season perspectives taking advantages of historical MODIS time series. It should be noted that this research adopted the guideline definition of annual cropland of the Joint Experiment for Crop Assessment and Monitoring (JECAM).

2.2 Materials

2.2.1 Remote Sensing Data

Four sites of 60x60 kilometers located in Argentina, Belgium, Ukraine and China were selected (see section 2.4). The remote sensing data is composed of imagery from three sensors: SPOT-4, Landsat-8 and MODIS. The high resolution imagery (SPOT-4 and Landsat-8) was acquired over all four sites to evaluate the performance of the classification method in different agrosystems. The coarse resolution data (MODIS) was only acquired over the Argentinian site to investigate the temporal generalization capabilities of the method.

SPOT-4 (20 m, every 5 days) surface reflectance data from February 2013 to May 2013 (Figure 2.1) were acquired during the SPOT-4 Take 5 experiment. This experiment consisted in bringing down the orbit of the satellite during five
2.2. Materials

57 months to reproduce the temporal resolution of the future Sentinel-2 mission. These data have been corrected from atmospheric effects, including adjacency effects and terrain effects. Clouds and associated shadows were removed from the original surface reflectance data with the multi-sensor atmospheric correction and cloud screening (MACCS) spectral-temporal processor (Hagolle et al., 2015). MACCS is based on multi-temporal method for cloud screening, cloud shadow detection, water detection as well as for the estimation of the aerosol optical thickness. Landsat-8 (30 m, every 16 days) surface reflectance data were acquired concomitantly from April to December 2013 (Figure 2.1). For the visible, near and short wave infrared bands, Landsat data underwent the same pre-processing as SPOT-4. The processing was enriched thanks to the additional Landsat-8 spectral bands: i) the 1.38 µm band enabled an enhanced detection of high and thin clouds and ii) the blue band provided an additional criterion to detect the aerosols –thanks to its quasi-constant relationship with the surface reflectances in the red wavelength above vegetation. The precision gain compensates for the loss due to the lower temporal frequency of Landsat-8 images. It should be noted that the thermal bands were not processed. The complementarity between SPOT-4 and Landsat-8 images was critical to grasp the beginning of the season. Yet in Argentina, the time series did not cover the period from December to February that is critical for crop development. For every spectral bands, the time series were smoothed using a spline filter.

Figure 2.1: Availability of the high resolution images and their temporal distribution over the four demonstration sites.

MODIS data were downloaded from 2005 to 2013 for the Argentinian site which exhibits a high field size/pixel size ratio. From daily quality controlled surface reflectance values of MODIS Terra and Aqua, 7-day mean composites (red, near-infrared) at 250-m were produced according to the procedure developed by Vancutsem et al. (2007). The mean compositing procedure produces spatially homogeneous cloud-masked composites with good radiometric consistency and does not require any model adjustment or additional parametrization (Bontemps et al., 2008).
2.2.2 Existing baselines and validation data

Instead of relying on in situ training data, the method privileges the use of so-called existing baselines. An baseline consists in the best possible land cover information available for each of the study sites (Table 2.1). This varies from crop information extracted from farmers’ declaration, to national land cover maps or to the information extracted from the global ESA-CCI Land Cover (http://www.esa-landcover-cci.org/). It must be noticed that for Belgium the baseline is a combination of multiple data sources. Crop types classes of the Land Parcel Identification System (LPIS) were converted to a binary cropland non-cropland legend that matched the JECAM definition. Then, it was rasterized and superimposed on the ESA-CCI Land Cover map.

Table 2.1: Source of the baseline data for each site. Training data are extracted from the baseline to calibrate the classifier.

<table>
<thead>
<tr>
<th>Site</th>
<th>Baseline</th>
<th>Epoch</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>ESA-CCI Land Cover</td>
<td>2010</td>
<td>300-m</td>
</tr>
<tr>
<td>Belgium</td>
<td>Land Parcel Identification System</td>
<td>2012</td>
<td>vector</td>
</tr>
<tr>
<td></td>
<td>ESA-CCI Land Cover</td>
<td>2010</td>
<td>300-m</td>
</tr>
<tr>
<td>China</td>
<td>GlobeLand30</td>
<td>2010</td>
<td>30-m</td>
</tr>
<tr>
<td>Ukraine</td>
<td>Land cover map provided by the site manager</td>
<td>2010</td>
<td>30-m</td>
</tr>
</tbody>
</table>

For the validation, independent in situ observations were available for each site thanks to the JECAM network. These were used for validating the resulting cropland masks. They consist in, for the cases of Argentina, Ukraine and China, in situ identification of agricultural fields and non-agricultural areas. These data were collected during the growing season of the year 2013. For Belgium, the validation data set was derived from the official farmers’ declaration of 2013. The in situ data sets were provided by the different site managers as a vector file, including the contours of the observed fields.

2.3 Methodology

2.3.1 Cropland Classification Method

The proposed methodology has three steps: 1) extract knowledge based temporal features for the satellite image time series, 2) clean the baseline of its potential misclassifications based on the temporal features and 3) classify the temporal features using the cleaned baseline as calibration data (Figure 2.2).

All steps operate at the pixel-level even though it has been shown that object-based classification offer a more generalized and more contiguous depiction of land cover, usually better matching human perception (Stuckens et al., 2000). In fact, the actual effects of object-based classifications seem to depend on the classes to be mapped (Dingle Robertson and King, 2011), on the classification methods (Duro et al., 2012) and the areas of interest (Matton et al., 2015);
2.3. Methodology

Figure 2.2: Methodology to derive the cropland map. Three steps are involved: 1) the extraction of the five temporal features (temporal synthesis of the reflectances observed when: a) the red is maximum, b) the NDVI is maximum, c) the NDVI was minimum, d) the maximum positive slope and e) minimum slope of the NDVI time series), 2) the cleaning of a land cover baseline to extract training samples and 3) the training and application of a SVM classifier on the temporal features.

Statistical significance of the differences appears also variable (Duro et al., 2012; Yan et al., 2006; Whiteside et al., 2011; Dingle Robertson and King, 2011). In addition, object-based classification is a time-consuming process which requires human intervention to select the appropriate segmentation parameters. The selection is generally carried out by means of trial and error, which appears prohibitive for frequent large scale mapping. Yet, it must be noticed that the method detailed here remains compatible with an object-based approach.

Knowledge-based feature extraction

With the development of state-of-the-art classifiers able to handle high-dimensional feature sets and therefore to cope with the curse of dimensionality, one could derive many features and let those algorithms mine the useful information by themselves (see Vieira et al. (2012)). Even for those algorithms, reducing the number of input features was shown to improve the classification accuracy (Löw et al., 2013). Therefore, this research focused on knowledge-based derived fea-
features to directly feed the classifier with key information to discriminate between what is cropland and not cropland. The overarching idea was to combine both the full discrimination potential proposed by the spectral bands of a sensor with the synoptic interpretation capabilities of the Normalized Difference Vegetation Index (NDVI). Four growing cycle characteristics were identified: typically, annual herbaceous crops i) grow on bare soil either resulting from a previous harvest or soil preparation, ii) have a higher growing rate than natural vegetation, iii) have a well-marked peak of photosynthetically active vegetation and iv) have a sharp reduction of green vegetation due to harvest or senescence. As bare soil has a high reflectance in the red (Tucker, 1979), the date of maximum red was also included after showing his great potential to discriminate bare soil after harvest (Table 2.2).

Table 2.2: Knowledge-based features and their interpretation. Two features target key moments preceding crop development, the three remaining features focus on characterizing the growth and senescence rate and the peak of photosynthetic activity.

<table>
<thead>
<tr>
<th>Compositing feature</th>
<th>Targeted crop cycle characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Red</td>
<td>Bare soil after harvest, before sowing</td>
</tr>
<tr>
<td>Maximum NDVI</td>
<td>Peak of photosynthetically active vegetation</td>
</tr>
<tr>
<td>Minimum NDVI</td>
<td>Minimum vegetation cover</td>
</tr>
<tr>
<td>Maximum Positive NDVI Slope</td>
<td>Rapid crop growth rate</td>
</tr>
<tr>
<td>Maximum Negative NDVI Slope</td>
<td>Rapid green vegetation reduction due to harvest/senescence</td>
</tr>
</tbody>
</table>

Following this conceptual framework, reflectances and NDVI time series were analyzed to translate those characteristics into temporal features (Figure 2.3). On a pixel basis, dates corresponding to the maximum red, the minimum NDVI as well as the maximum positive and negative slopes of the NDVI curve were extracted. In a second step, the corresponding reflectances (green, red, near infrared and shortwave infrared) coming from these dates were composited. The final temporal features are thus the reflectances observed when: a) the red is maximum, b) the NDVI is maximum, c) the NDVI was minimum, d) the maximum positive slope and e) maximum negative slope of the NDVI time series. For a given date, the slopes were derived from the NDVI values preceding and following it: \[ \text{Slope}_t = \text{NDVI}_{t+1} - \text{NDVI}_{t-1}. \] Figure 2.4 illustrate the identification of the feature compositing date.

Cleaning imperfect and outdated baselines

Several methods to mitigate of the issues hampering the use of land cover maps as a baseline have been devised such as iterative trimming to identify statistical outliers (Desclée et al., 2006; Radoux and Defourny, 2010) and erosion filters (Radoux et al., 2014). A major drawback of iterative trimming lies in the fact that it operates in a class-specific approach: in the case of a class dominated by mislabeled pixels, well-labeled pixels are consequently considered as outliers. The proposed approach assumes that a baseline is imperfect because it is affected by labeling errors or land cover changes –which is reasonable as most
2.3. Methodology

Figure 2.3: Procedure for the knowledge-based feature extraction. NDVI time series is first extracted from the input reflectance time series. Dates corresponding to the features are identified per pixel. The reflectance corresponding to those dates are finally composited.

Figure 2.4: Temporal NDVI profiles for two cropland (a,b), grassland (c) and forest (d). Reflectance was interpolated (green dots and lines) prior to the feature extraction (A: maximum value of NDVI; B: minimum value of NDVI; C: maximum slope; D: minimum slope). Source: Matton et al. (2015).
baselines are derived from remote sensing for a given epoch. The overarching idea of this cleaning step is to exploit spectral information from other classes to identify and discard the likely mislabeled pixels.

The algorithm builds a cleaned training set from a baseline land cover map and the extracted five temporal features. Operating class by class, a random sample of maximum 1000 vectors from class $i$ is drawn. As this set is imperfect, it contains not only samples correctly labeled as $i$ but in a lesser extent samples wrongly labeled as $i$. A second sample of twice the size of the first set is drawn among the remaining classes $\{j,k,l,\ldots,z\}$ regardless of their label. This second set is also assumed imperfect and contains samples correctly belonging to classes $\{j,k,l,\ldots,z\}$ but also samples belonging to $i$ even though they are not labeled as such.

Mislabeled samples can be identified and set aside because of their overlapping spectral signatures with other classes. Clustering the two sets together based on the temporal features and computing the probability of a cluster to be of class $i$ allows discarding those imperfect pixels labeled as $i$ but belonging to a $\{j,k,l,\ldots,z\}$ cluster. A self-organizing map clusters this data set and the probability of a cluster to be of class $i$ is thus computed. Kohonen self-organizing map (Kohonen, 2001) is a machine learning clustering method that handles high dimensions; it has already been applied successfully in land use/land cover classification (Ji et al., 2000). Samples describing class $i$ and belonging to clusters with a probability of belonging to $i > 50\%$ are set aside in a cleaned training database. The operation is then repeated for each class in the baseline.

Classification method

Once the baseline cleaned and the reliable samples automatically extracted from it, state-of-the-art classifiers can be trained. In this study, a Support Vector Machine (SVM: Vapnik (2000) was trained as it has been shown to generally outperform other classifiers (Huang et al., 2002a; Mountrakis et al., 2011; Pal and Mather, 2005). A Support Vector Machine uses a nonlinear mapping algorithm to transform the original training data into a higher dimension. In this new dimension, it searches for the linear optimal separating hyperplane, i.e. a decision boundary. With an appropriate nonlinear mapping, such a separating hyperplane always exists. The SVM identifies this hyperplane using support vectors ("essential" training vectors) and margins (defined by the support vectors). The cost-support vector classification was trained with Gaussian radial basis kernel functions whose widths were defined using heuristics (Caputo et al., 2002) in order to ensure a high level of automation. Note that the classification was applied on all the land cover classes present in the baseline and then reclassified in a binary cropland/non-cropland map. Even though suffering from certain drawbacks such as overfitting and the selection of the hyper-parameters, SVMs are particularly appealing in the remote sensing field due to their ability to generalize well even with limited training samples—a common limitation for
2.3. Methodology

remote sensing applications (Mountrakis et al., 2011)– and high-dimensional
data (Pal and Mather, 2005).

2.3.2 Accuracy assessment and spatial thematic uncertainty based on high resolution time series

*In situ* data were used to assess the accuracy of the classification method. Confusion matrices were computed for each site as well as standard accuracy measures such as the overall accuracy, the omission and commission errors (Congalton, 1991), the quantity and allocation disagreement (Pontius and Millones, 2011) as well as the F-score (Labatut and Cherifi, 2012).

In addition to the final class, classifiers such as SVMs can provide a vector of the class membership degree \( p = p^1, p^2, ..., p^M \) of a given pixel where \( M \) the number of class (Wu et al., 2004). Those membership degrees can be regarded as the probability of a pixel to actually belong to a given class (Löw et al., 2013). To evaluate the uncertainty of the classification, one could consider the membership degree of the final class solely but this measure does not consider the distribution of the class probabilities. To summarize all the information contained in \( p_{ij} \) and to commit the probabilities of the other classes in the uncertainty evaluation, an uncertainty measure was computed using the entropy of the class probability distribution (Dehghan and Ghassemian, 2006). For a given pixel, the Normalized Uncertainty Criterion (NUC) based on the above entropy criterion reads as follow:

\[
\text{NUC} = 1 - \frac{\log_2(M) - \text{entropy}(p)}{\log_2(M)}
\]

where \( M \) is the number of classes and \( p \) the vector of class membership and is bounded between 0 and 1. For low entropies, the performance of classifier is good and the associated uncertainty of the results is low, i.e. a pixel with a low entropy criterion is powerfully labeled to a class and conversely. Hence, NUC gives a measure of doubt that can be used to quantify the thematic map uncertainty.

2.3.3 Temporal robustness analysis based on MODIS time series

As agriculture monitoring requires timely information, it is of paramount importance that classification methods can ingest data in near real time and update the cropland map as within-season information accumulates. This section places the method in an operational context and seeks to provide answers to two main questions: 1) how can the method be adapted so that it may be used from one year to another without recalibrating the parameters of the SVM, and 2) how robust is the method from one month to another without recalibration. As the high resolution time series was only available for 2013, MODIS was utilized instead as it offers a long history of observation. Note that the generalization capabilities of the method are investigated over the Argentinian site only.
Year-to-year generalization

The knowledge-based features are designed to remain similar from one year to another. However, slight differences might occur in the reflectances of the features due to changes in the acquisition dates, cloud conditions, residual errors of the cloud, shadow and snow masks. It was thus assumed that deriving the features on interpolated reflectances rather than on the measured reflectances would further stabilize the features through time. In this case, “interpolated” means that instead of identifying key dates from the observed time series, e.g., the date corresponding to the maximum red), the key dates are extracted based on a daily interpolated time series derived from the observed time series and based on cubic spline model. To verify this assumption, two sets of features (observed and interpolated) were derived from the 2013 MODIS time series. The first set of features was derived following the approach detailed in section 2.3.1, the second by extracting the feature based on the interpolated spline model.

A McNemar’s test, a test for the difference of two proportions, was used to evaluate the impact of the interpolation. The McNemar’s test assesses if the difference in accuracy by classifying on interpolated reflectance data was statistically significant. It is based on a $\chi^2$ test with one degree of freedom for goodness-of-fit that compares the distribution of counts expected under that null hypothesis to the observed counts. This test relies on a contingency table presenting the number of misclassified samples by one classifier, by both and well classified in both. It incorporates a continuity correction term to account for the fact that the statistic is discrete while the $\chi^2$ distribution is continuous:

$$\chi^2_{0.95,1} = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}}$$

where $n_{01}$ is the number of samples misclassified by classifier A but not by classifier B and $n_{10}$ the converse. If the null hypothesis is correct (the error rate are the same, i.e. $n_{01} = n_{10}$), then the probability that this quantity is greater that $\chi^2_{0.95,1} = 3.84$ is less than 0.05. So if $\chi^2_{0.95,1}$ is larger than 3.84, then with 95% confidence, one can reject the null hypothesis that the two classifiers have the same error rate. If no significant difference is found, the direct implication is that one could safely use interpolated features and use the same classifier years after years without retraining.

Two cropland maps were derived for 2013, one with the measured features and the other with the interpolated features following a cubic spline scheme. Then, cropland maps were then produced from 2005 to 2012 using the classifier derived for 2013 without recalibration. The overall accuracy and the commission and omission errors were also assessed for the different years of interest and compared to those obtained with year-specific algorithms.

Near real time generalization

If derived in a monthly moving window fashion, the knowledge-based temporal features are also expected to remain stable along the season because the entire
12-month crop cycle would constantly be represented which should limit their variations. To verify this assumption, the cropland mapping methodology was applied at the end of the 2012 growing season and the resulting classifier was applied on a monthly basis along the 2013 season. The resulting maps were compared to those obtained with classifiers trained specifically for each month.

### 2.4 Study sites

To illustrate the generic character of the approach, four sites of 60x60 kilometers belonging to the JECAM network† and located in diverse agrosystems were selected (Figure 2.6). The overarching goal of JECAM is to reach a convergence of approaches, develop monitoring and reporting protocols and best practices for a variety of global agricultural systems. It enables comparing results over a variety of global cropping systems.

![Figure 2.5: Distribution of the four sites of interest. The location of the sites is overlaid on top of the Unified Cropland Layer in green.](image)

In the first site located in Belgium, the typical field size ranges from 3 to 15 ha and the dominant crop types are wheat, barley, potatoes, sugar beet, and corn. Winter crops are sown in October and harvested in August at the latest whereas summer crops are sown in April and harvested in September. The landscape topology is flatlands and hills. The climatic zone is temperate with annual rainfall of about 780-mm and relatively well distributed over the year therefore irrigation infrastructure is not frequent.

The second site in Ukraine covers the administrative region of the Kyiv oblast. The Northern part of the region is dominated by forests and grasslands, while its central and southern parts are dominated by agriculture. The climate

is humid continental with approximately 700 mm of annual precipitations. Flat terrain dominates the landscape, only near 10% of the territory is hilly. Fields in the region are quite large with size generally ranging up to 250 ha. The crop calendar starts from September to July for winter crops, and from April to October for spring and summer crops. Dominant crop types include maize (25.1% of total cropland area in 2013), winter wheat (16.1%), soybeans (12.6%), vegetables (10.3%), sunflower (9.3%), spring barley (6.8%), winter rapeseed (4.0%), and sugar beet (1.3%). It should be noted that vegetables are mainly (approximately 96%) produced by small farms and around villages for self-consumption purposes (so-called family gardens (Gallego et al., 2014)). Those fields have very limited size (less than 0.1 ha) and are thus discarded by the definition. There is no typical simple crop rotation: it depends on specialization.

The third site is located in the Northwest Shandong province in China (Jia et al., 2012). This area has a temperate, semi-arid monsoon climate, with mean annual temperature of 13.1° and precipitation of 582 mm concentrated from

---

**Figure 2.6:** Zooms of the four study sites. The imagery provided is a 20-m false color composite derived from SPOT-4.
late June to September. Land use is dominated by agriculture, forestry and residential uses. The typical field size ranges from 0.2 to 0.8 ha. The typical crop rotation is winter wheat followed by summer maize. Winter wheat is sown in early October and harvested in early or mid-June the following year. Summer maize is sown in mid-June and harvested from the end of September to early October. The rotation is then repeated (Meng et al., 2013). Tillage practices vary from intensive tillage with very low residue cover to conservation tillage (including no-till) with little disturbance of the residue.

The fourth site is located in the Pampas, a region of Argentina with gentle slopes (lower than 3%) and rivers. Climate is humid temperate with an isohygro precipitation regime and annual mean of about 1000 mm. Most of the land is dedicated to agriculture both crop and forage, grassland is also present in a lesser extent. Main grain crops are soybean, maize and wheat. Wheat is planted from June and its harvest starts at the beginning of December. Soybean could follow in the rotation (sown in December and harvested in April, thus producing two crops in one season). As one-season crop, soybean is planted in November and harvested in March/April and whereas corn is planted in October and harvested in March (Veron and de Abelleyra, 2014). Late maize crops can be planted in December after a fallow. Agriculture is developed under no-till systems (only a reduced till is done together with planting leaving crop residues over soil surface) and mostly without irrigation. Typical field size is 20 ha but the field variability in size is high.

2.5 Results

2.5.1 High resolution temporal features

Five temporal features were extracted from the SPOT-4 and Landsat-8 time series (Figure 2.7). They target key phases of the crop cycle such as 1) the bare soil after harvest or before sowing, 2) the peak of photosynthetically activity, 3) the minimum vegetation cover, 4) the growth rate and 5) the green vegetation reduction due to harvest or senescence.

Cropland appears clearly distinct from other classes. Indeed, there is a high match rate between the cropland class (in white on Figure 2.7) and the blueish areas on the minimum NDVI feature which corresponds to bare soil. The correlation between features seems to vary from site to site (Figure 2.8). Cropland samples display a stronger between feature correlation than non-cropland samples. Depending on the overlap between the time series and the crop cycle, specific features tend to give a homogeneous response over the cropland regardless of the crop types. For Belgium, the time series does not cover the end of season of the summer crops. As a result, some field looking as fallow in the maximum NDVI feature (blue shades) have not yet reached their maximum NDVI value. The opposite is observed in Argentina where the maximum red and minimum NDVI underestimate cropland as the start and the end of the cycle. These phenomena are dependent of the time series length.
Figure 2.7: Baseline map and false color infra-red knowledge-based features over the Belgian site. The white class in the baseline map corresponds to cropland. It generally occurs very distinctly from the other class in the features.
2.5. Results

Figure 2.8: Scatter plot of red band of the five knowledge-based features for (a) Argentina and (b) China. maxNDVI.red, maxRed.red, maxSlope.red, minNDVI.red and minSlope.red correspond to the red band of the maximum NDVI feature, the maximum red feature, maximum slope feature, the minimum NDVI feature and the minimum slope feature, respectively. The correlation between features seems to vary from one site to another. The feature correlation seems higher for the cropland class than for the non-cropland class.
and are a source of additional noise for the classification. In general, features based on slopes are more sensitive to noise and might produce patchy results.

### 2.5.2 Cropland classification on high spatial resolution features

Cropland maps were generated with a support vector machine classifier trained on knowledge-based temporal features derived from SPOT-4 and Landsat-8 time series and a baseline after a statistical cleaning (Figure 2.9). The accuracy of the classification maps was assessed with independent validation sets. The overall accuracies range from 80 to 92% and the F-scores of both classes exceeded 80% with the exception of the non-cropland class in Ukraine (Table 2.3). Besides, quantity and allocation disagreement were found balanced for each case, i.e. the amount of over and under-estimation was similar to that of mismatched pixels in each mapped class. It should also be noted that the accuracy does not increase with the average field size.

| Table 2.3: Accuracy indices of the high resolution classifications: Overall Accuracy (OA), Quantity Disagreement (QD), Allocation Disagreement (AD), Producers’ Accuracy (PA), Users’ Accuracy (UA). The overall accuracies range from 80 to 92% and the F-scores of both classes exceeded 80% except for the non-cropland class in Ukraine. Besides, quantity and allocation disagreement were found balanced for each case. |
|---|---|---|---|---|---|
| Site | OA | QD | AD | PA | UA | F-score |
| Argentina | 80.03 | 10.01 | 9.95 | crop | 88.80 | 72.47 | 79.81 |
| | | | | non crop | 73.00 | 89.06 | 80.23 |
| Belgium | 92.12 | 3.26 | 4.60 | crop | 96.18 | 91.24 | 93.64 |
| | | | | non crop | 85.96 | 93.67 | 89.64 |
| China | 89.03 | 4.81 | 6.15 | crop | 75.68 | 88.86 | 81.74 |
| | | | | non crop | 95.44 | 89.09 | 92.15 |
| Ukraine | 89.05 | 5.47 | 5.48 | crop | 96.77 | 90.91 | 93.75 |
| | | | | non crop | 45.80 | 71.68 | 55.89 |

Argentina appears as the site with the lowest overall accuracy (80%) which might be explained by the poor cropland delineation in the baseline (Figure 2.9a) as well as by the mismatch between the period covered by the time series and the crop calendar. Nonetheless, the method succeeds in discarding non-cropland areas—mainly grassland—labeled as cropland in the baseline. There is a remarkable increase in the natural vegetation class compared to the baseline which fits the actual land cover patterns. Commission errors for the cropland also include fallow fields. This site was the most affected by the salt and pepper effect due to pixel-based classification and might benefit from an additional spatial filtering step.

Despite a small average field size, a large crop diversity and a less dense time series due to the persistent cloud coverage, results over Belgium indicate an overall accuracy of 92.12%. Omission and commission errors are also
2.5. Results

![Baseline Argentina](image1)

![Map Argentina](image2)

![Uncertainty Argentina](image3)

![Baseline Belgium](image4)

![Map Belgium](image5)

![Uncertainty Belgium](image6)

![Baseline China](image7)

![Map China](image8)

![Uncertainty China](image9)

![Baseline Ukraine](image10)

![Map Ukraine](image11)

![Uncertainty Ukraine](image12)

**Figure 2.9:** Baseline, classification map and uncertainty map for the four sites.
Chapter 2. Automated Cropland Mapping

balanced. Such good results might be attributed to the quality of the cropland baseline derived from the Land Parcel Identification System (LPIS). However, the LPIS delineates only the cropland and the remaining land cover information was derived from the land cover CCI which has a much coarser spatial resolution.

In China, the accuracy of the cropland map reaches 89.03% and the cropland class is more subject to both omission and commission errors than the non-cropland class. The cropland spatial distribution is consistent with the site's general land cover patterns; compared to the baseline the cropland map gains in delineation of the built-up and water classes.

The cropland map of the Ukrainian site yields an overall accuracy of 88.49% yet accuracy measures display that its non-cropland class is the poorest compared to the other sites. Producer's and user's accuracies are 96.77% and 90.91% for the cropland class and 45.80% and 71.68% for the non-cropland class. The cropland class has considerably higher errors of omission which suggested some difficulty in discriminating cropped areas from non-crop vegetation. In fact, most of the commission errors are due to fallows; forest, grassland and urban areas are generally well discriminated on the map—but not sampled proportionally to their occurrence in the validation data set.

The performance of the classification was further assessed by looking at the normalized uncertainty criteria outputs. The thematic uncertainty of the final crop maps occurred in the same order of magnitude across sites (Figure 2.9). These uncertainty maps can serve to allocate and discuss the causes of possible classification errors. A clear and similar spatial pattern appeared among the four sites, yet in a lesser extent in Argentina: the cropland class is characterized by low uncertainties whereas the non-cropland is characterized by higher level of uncertainty. This contrasted situation is more diffuse in Argentina (Figure 2.9c). This needs to be related to the fact that the temporal features were designed for cropland recognition. Thus, cropland is clearly discriminated from the non-cropland classes but the separability among those non-crop classes is low which does not matter when focusing on cropland mapping solely.

To evaluate the classification confidence regarding the thematic uncertainty, one might simply test if the correctly classified pixels are characterized by low uncertainty values and conversely. If this assumption is verified, the uncertainty values of correctly classified cases should be equal to or lower than the uncertainty values of the misclassified cases (Zhu, 1997). Empirical frequency distribution of normalized uncertainty criterion were extracted for correctly classified validation fields (Figure 2.10a) and omitted validation fields (Figure 2.10b). The shape of these distributions gives an indication on the reliability of the algorithm and the quality of the cropland maps, as proposed and demonstrated for decision trees (McIver and Friedl, 2001), random forest (Loosvelt et al., 2012), and support vector machine (Löw et al., 2013). Correctly classified pixels tend to display relatively low values of uncertainty but the degree to which this is verified differs from one site to another. Overall these figures confirm the
2.5. Results

high classification confidence of the algorithm over all sites and indicates that it already operates at a high level of confidence, with a limited potential for significant improvements. The largest difference between the four sites occurs in Argentina where the 40-60% interval is the most densely populated. Regarding the pixels wrongly classified as non-cropland, one would expect those to have high uncertainty values. The present case respect this expected behavior but in a less pronounced fashion and with exceptions in China and Belgium that show peaks for the 0-20% and the 20-40% interval. Overall, the quality of the baseline seems to affect more strongly the classification strength than the label of the classification itself.

![Figure 2.10: Distribution of the uncertainty criterion (NUC). Correctly classified pixels tend to display relatively low values of uncertainty but the degree to which this is verified differs from one site to another.](image)

2.5.3 Temporal generalization with coarse resolution time series

To assess the ability of the method to generalize in time, it was applied on MODIS time series at 250 m over the Argentinian site. To further reduce the time series variability in time, it was assumed that the feature extraction would be more stable in carried out on simulated reflectance rather than on the observed reflectance. If this assumption holds true, the difference in accuracy between the two approaches should not be significantly different. Prior to applying this approach along the season or on multiple year, a non-parametric McNemar’s test was used as a direct comparison method for testing whether classifications with the two types of features differed significantly among themselves (Table 2.4). Statistically, based on the McNemar’s test, the two classification algorithms do have the same error rate at significance level \( \alpha = 95\% \) because this value was lower than \( \chi^2_{0.05,1} = 3.84 \). Hence, interpolated features could be extracted
Chapter 2. Automated Cropland Mapping

and used without deteriorating the classification accuracy. To evaluate the multi-year generalization potential, a classifier was trained on features extracted for 2013 and applied back in time on features extracted from 2005 to 2012. The overall accuracy ranged from 76 to 91%; confusion and omission errors varied generally from 10 to 25% for both the crop and non-crop classes (Figure 2.11a). This variability might also result from MODIS data itself, noisier than the high resolution imagery because of the variability of the pixel purity resulting from its sensor (Duveiller et al., 2011).

Table 2.4: Confusion matrices for the MODIS classification on the observed and interpolated reflectances and the associated McNemar’s test. McNemar’s test ($\chi^2 = 0.2748$ and $p$-value = 0.6001) led to the rejection of the hypothesis that the two classifications significantly differ from one another.

(a) Confusion matrix observed reflectances

<table>
<thead>
<tr>
<th></th>
<th>Non-cropland</th>
<th>Cropland</th>
<th>UA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-cropland</td>
<td>593</td>
<td>32</td>
<td>94.88</td>
</tr>
<tr>
<td>Cropland</td>
<td>289</td>
<td>1016</td>
<td>77.85</td>
</tr>
<tr>
<td>PA</td>
<td>67.23</td>
<td>96.94</td>
<td>OA: 83.37</td>
</tr>
</tbody>
</table>

(b) Confusion matrix interpolated reflectances

<table>
<thead>
<tr>
<th></th>
<th>Non-cropland</th>
<th>Cropland</th>
<th>UA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-cropland</td>
<td>631</td>
<td>63</td>
<td>90.92</td>
</tr>
<tr>
<td>Cropland</td>
<td>251</td>
<td>985</td>
<td>79.69</td>
</tr>
<tr>
<td>PA</td>
<td>71.54</td>
<td>93.99</td>
<td>OA: 83.73</td>
</tr>
</tbody>
</table>

(c) McNemar’s confusion matrix

<table>
<thead>
<tr>
<th></th>
<th>Obs. true</th>
<th>Obs. false</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int. true</td>
<td>252</td>
<td>62</td>
</tr>
<tr>
<td>Int. false</td>
<td>69</td>
<td>1547</td>
</tr>
<tr>
<td>McNemar’s $\chi^2 = 0.2748$</td>
<td>$p$-value = 0.6001</td>
<td></td>
</tr>
</tbody>
</table>

For the classification along the season without recalibration in a moving window fashion (Figure 2.11b), the overall accuracy oscillates between 72% and 78%. It should be noted that the lower accuracy results compared to the high resolution might be attributed to the coarser spatial resolution and to MODIS’ limited number of bands at 250 m. Its minimum occurs in December and is associated with an overestimation of the cropland class around the time of the sowing of summer crops. Compared to the year to year experiment in which commission and omission errors were balanced, there is a clear tendency towards cropland overestimation. As the landscape switches from winter crops to summer crops, the noise in the features might increase as a result of mixed pixel information especially for the features related to the slope. The magnitude of this phenomenon should be reduced with high spatial resolution images.
2.6 Discussion

Efficient annual cropland mapping approaches for operational crop monitoring must comply to several requirements such as timeliness, accuracy, a high degree of automation and cost-effectiveness. The proposed methodology has been applied on four contrasted sites and was found able to deliver accurate cropland maps without *in situ* data. The process is fully automated and reached accuracies around 80% or more in all four sites. Yet, challenges were apparent to accurately distinguish fallows and in a lesser extent grassland during the growing season. This might be achieved by improving the temporal coverage of the time series and their density. As the method handles well high dimensionality, perspectives of improvements include considering more spectral bands and de-

![Multi-year generalization](image1)

![Along the season generalization](image2)

**Figure 2.11:** Generalization accuracy of the method: a) accuracy over time for a classifier trained for 2013 and applied to the 2005-2012 period, b) accuracy over time when the method is applied along the season on a 12-month moving window. Indices with the “recal” suffix corresponds to the benchmark method that was re-calibrated for each year or month.
developing new temporal features to better capture local specificities. While only four spectral bands were used, recent and upcoming sensors offer a larger range of channels. Landsat-8 collects information in nine shortwave spectral bands and two thermal bands (Irons et al., 2012) and Sentinel-2 will carry thirteen spectral bands spanning from the visible and the near infrared to the short wave infrared (Drusch et al., 2012). However, the strong relationships observed between the knowledge-based features might suggest multicollinearity. Multicollinearity occurs when complex correlation patterns exist among variables, which leads to several undesirable consequences. This multicollinearity decreases the intrinsic dimensionality of the dataset, which would mean that fewer observations are required and can lead to a degradation of the prediction accuracy. Generally, feature reduction (such as principal components transformation and partial least squares) or feature selection (stepwise linear regression, band shaving (Clevers et al., 2007; Verzakov and Duin, 2004)) techniques are applied to overcome these problems (Bruzzone and Serpico, 2000). The advantage of feature selection is that the derived features are easily interpretable from a physical point of view. Nonetheless, previous work suggests that SVMs are capable of handling a high degree of collinearity (Howley et al., 2006; Kotsiantis, 2007). Additional steps might be considered to further improve the accuracy of the cropland map such as Markov Random Fields (Kasetkasem et al., 2005) or filtering based on the class membership.

This study demonstrated the ability to extract automatically training samples to capitalize on the available cropland information and alleviate the burden of within-season data collection. The quality of the baseline was found to affect more the certainty of the classification rather than the result of the classification itself: the more detailed the baseline the more certain the classification. In fact, Pelletier et al. (2017) showed that SVM was little influenced for low random noise levels up to 25%-30%, but their performances drop down for higher noise levels. Thus, this suggests that the concepts developed here can be extended to other geographic regions such as the priority areas for cropland mapping identified by Waldner et al. (2015c) or even globally. The application to a larger scale could gain efficiency by replacing the random sampling by a SVM-specific sampling strategy as proposed by Mathur and Foody (2008). Candidate land cover data sets to serve as baseline are numerous. A key parameter in their selection is the thematic adequacy of the cropland class (Waldner et al., 2015c). Baselines with different legend definition, coarse spatial resolution or out-of-date might be considered thanks to the cropland specific temporal features and the cleaning step. Since the uncertainty of the method depends on the initial land cover map, it means that the classification might provide spatially heterogeneous maps according to the available input data. One of the major challenges in upsampling the annual cropland mapping to the regional, continental or global scales is the availability of appropriate validation data. Indeed, the validation samples need to be renewed annually as fields could be set aside during the next season depending on crop practices.
2.7. Conclusions

In the framework of food security, cropland maps are required to mask the agricultural areas for crop condition monitoring and should be therefore available early in the season. In a similar direction as for the seasonal cropland mapping advanced by Wu et al. (2014b), the presented methodology tackles this issue in three ways. First, knowledge-based features were designed and it was shown that they can be extracted with stability along the season (exhibit a high stability throughout the season and from year to year). To further increase the robustness, features were extracted at interpolated dates rather than at an imaged date—as dates of cloud-free acquisition changes from year to year. For large scale mapping, such features offer also a simple and comprehensive framework to integrate images from different orbits without losing temporal details. Indeed, as the method operates the compositing of the features at the pixel level, it tolerates time series of different lengths which would increase temporal resolution and consequently the feature extraction. Other state-of-the-art methods could also be implemented to derive the features (see Zhu et al. (2015a)). Second, thanks to the generalization capabilities of the classifier, it might be trained once and achieve reasonable accuracies (70%) without being retrained which ensures maps delivery with a minimal time-lag to inform decision makers. If in situ data happen to be available for a specific area, they could easily be ingested by the method and utilized repeatedly for multiple years without the need for year-to-year reformulation, mapping cost would be reduced substantially and the timeliness of the map products would be improved. Third, the dependence of most existing methods on ground truth data for algorithm calibration leads to considerable cost in time and labor resources. This study successfully demonstrates that this deleterious dependence can be alleviated using already available land cover information. The approach along the season can be fully automated allowing its operational use to generate annual cropland maps with good agreement to the validation data.

2.7 Conclusions

In the context of food security, global, timely, accurate and cost-effective cropland mapping is a prerequisite for accurate crop condition monitoring. This chapter presented a methodology that delivers high accuracy cropland maps (>80%) by introducing 1) five knowledge-based temporal features that remain stable over time and 2) a cleaning method that identifies pixels of baseline land cover maps that can be used from training a new classifier. Besides, the quality and accuracy of the baseline seem to have a stronger effect on the classification certainty than on the classification outcome itself. The independence of ground truth data makes this approach particularly promising for regular mapping over large areas. This potential for large area mapping has still to be verified and will be the main topic of the next chapter. The knowledge-based features are also quite stable over time both from year to year and along the season. It indicates that re-calibration of the classifier from one year to another or from month to month is not necessary. This behavior could also be particularly interesting to enable the re-use of ground truth data.
Chapter 3

National-scale cropland mapping based on spectral-temporal features and historical land cover information

Highlights

- A method is presented for national-scale cropland mapping based on multi-year imagery and historical land cover information.
- Spatial consistency is achieved by using seamless input features, stratifying the landscapes to reduce the within-class spectral diversity, and fusing stratum-specific map based on pixel-level class membership.
- The overall accuracy is 92% but a province-level accuracy assessment as well as a validation using spatially constrained confusion matrices revealed large variation across the landscapes.
- Accuracy was the lowest in smallholder farming systems.

Abstract. The lack of availability of ground truth data has always been a constraint for supervised learning, thereby hindering the generation of up-to-date thematic maps from of satellite imagery. This is all the more true for those applications requiring frequent updates over large areas such as cropland mapping. Therefore, we present a method enabling the automated mapping of spatially consistent cropland maps at the national scale based on spectral-temporal features extracted from Landsat time series and historical land cover information. This method extracts reliable calibration pixels based on their label in the historical map and their spectral signature. To ensure spatial consistency and coherence in the map, we propose several data preparation, classification and post-processing steps. First, we generate seamless input images by normalizing the time series and
Chapter 3. National-scale cropland mapping

deriving spectral-temporal features that target salient cropland characteristics. Second, we reduce the spatial variability of the class signatures by stratifying the country and by classifying each stratum independently. Finally, we further remove speckle thanks to a weighted majority filter. Capitalizing on a wall-to-wall validation data set, the method was tested in South Africa using a 16-year old land cover map. The overall accuracy reached 92% but a spatially explicit validation revealed large variations across the country. Results suggest that intensive grain production areas were better characterized than smallholder farming systems. Besides, we found that the informative features in the classification process vary from one stratum to another. Nonetheless, features targeting the minimum of vegetation as well as the short wave infrared features were consistently important throughout the country.

3.1 Introduction

South African households’ vulnerability to hunger has declined in the past ten years from 24% to 12% in 2011 (Stats, 2012; Labadarios et al., 2011). In 2013, 2.8 million households –comprising 11 million people– were deemed food insecure (Government Communication and Information System, 2013). The measures and programs initiated by the South African government appear beneficial even though Aliber and Hart (2009) argued they could be run more effectively. They emphasized that the lack of access to land must be addressed through sustainable, income-independent measures, such as the promotion of subsistence farming. Besides, progress in achieving food security is in jeopardy as the agriculture sector faces considerable impact from climate change: South Africa, on average, has been hotter and drier during the last 10 years than during the 1970s. Those changes in climate and water use affect the livelihoods of the vast majority of people, especially those already vulnerable (Government Communication and Information System, 2013). Blignaut et al. (2009) employed an econometric model to estimate how sensitive the nation’s agriculture may be to changes in rainfall. For the country as a whole, they concluded that each 1% decline in rainfall is likely to lead to a 1.1% decline in the production of maize and a 0.5% decline in production of winter wheat. Reducing risk through raising awareness as well as strengthening early warning systems and warning dissemination helps to build resilient farming communities. Therefore, the Department of Agriculture, Forestry and Fisheries has developed and implemented an Early Warning System disseminating extreme weather warnings (Government Communication and Information System, 2013).

Up-to-date and dependable satellite-derived cropland maps are one crucial element of early warning systems because they allow subsequent analyses such as crop inventory, crop status assessment, and yield forecasting to focus on actually cropped areas. Operational cropland mapping must comply to several requirements such as timeliness, accuracy, automation and cost-effectiveness (Waldner et al., 2015c). A critical limitation to achieving timeliness and cost-effectiveness is the availability of in situ data to calibrate supervised classifiers. The reliance
on within-season *in situ* data or on human interpretation of spectral signatures makes the classification process resource-intensive, time-consuming, and difficult to repeat over space and time. Several strategies were devised to cope with the limited availability of calibration data such as increasing the amount of field data by identifying homogeneous regions around based on aerial photography (Mannel et al., 2006) or by implementing positive and unlabeled learning algorithms (Chen et al., 2015b; Guo et al., 2012). These one-class classifiers are particularly interesting as the cost of unlabeled samples tends to zero and can thus have a much larger size than the positive sample set. Extracting calibration data from existing land cover maps (Jiang et al., 2012; Radoux et al., 2014; Waldner et al., 2015a) is especially interesting because such maps are already available globally.

Another challenge for national-scale cropland mapping is to achieve spatial continuity and consistency in the final map. There are two main sources of spatial inconsistencies: heterogeneity in the imagery (different orbits, acquisition dates, cloud/shadow contamination) and within-class spectral variability due to changes in environmental conditions, management decisions and practices. Given the amount of data required to cover large areas, this heterogeneity is likely be propagate in higher level products. Efficient strategies to cope with remote sensing image data heterogeneity and spectral variability are therefore crucial. One strategy to reduce the spectral variability is to derive temporal or spectral-temporal features from the time series (Löw et al., 2013; Zhong et al., 2014; Waldner et al., 2015a; Matton et al., 2015). Spectral-temporal features are composites of the spectral reflectances measured at a specific stage in the season. They summarize events that did not necessarily co-occur in composite images. This facilitates the discrimination between classes by reducing the within-class heterogeneity and improves the classifier’s extendability (Waldner et al., 2015a; Zhong et al., 2014). Drawbacks of spectral-temporal features are related the amount of available cloud-free images and their quality. Dense time series are required to be able to extract stable spectral signatures at the key-moments in the season. Besides, poor cloud/shadow screening results inevitably to noisy features. Classifiers’ accuracies are affected by the landscape diversity over large areas (Pelletier et al., 2016). In fact, the specific characteristics of the agro-systems to be mapped tend to have a stronger influence on the classification accuracy than the classification methods themselves (Waldner et al., 2016). Therefore, a second strategy to achieve spatial consistency is to stratify the area of interest, e.g., by agro-environmental conditions, and to calibrate stratum-specific algorithms (Vintrou et al., 2012a; Lambert et al., 2016; Bartalev et al., 2016). Nonetheless, local training is generally achieved at a higher processing cost and achieving seamless transitions between strata can be challenging (Radoux et al., 2014).

With the dearth of *in situ* data and the requirement of achieving spatial consistency as a backdrop, we present a method to derive automatically national-scale cropland maps based on multi-sensor Landsat time series and outdated land cover information. Given the merits of the aforementioned strategies for
large-scale cropland mapping, we detail i) how consistent spectral-temporal features can be derived from high resolution time series to capture the salient characteristics of cropland, ii) how calibration data can be selected from a historical land cover map, and iii) how the classifiers’ soft outputs can be used to fusion stratum-specific classifications and to improve the majority spatial filter. We tested the method in South Africa to capitalize on a wall-to-wall validation data set, i.e., field boundaries, as well as to assess the method performance in space, document the errors and identify the drivers of accuracy. It is worth noting that we do not present a product but a procedure for mapping national-scale cropland maps in a consistent and reproducible way.

3.2 Study area

South Africa is located at the southern tip of Africa and lies between latitudes 22° and 35°S, and longitudes 16° and 33°E spreading over 1,221,037 km² (Figure 3.1). The country is divided in nine provinces and has a wide variety of climates ranging from arid to sub-tropical, temperate or Mediterranean. The agricultural economy is a dual, with both well-developed commercial farming and more subsistence-based production in the remote rural areas. The dominant activities range from intensive crop production and mixed farming in areas characterized by winter rainfall and high summer rainfall, to cattle ranching in the bushveld and sheep farming in the arid regions (Figure 3.1). About 12% of the territory can be used for crop production but only 22% of this is of high-potential. The main growing regions lie along the more fertile soils of the Western Cape valleys and the KwaZulu-Natal province in the West. Agricultural systems have been primarily developed under arid and semi-arid climatic conditions where droughts are common (Bennie and Hensley, 2001). Irrigation agriculture is by far the largest consumer of water (Blignaut et al., 2009), and is responsible for 30% of the total crop production (Bennie and Hensley, 2001). The majority of grain is irrigated under center-pivot systems, and in many cases based on a double cropping rotation with winter wheat followed by summer maize. Most of the dryland crop production occurs in the semi-arid zones that can be divided into winter and summer rainfall regions.

The largest area of cropland is planted with maize, followed by wheat, and to a lesser extent sugarcane and sunflower (Government Communication and Information System, 2013; Department of Agriculture and Fisheries, 2011). It is estimated that over 8,000 commercial maize producers are responsible for the majority of the South African crop (10.8 Mt of maize produced in 2011/12 on 2.7 million ha of land), while thousands of small-scale producers are responsible for the rest. The “maize quadrangle” in the North West Province and northwestern Free State produces 75% of the country’s maize. Half of the production consists of white maize for human food consumption. Wheat is produced mainly in the winter rainfall areas of the Western Cape and the eastern parts of the Free State (2.0 Mt produced on 0.6 million ha in 2011). Sorghum is cultivated mostly in the drier parts of the summer rainfall areas such as Mpumalanga, the Free
3.3 Data

3.3.1 Satellite data pre-processing

All Landsat-5, -7 and -8 data from 2013 to 2015 falling into the area of interest were acquired and pre-processed following the procedure implemented in Hansen et al. (2013); Potapov et al. (2014, 2015). Note that four tiles (path/row: 176/77, 175/78, 174/78, 174/79) were discarded because no crop production occurs there. Four spectral bands were kept: the red, the near-infrared (NIR), and the two short-wave infrared (SWIR) bands. The blue and green bands were discarded due to their sensitivity to atmospheric effects (Ouaidrari and Vermote, 1999). We applied a three-step procedure to normalize the radiometry. First, Landsat data were converted to top-of-atmosphere reflectance (Chander et al., 2009) and then normalized by taking the corresponding MODIS top-of-canopy reflectance data as target (Potapov et al., 2012). Third, we adjusted cross-track surface anisotropy effects by modeling the Landsat reflectance per spectral band as a function of the viewing angle (Hansen et al., 2008; Potapov et al., 2012;
3.3.2 Historical land cover map, validation and ancillary data

The National Land Cover (NLC) 2000 map was generated from Landsat imagery acquired primarily from 2000-2001 (Van den Berg et al., 2008). It describes the South African territory with 45 land cover classes and an accuracy of 66% (Van den Berg et al., 2008). The minimum mapping unit is 2 ha (approx. 22 Landsat pixels). We translated the NLC-2000 native legend into a simplified ten-class legend: cropland, irrigated cropland, forest, shrubland, wetland, urban, bare soil, and water bodies.

For validation, the 2014 national field boundary data set was sourced from the Department of Agriculture, Forestry and Fisheries. It was created by digitizing all fields throughout the country based on 2.5 m resolution, pan-merged SPOT-5 imagery acquired between 2013 and 2015. Field polygons were rasterized at 30 m so that it matched Landsat’s grid, providing a wall-to-wall validation data set rich of >2 billions reference pixels.

Ancillary data were collected in order to assess if local accuracy patterns can be explained by proximity to specific landscape features and/or environmental parameters. Those include the hole-filled Shuttle Radar Topography Mission digital elevation model (90 m, 2003; Jarvis et al. (2008)), the annual precipitation and the mean temperature from the WorldClim database (30 arc seconds, 1960-1990; Hijmans et al. (2005)), the IFPRI-SPAM crop type distribution map (5 arc minutes, 2000; You et al. (2014)), the irrigated areas coming from the irrigated area map of Africa (250 m, 2010; International Water Management Institute (2016)), the settlement locations from the Global Insight 2012 data set, and the water courses as well as the road network from OpenStreetMap.

3.4 Methods

The method section is structured in four parts. In section 3.4.1, we introduce the classification scheme that was developed to derive the national-scale cropland map of South Africa for the 2013-2015 period. Section 3.4.2 presents the map accuracy assessment using the wall-to-wall validation data set. Then, we detail how we related the spatial patterns of accuracy with explanatory variables in section 3.4.3. Finally, section 3.4.4 presents how we assessed the respective importance of the spectral-temporal features in the classification process.
3.4. Methods

3.4.1 Classification scheme

The main originality of the classification scheme is its ability to deal with large territories and thus large volumes of data. It is a fully automated procedure designed to be generic. The classification scheme includes four main steps (Figure 3.2):

1. extraction of spectral-temporal features from the Landsat time series;
2. stratum-specific random forest classifications based on reliable pixels identified in the outdated land cover map;
3. fusion of the stratum-specific maps based on pixel-level class memberships;
4. speckle removal with a weighted majority filter that takes into pixel-level classification confidence.

Even though the goal is to produce a cropland vs. non-cropland map, the methodology works at the level of land cover classes to enhance between-class discrimination, e.g., between rainfed and irrigated cropland. While this method is largely based on the developments proposed in Chapter 2, several tweaks were necessary to better deal with large area mapping. In particular, Support Vector Machines were replaced by Random Forest because, they are faster, less sensitive to noise and parameters, and provide consistent class membership estimates. Besides, spectral-temporal features still target cropland’s salient characteristics but had to be adapted to deal with multi-year time series.

Extraction of spectral-temporal features

Three spectral-temporal features were derived from all exploitable pixels in the normalized time series, that is, pixels not affected by clouds, cloud shadows, adjacent clouds and quality flags. These features were defined to capture salient crop characteristics:

1. the median reflectance value over the three-year time series (medNDVI.red, medNDVI.nir, medNDVI.swir1, medNDVI.swir2);
2. the average reflectance of all pixels belonging to the first decile of stacked NDVI values (minNDVI.red, minNDVI.nir, minNDVI.swir1, minNDVI.swir2);
3. and the average reflectance of all pixels belonging to the last decile of stacked NDVI values (maxNDVI.red, maxNDVI.nir, maxNDVI.swir1, maxNDVI.swir2).

There were thus twelve input features for the classification (three temporal features of four spectral bands each). Figure 3.1 presents a false color composite (minNDVI.red, minNDVI.nir, minNDVI.swir1) of the study area.
Chapter 3. National-scale cropland mapping

Figure 3.2: Flowchart of the classification scheme to derive national-scale cropland maps based on outdated land cover information and spectral-temporal features.

Identifying reliable pixels from the historical land cover map

We performed a selection procedure to mitigate the classification errors contained in the NLC-2000 map and to account for potential land cover changes since the map production date. First, a class-specific erosion filter removed all class boundary pixels to account for a potential imperfect co-registration. Second, we identified “reliable pixels” for training based on an unsupervised clustering of the spectral-temporal features (Waldner et al., 2015a). “Reliable pixels” refers to pixels that are correctly labeled in NLC-2000 during the period of interest. The rationale to identify reliable pixels is that the cluster purity, i.e., the proportion of pixels of a certain class within a cluster, is a good indicator of the pixels’ reliability because pixels with similar spectral properties will belong to the same cluster. Therefore, mislabeled pixels are characterized by a low purity because they will likely belong to clusters dominated by pixels from another class which has similar spectral properties are similar (their actual class), and conversely. For each stratum, a random sample of 5,000 pixels belonging to a known class, e.g., class c, was drawn as well as a sample of 10,000 pixels from all remaining classes. These two sets were merged and clustered using self-organizing maps and the spectral-temporal features. The cluster purity was computed and all pixels labeled as c and belonging to clusters with a purity of at least 75% were flagged as reliable pixels. Additionally, we filtered out pixels
3.4. Methods

Figure 3.3: False color composite (minNDVI.red, minNDVI.nir, minNDVI.swir1) of the study area for the years 2013-2015. Forests are shown in red tones, light color depths represent bare soil (including annual cropland), greenish and blueish areas are grassland and shrubland, dark blue is water.

strongly deviating from the class distribution (95% of confidence interval) in at least one of the four bands of the three features. This reliable pixel selection was repeated for all classes present in the stratum. The final training set was constructed by sub-sampling the reliable pixels from all classes in the stratum in order to i) regain the initial class proportions as suggested by Zhu et al. (2016); Waldner et al. (2017), and ii) ensure a sample size of 5,000 pixels. The sub-sampling was carried out as to maximize the intra-class dissimilarity using the approach developed by Willett (1999). The class proportions were derived from the NLC-2000 map.

Stratum-specific classification to handle the spectral diversity

Because of vegetation dynamics and the variability of spectral signatures with environmental gradients and management practices (Bégué et al., 2014; Hentze et al., 2016), we stratified South Africa into nine zones according to the province delineation. Province boundaries tend to follow environmental boundaries in some cases and in others, they result in finer spatial units than the existing environmental zonations.

We calibrated stratum-specific random Forest (RF) classifiers. RF is a non-parametric, ensemble classifier based on a large set of decision trees and bootstrapping with replacements (Breiman, 2001). As each tree predicts a class, the RF output class is defined by taking the majority vote of all trees. RFs are particularly attractive because they require little guidance for parametriza-
Chapter 3. National-scale cropland mapping

RFs do not overfit and can handle high-dimensional inputs as well as multicollinearity (Hastie et al., 2009; Rodriguez-Galiano et al., 2012b). Finally, they achieve high robustness to random and systematic label noise up to 25%-30% (Pelletier et al., 2017). This relative insensitivity to noise is especially desirable in this case as undetected mislabeled pixels could occur in the training data. For more information on random forests, the reader is referred to Belgii and Drăguţ (2016). Practically, the number of trees was set to 500 and the number of random split variables to the square root of the number of input variables which conforms to the guidelines provided by Rodriguez-Galiano et al. (2012b) and Belgii and Drăguţ (2016).

We applied a buffer zone of one third of a degree to minimize boundary artefacts due to the stratum-specific training. In overlap areas between strata, there are several stratum-specific maps. To integrate them, we relied on per-pixel class memberships, i.e., the vote distribution of the trees between the input classes: \( p = \{p_1, p_2, ..., p_i, ..., p_n\} \) where \( p_i \) is the estimated membership of a given pixel to class \( i \), and \( n \) the number of classes. We fused the class memberships using a geometric mean operator (Kittler et al., 1998) and attributed the final class following the maximum likelihood principle. In areas without overlap, fusion was not necessary and the final class corresponds to the only stratum-specific classification available.

**Post-classification filtering**

Post-processing methods such as spatial filtering are often applied to classified maps to remove speckle and can be an important step in improving the map quality (Harris and Ventura, 1995; Murai and Omatu, 1997; Lu and Weng, 2004; Stefanov et al., 2001; Alganci et al., 2013b). An oft-used spatial filter is the majority filter which replaces isolated pixels by the majority class in the moving spatial window.

\[
M(s) = \text{argmax}_i \sum_j I(h_j(s) = i)
\]

(3.1)

where \( h_j \) are votes, \( I(\cdot) \) is an indicator function and \( n \) is the number of classes and \( M(s) \) is the final label for a given spatial window \( s \). However, conventional majority filtering results in inevitable information loss and classification errors at boundary are not dealt with effectively (Townsend, 1986; Wilson, 1992). The main reason is that this type of spatial filtering applies arbitrary weights to all locations and it could be enhanced by accounting for pixel-level classification confidence. Pixel-level classification confidence measures can be derived from the soft outputs of the classifiers. Rooted in information theory, the Equivalent Reference Probability (ERP; Bogaert et al. (2016)) is particularly interesting because it accounts for the full set of probabilities while remaining consistent with the maximum probability. Pixels classified with high confidence have an
3.4. Methods

ERP close to unity. A modified version of the filter is proposed in order to account for pixel-level classification confidence information into:

\[ M_w(s) = \text{argmax}_i \sum_j \omega_j I(h_j(s) = i) \]  \hspace{1cm} (3.2)

where \( \omega_1, \ldots, \omega_n \) are the weights. They are obtained by normalizing the ERP values within the moving windows so that they sum to 1. We implemented the weighted majority filter with a moving window of 3x3 pixels and reclassified the land cover map into a binary cropland/non-cropland map after filtering. For completeness, we compared the weighted majority filter with conventional majority filter and no spatial filtering at all.

3.4.2 Evaluation of the classification accuracy

The classified pixels were compared to the pixels from the wall-to-wall validation samples. Accuracy measures such as the overall accuracy (OA) (Congalton and Mead, 1983) as well as the F-scores for the cropland class (FS\(_C\)) and the non-cropland class (FS\(_{NC}\)) (Powers, 2011) were derived from the confusion matrix. The F-score is a class-specific accuracy metric mathematically defined as the harmonic mean of the users’ and producers’ accuracies of the class being evaluated. Standard errors of the overall accuracy estimates were also provided.

Map accuracy is known to vary in space (Foody, 2005; Waldner et al., 2015b; Renier et al., 2015; Lambert et al., 2016). Capitalizing on the wall-to-wall data set, the local variations of the accuracy measures were characterized by constraining geographically the reference data used for validation (Foody, 2005). Following a regular grid of points spaced 40 km apart, local accuracy measures were computed by considering all pixels falling within a 90x90 km\(^2\) spatial window. We interpolated the accuracy measures with an inverse distance weighting approach.

3.4.3 Explanatory variables of the classification accuracy

The potential of several variables to explain and predict the spatial variability of classification accuracy was evaluated. Together, these variables describe landscape and climate characteristics as well as physical, environmental and agricultural management conditions. They were selected because of their potential to describe different cropping practices, e.g., irrigation is highly likely to occur in areas close to rivers and intensive fields generally occur in accessible areas. By extension, different cropping practices have different spectral signatures with specific recognition ability for a classifier. They can be divided into three groups (Table 3.1):

1. **Site-specific characteristics** describe the physical and climate conditions. They also characterize local cropping practices such as crop diversity and
irrigation. The number of crops derived were interpolated to the whole country using a minimum distance algorithm.

2. **Density characteristics** relate to the intensiveness of agriculture (field density) and urbanization (road density) as well as to the potential for irrigation. They were computed using a kernel density approach that fitted a smoothly tapered surface to points or lines. The search radius is computed specifically to the input data set using a spatial variant of Silverman’s Rule of thumb.

3. **Proximity characteristics** including distance to roads, rivers, settlements or agriculture were also calculated. Factor maps depicting distances were calculated as the Euclidean distance to the nearest feature.

| Table 3.1: Potential explanatory variables of accuracy. These variables are of three types: site-specific characteristics, density characteristics and proximity characteristics. |
|---|---|
| **Variable** | **Description** |
| **Site-specific characteristics** | |
| Elevation | Elevation (m) |
| Slope | Slope (%) |
| Mean annual rainfall | Average annual temperature (mm) |
| Mean annual temperature | Average annual temperature (°C) |
| Crop diversity | Average number of crops (number of crops/km²) |
| Irrigation proportion | Proportion of crops under irrigation [%] |
| **Density characteristics** | |
| River density | Density of rivers (km/km²) |
| Road density | Density of roads (km/km²) |
| Field density | Density of fields (number of fields/km²) |
| **Proximity characteristics** | |
| Distance to roads | Euclidean distance to roads (km) |
| Distance to rivers | Euclidean distance to the nearest river (km) |
| Distance to settlements | Euclidean distance to the nearest settlement (km) |
| Distance to fields | Euclidean distance to nearest field centroid (km) |

We evaluated the degree of association of the explanatory variables with the overall accuracy and the F-scores with Multivariate adaptive regression splines (MARS) (Friedman, 1991). MARS is a non-parametric statistical method relying on a divide and conquer strategy that portions training data sets into separate piece-wise linear segments (splines) of differing gradients (slope). We calibrated one MARS model per accuracy measure. The importance scores of a predictor variable were calculated by refitting the model after dropping all terms involving the variable in question and tracking the corresponding reduction in goodness-of-fit. The best predictor variable degraded the model fit the most,
3.5. Results

and conversely (Steinberg et al., 1999). Three statistics of the MARS model express the variable importance: i) the generalized cross-validation statistic (GCV) (Hastie et al., 2005), ii) the residual sum of squares (RSS) and the number of times that each variable is involved in an optimal subset, iii) the number of times that each variable is involved in a subset (in the final, pruned model). We extracted the accuracy measures and the predictor variables at 4,000 locations following a systematic sampling scheme. Two thirds were used for calibration and one third was set aside for independent validation.

3.4.4 Remote sensing features of importance

RF provides measures of the feature importance in the classification process such as the Gini index (Breiman, 2001). Gini indices were thus analyzed to identify the influential spectral and temporal features for each stratum. A Friedman test (Friedman, 1940) and a post-hoc Nemenyi test (Nemenyi, 1962) were performed to determine if the ranking of the features was stratum-specific.

![Updated South African Cropland map and the corresponding pixel-level thematic confidence map. Red points illustrate the locations of the four zooms provided in Figure 3.5.](image)

Figure 3.4: Updated South African Cropland map and the corresponding pixel-level thematic confidence map. Red points illustrate the locations of the four zooms provided in Figure 3.5.

3.5 Results

3.5.1 Visual assessment

The updated national-scale South African cropland map depicts the typical patterns of the country’s cropland with an L-shaped intensive areas in the Western Cape and another intensive area centered on the Free State (Figure 3.4). Two large patches of high pixel-level confidence are visible: the first one includes most cropped areas in the Free State, the second incorporates most of the Western
and Northern Cape provinces. Overall, a good level of spatial consistency is achieved in both maps. The four red points marked the locations of the four zooms provided in Figure 3.5. Each subset illustrates a different landscape ranging from intensive large fields in the western Free State and in the Western Cape provinces (Figure 3.5a and Figure 3.5c) to fields under pivot irrigation in the Northern Cape province (Figure 3.5d) and smaller fields of the Eastern Cape (Figure 3.5b). The left-hand side image provides a synoptic view of the classification and of its accuracy since areas in red represents omission errors, and areas in blue commission errors. The right-hand side image is a false color composite of the maxNDVI feature (maxNDVI.swir1, maxNDVI.nir, maxNDVI.red). As this composite corresponds to a maximum NDVI composite, i.e., illustrating the maximum photosynthetic activity of the period of interest, reddish areas are associated with high photosynthetic activity where blueish areas correspond low photosynthetic activity (bare soil, fallow). Dark blue colors are burned areas.

The analysis of Figure 3.4 yields the following observations. Besides obvious misclassifications, omission errors (in red) seem to occur in areas that were previously cropped or under fallow during the period of interest (see the large red patches in Figure 3.5a and Figure 3.5b for instance). Commission errors (in blue) with natural vegetation occurs along the river which highlights the challenge of discriminating irrigated crops from riparian vegetation. They also tend to consistently affect pixels close to field boundaries and between fields. This might be related to the similarities between the spectral-temporal signatures of cropland and the surrounding grassland. Yet, this is insufficient to explain the inability to separate road pixels (Figure 3.5c). Commission might thus also be attributed to the net point spread function of the sensor for two reasons. First, it decreases the separability between pixels by mixing their signal (Huang et al., 2002b). Second, it was neglected when resampling the field boundaries to Landsat’s spatial resolution, thereby introducing a pessimistic bias on the accuracy estimation. Finally, some commission errors are due to pixels that are cropped on the composite but not digitized in field boundary dataset. The multi-year component of both the features and the validation data set as well as the imperfect co-registration of the imagery with the wall-to-wall validation data are sources of a pessimistic bias in the accuracy estimation.

### 3.5.2 Map accuracy assessment

The accuracy of the national-scale cropland map reached 92.0% (Table 3.2). From a class specific point of view, the F-scores reached 63.3% for the cropland class and 95.3% for the non-cropland class. The weighted majority filtering slightly improved the accuracy compared to the conventional majority filtering or to no filtering. Due to the high number of pixels used for the computation of the accuracy measures (>2 billion pixels), the standard deviation of the overall accuracy is very small and all differences between measure can be considered as highly statistically significant. Bearing in mind the limitations of the validation data set, we re-computed the accuracy after discarding all boundary pixels to evaluate the magnitude of the pessimistic bias due to the imperfect
3.5. Results

(a) Classification and the associated errors for subset 1

(b) Classification and the associated errors for subset 2

(c) Classification and the associated errors for subset 3

(d) Classification and the associated errors for subset 4

Figure 3.5: Selected zooms over four contrasted sites at the 1:200,000 scale. Left-hand side images for the four subsets provide a synoptic view of South African Cropland as well as its accuracy since areas in red represents omission errors, and areas in blue commission errors. Right-hand side images are false color composites of the maximum NDVI Landsat feature (maxNDVI.swir1, maxNDVI.nir, maxNDVI.red).
co-registration. The overall accuracy reached 95.7% and the F-score for the non-cropland class 97.7%. The most significant improvement was observed for the F-score of the cropland class (71.5%). Further, we calculated Mann-Whitney-Wilcoxon tests to assess if the population of pixel-level confidence differed for well-classified and misclassified cropland pixels. For each stratum, well-classified cropland pixels tend to have a higher ERP value than misclassified pixels at the .05 significance level.

Table 3.2: Accuracy measures for the classification without filtering and with two different spatial filters. The overall accuracy (OA) is given with the standard deviation (SD) of its estimation.

<table>
<thead>
<tr>
<th>Post-processing</th>
<th>OA +/- SD [%]</th>
<th>FS$_C$ [%]</th>
<th>FS$_{NC}$ [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted majority filter</td>
<td>92.0 ± 4.20e-9</td>
<td>64.2</td>
<td>95.5</td>
</tr>
<tr>
<td>Weighted majority filter/no edge pixels</td>
<td>95.7 ± 4.20e-9</td>
<td>71.5</td>
<td>97.7</td>
</tr>
<tr>
<td>No filter</td>
<td>91.7 ± 4.34e-9</td>
<td>63.3</td>
<td>95.3</td>
</tr>
<tr>
<td>Majority filter</td>
<td>91.9 ± 4.33e-9</td>
<td>63.9</td>
<td>95.5</td>
</tr>
</tbody>
</table>

Local measures of the thematic accuracy were derived from spatially constrained confusion matrices to illustrate the local variation of accuracy in the map (Figure 3.6). The overall accuracy and the F-score for the non-crop class follow generally the same spatial patterns. This behavior was expected given prevalence of non-cropland pixels. Cold spots of those measures coincide with areas where the crop proportion increases, i.e., where the probability of misclassification is likely. Regarding the accuracy of the cropland class, two hot spots of accuracy are visible and coincide with the two intensive grain producing areas of the country, one in the Western Cape province and the other in the maize quadrangle (recall Figure 3.1). Irrigated areas along the Orange river (from the center of the country to the Namibian border) are also mapped with high accuracy. Cold spots of accuracy occur in landscapes with low cropland proportion dominated by smallholder farming.

3.5.3 Drivers of accuracy

Table 3.3 presents the generalized cross-validation estimate of error of the MARS models, the residual sums of squares, and number of times that each predictor variable is involved in a subset of a pruned model. A very good fit was achieved for all three models, indicating a high prediction power of the variables. The Pearson’s correlation coefficients between the predicted and the observed accuracies were superior to 0.84. Besides, the root mean square errors were particularly low; 0.015, 0.02, 0.085 for the models based on the overall accuracy, the FS$_{NC}$ and the FS$_C$, respectively. The overall accuracy tends to decrease as the annual precipitation, the distance to fields, the field density and settlement density increase. For the non-cropland class, the F-score decreases as the altitude, temperature, field density and settlement density increase. Besides, it diminishes as the precipitation increase until it reaches 600 mm, then the trend is inverted. For the cropland class, the F-score increases with field density.
3.5. Results

but decreases as annual precipitation and settlement density increase. The analysis of the variables driving the accuracy converges with the previous spatial analysis and suggests that the method performed best in farming systems with a high productivity.

3.5.4 Spectral-temporal features of importance

The variable importance was quantified with the Mean Decrease Gini at the province level. First, a Friedman test was performed and lead to the rejection of the hypothesis that all features are equivalent at the .05 significance level (Friedman’s chi-squared = 40.111; p-value = 3.424e-05). We applied a post-
Table 3.3: Variable importance derived from the MARS models. Three parameters are provided: the generalized cross-validation (GCV) estimate of error, the residual sums of squares (RSS) as terms are added, and number of times that each variable is involved in a subset in the final, pruned model (nsubsets).

(a) Overall accuracy

<table>
<thead>
<tr>
<th>nsubsets</th>
<th>GCV</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual precipitation</td>
<td>23</td>
<td>100.0</td>
</tr>
<tr>
<td>Field density</td>
<td>22</td>
<td>44.4</td>
</tr>
<tr>
<td>Settlement density</td>
<td>21</td>
<td>35.2</td>
</tr>
<tr>
<td>Elevation</td>
<td>19</td>
<td>29.3</td>
</tr>
<tr>
<td>Distance to field</td>
<td>19</td>
<td>29.3</td>
</tr>
<tr>
<td>Annual temperature</td>
<td>18</td>
<td>27.5</td>
</tr>
<tr>
<td>Distance to rivers</td>
<td>17</td>
<td>22.1</td>
</tr>
<tr>
<td>Crop diversity</td>
<td>12</td>
<td>12.3</td>
</tr>
<tr>
<td>Slope</td>
<td>7</td>
<td>6.5</td>
</tr>
</tbody>
</table>

(b) F-score cropland

<table>
<thead>
<tr>
<th>nsubsets</th>
<th>GCV</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field density</td>
<td>23</td>
<td>100.0</td>
</tr>
<tr>
<td>Settlement density</td>
<td>22</td>
<td>76.5</td>
</tr>
<tr>
<td>Distance to rivers</td>
<td>21</td>
<td>58.2</td>
</tr>
<tr>
<td>Annual temperature</td>
<td>20</td>
<td>48.0</td>
</tr>
<tr>
<td>Annual precipitation</td>
<td>19</td>
<td>40.9</td>
</tr>
<tr>
<td>Crop diversity</td>
<td>19</td>
<td>40.9</td>
</tr>
<tr>
<td>Elevation</td>
<td>17</td>
<td>32.3</td>
</tr>
</tbody>
</table>

(c) F-score non-cropland

<table>
<thead>
<tr>
<th>nsubsets</th>
<th>GCV</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field density</td>
<td>26</td>
<td>100.0</td>
</tr>
<tr>
<td>Annual precipitation</td>
<td>25</td>
<td>53.4</td>
</tr>
<tr>
<td>Annual temperature</td>
<td>24</td>
<td>43.0</td>
</tr>
<tr>
<td>Elevation</td>
<td>23</td>
<td>40.2</td>
</tr>
<tr>
<td>Distance to field</td>
<td>22</td>
<td>36.4</td>
</tr>
<tr>
<td>Slope</td>
<td>22</td>
<td>36.4</td>
</tr>
<tr>
<td>Settlement density</td>
<td>20</td>
<td>31.1</td>
</tr>
<tr>
<td>Road density</td>
<td>16</td>
<td>23.7</td>
</tr>
<tr>
<td>Irrigation proportion</td>
<td>10</td>
<td>17.6</td>
</tr>
</tbody>
</table>

hoc Nemenyi test for a pairwise comparison of the average ranks of all 12 features. According to this test, the feature importance is significantly different (at $\alpha = 0.05$) if the average ranks differ by at least the critical difference $CD = 5.6941$. As proposed by Demšar (2006), the critical distance diagram summarizes these comparisons (Figure 3.7). A connecting line between features means that the null hypothesis of them being significantly different from one another could not be rejected. Despite the inter-province variability, three groups can be identified. The first is constituted of the first SWIR band of maxNDVI which seems important regardless of the province. A second group stands out with minNDVI.nir, minNDVI.red, medNDVI.red, minNDVI.swir2,
3.6. Discussion

In order to provide up-to-date national-scale cropland information in the absence of within-season ground truth data, we developed a strategy to select reliable calibration pixels from an outdated land cover map based on their spectral signatures. To ensure spatial consistency in the map, we derived seamless spectral-temporal features capturing the salient characteristics of crops from normalized Landsat time series. We also stratified the country to reduce the within-class variability of the spectral-temporal signature. The class memberships derived from the random forest were instrumental in combining stratum-specific classification as well as to calibrate the weighted majority filter. We applied this method in South Africa and the overall accuracy of the map yielded 92% with some local variations (86 to 99% at the stratum-level). Such accuracy levels are comparable to those attained by similar studies (Duro et al., 2012; Müller et al., 2015; Jepson, 2005; Brannstrom et al., 2008). Müller et al. (2015) concluded that spectral-temporal features were found instrumental reach >90% accuracy and to minimize outliers. Frequent errors are observed due to spectral confusions with similar classes such as grassland (Waldner et al., 2015a; Eggen et al., 2016) and pasture (Jepson, 2005; Brannstrom et al., 2008). Yet, those research have in common that the calibration data was collected by photo-interpretation or on the ground and that they significantly smaller extent. Thus, extracting data from existing maps appear as a competitive option for cropland mapping, especially where ground truth data is lacking and cannot be collected.
We observed marked accuracy patterns across the country. The overall accuracy and the F-scores for the non-cropland class were high across the country except in areas with high cropland proportions. F-scores for the cropland class were the highest in the two intensive agricultural areas (Western Cape province and maize quadrangle) as well as along the Orange river. Annual precipitation, field density, and settlement density were found to be important drivers of accuracy. Batista et al. (1985) already noted a correlation of larger field size with higher classification accuracy and potentially, with a broader range of significant variables including the proportion of crop in the scene, crop diversity, soil order and drainage class, percent slope, maximum yield, geographic location, weather, and crop development stage. In a Sudano-Sahelian landscape, Lambert et al. (2016) explained 41% of the variance of the classification with eight explanatory variables describing the landscape, the site location and the data availability.

Both the spatially explicit validation and the explanatory variables of accuracy pointed out to the fact that smallholder farming systems were the least accurately mapped. Smallholder farming systems have reportedly been noted as challenging to map with accuracy (Eggen et al., 2016). This is well illustrated by Delrue et al. (2013) who concluded that a classification method that yielded good results in commercial farming systems could not deal smallholder systems due to the small field size. Similarly, accuracies obtained by Inglada et al. (2015) were always higher than 80% for sites of intensive farming and stalled at around 50% for sites dominated by smallholder agriculture. While dependable information on commercial farming systems is critical to reduce uncertainty in the global commodity markets, traditional smallholder farming systems dominate the savanna range countries of sub-Saharan Africa and provide the foundation for the region’s food security (Sweeney et al., 2015). More generally, estimates suggest that in the rural areas of developing countries around half of the population is smallholder farmers with up to three hectares of cropland (Morton, 2007). For complex landscapes, methods could benefit from the addition of very high spatial resolution imagery (Vaudour et al., 2015) with good temporal information (Lebourgeois et al., 2017) or from any other satellite-derived environmental information, such as elevation data (Sesnie et al., 2008). The spatial resolution of Landsat time series limits without any doubt the accuracy with which fields can be resolved because of the mixed pixels and the resolution bias they introduce (Boschetti et al., 2004). In fact, Waldner and Defourny (2017) showed that for pixel counting area estimates from Landsat data could not reach a 10% accuracy target in most South African landscapes, expect the two intensive grain producing areas. This highlights that the achievable accuracy is strongly constrained by the resolution and the fragmentation of the cropland.

Spatial variations of accuracy were mapped thanks to the field boundary data. In general, validation data are not available in such abundance preventing the implementation of local accuracy assessment. In those cases, pixel-level certainty information received growing interest in the remote sensing community (Waldner et al., 2015a; Schultz et al., 2015; Schmedtmann and Campagnolo,
3.7 Conclusions

2015) because they can inform the users of the map of the spatial variations of the quality. There are obvious links between accuracy and per-pixel confidence (Loosvelt et al., 2012; Löw et al., 2013); for instance, we found that well-classified cropland pixels have on average a statistically confidence value. Nonetheless, the information provided by confidence measures such as the equivalent reference probability remains complementary to accuracy measures.

The feature importance analysis underlined the importance the SWIR band for crop classification as already reported (Immitzer et al., 2016; Guerschman et al., 2003; Büttner and Csillag, 1989; Yang et al., 2011; Sharma et al., 1995; Peña-Barragán et al., 2011; Lambert et al., 2016). The importance of the SWIR band ought to be related to a differential leaf water content between crops and natural vegetation (Tucker, 1980), especially in irrigated areas as well as to its specific links with canopy structure and crop residues. From a temporal perspective, three out of the top five spectral-temporal features come from the minimum NDVI which confirms that cropland is most separable when the soil is bare or prepared for sowing (Matton et al., 2015; Waldner et al., 2015a).

The availability of 10-m satellite data such as Sentinel-1 and Sentinel-2 provides perspectives of improvement to increase the accuracy of the proposed classification scheme, especially in smallholder farming systems where a higher spatial resolution is required. A higher density of image within the season would also allow to move towards annual cropland mapping, thereby reducing confusions due to land cover and land use change. The red-edge bands available with Sentinel-2 could be instrumental to enhance discrimination with grassland and wetland vegetation (Schuster et al., 2012). Besides more accurate and up-to-date land cover data could be used instead of the NLC-2000, e.g., GlobLand 30 (Chen et al., 2015a), and ancillary data geographic databases such as OpenStreetMap could also be included. Advanced filtering method of the reference land cover map such as that proposed by Radoux et al. (2014) should be tested. Other uses of per-pixel class membership or confidence information should be investigated, e.g., in a scheme to fuse the outcomes of multiple classifiers (Huang and Zhang, 2013; Hao et al., 2015; Löw et al., 2015a) or in an iterative classification process.

3.7 Conclusions

We proposed a fully automated method to map the cropland extent over large areas based on outdated land cover information and spectral-temporal features. The classification scheme was demonstrated over South Africa—a country of 1.221 million km$^2$—with multi-sensor imagery of Landsat-5, -7 and -8. To ensure spatial consistency and coherence in the map, we normalized the Landsat time series and derived spectral-temporal features to obtain seamless input data. The spatial variability of the class signatures was reduced by stratifying the country into homogeneous strata that were classified independently. The stratum-specific maps were fused based on pixel-level class membership and which limited artefacts at the stratum boundaries and a weighted majority filter
based on pixel-level classification confidence further removed the speckle. The overall accuracy of the map reached 92% with F-scores of 96% and 64% for the non-cropland and the cropland class respectively. This level of accuracy is similar or higher than what most state-of-the-art methods can achieve when ground truth data are available. In addition, imperfect co-registration and land use land cover changes during the period of interest are artificial sources of discrepancies between the validation data and the imagery, resulting in a pessimistic accuracy estimation. Smallholder farming systems were more challenging to map than the intensive producing areas. Dedicated approaches in terms of methodology and Earth Observation data, e.g., decametric or metric time series, should be investigated to lower confusions in those complex farming systems.
Chapter 4

Optimizing the spatial resolution of Sentinel-2 to reduce the burden of data volume and its implications for cropland mapping over large areas

Highlights

• The data load can be optimized by aggregating satellite imagery to a coarser resolution but no single spatial resolution equally fits all landscapes.
• A method is introduced to enable the identification of an optimized spatial resolution that the associated increase of resolution-dependent errors remains below a certain user-defined threshold.
• Spatial resolution is not the only option to reduce resolution-dependent errors as the sensor’s spatial response (PSF) is itself a source of classification errors.

Abstract. Sentinel-2 has opened a new era for the remote sensing community where decametric imagery are freely available with a 5-day revisit frequency and a systematic global coverage. Having both frequent and detailed observations across large geographic areas are ideal characteristics that can potentially revolutionize applications such as crop mapping and monitoring. However, such large volumes of high resolution data pose challenges to users in terms of problem complexity, computational resources and processing time, beckoning the increasingly relevant question: at which resolution should this imagery be processed? Here, we develop a methodology to characterize classification resolution-dependent errors in cropland mapping and explore their behavior when we move across different
Chapter 4. Optimization of the spatial resolution

spatial scales and with different levels of landscape fragmentation, taking special
care to include the effects of the instrument’s point spread function (PSF). Results
confirm that upscaling 10 m imagery such as Sentinel-2 to 30 m mitigates most of
the adverse effects generated by the PSF when comparing it to imagery natively
at 30 m, such as Landsat-8. Extending this logic, we realized two nationwide
demonstration exercises to calculate maps of the optimal spatial resolution that
ensures cropland classification errors remain below a given threshold. Based on
these maps we estimate that 31% of Belgium and 59% of South Africa could be
processed at a reduced resolution of 20 m instead of 10 m, which would reduce
data volume and processing time requirements by 23% and 44% respectively,
while keeping the classification error below 3%.

4.1 Introduction

The Sentinel-2 constellation has opened a new era for satellite Earth Observa-
tion. Systematic observations with a decametric spatial resolution, i.e., 10 m,
and a 5-day revisit frequency are now available at no cost and with a global
coverage. This unique observational configuration guaranteed by the operati-
onal commitment of the Copernicus Program of the European Union has the
potential to revolutionize many applications as the deep-rooted bottleneck of
data availability is being lifted. This may bring a paradigm shift towards more
data intensive scientific research and discovery (Hey et al., 2009), as well as
enabling the development of many new services for society as a whole.

This great opportunity to increase knowledge of the Earth System also
comes with great challenges for both scientists and information technology
experts (Nativi et al., 2015). In operational applications, a new bottleneck of a
different nature has emerged: timeliness. Data availability can take the form of
a data tsunami, posing serious challenges to store, process and deliver remote
sensing products to users in due time. This can be a problem for stakeholders
as diverse as cloud services providers, who need to optimize on-the-fly data pro-
cessing and delivery across the Internet, and local users in developing countries,
with limited computational and communication infrastructures. Yet, for many
applications the spatial resolution requirements may vary spatially depending
on the landscape structure and fragmentation (Ozdogan and Woodcock, 2006b;
Duveiller and Defourny, 2010; Löw and Duveiller, 2014). Could we not reduce
the burden of overload in this new data-rich environment by finding an opti-
mized resolution that can be spatially adjusted, and to which imagery can be
aggregated to, thereby reducing the cost and time of storage, processing and
distribution while minimizing the loss of precision?

Before addressing this question directly, it is worth recalling some key con-
cepts concerning spatial resolution and scale in remote sensing. Scale can be
defined as the number and size of the spatial sampling units used to partition
a geographic area (Lam and Quattrochi, 1992). However, a distinction must
be made between the physical sampling of the observations by the instrument
and the spacing of the grid in which the data is provided. Here, we refer to the on-ground distance between the centers of two observations as the Ground-projected Sampling Interval (GSI; see Figure 4.1 and Table 4.1 for definitions of the main scale-related terms). The spacing between the ground projection of two pixels is referred to as pixel size (and denoted $\nu$). These can differ even on a single scan line on whiskbroom instruments such as the MODerate resolution Imaging Spectroradiometer (MODIS), where increasing viewing angle combined with Earth’s curvature lead to larger GSI at the edge of the swath than at nadir, while the pixel size of the delivered image remains the same. Another common misconception is that the shape of the observation footprint is not the same as the rectangular ground projection of the pixel (Cracknell, 1998). Instead, a substantial portion of the measured radiance originates from surrounding areas (Forster and Best, 1994; Townshend, 1981). At every ground sampling interval, a detector measures the incoming radiance within its instantaneous field of view (IFOV) during a specific time interval. The IFOV is an angular measure and its ground projection is known as the GIFOV. The width of the GIFOV does not exactly match the GSI because of several factors such as the optics of the instrument, the electronics of the detector, and the image motion (Markham and Barker, 1986; Schowengerdt, 2006; Kavzoglu, 2004). Thus the image of the scene viewed by the sensor is not a completely faithful reproduction of the real ground features. Small details are blurred relative to larger features and this blurring effect can be characterized by the net sensor
Chapter 4. Optimization of the spatial resolution

Point Spread Function (PSF). Alternatively, the PSF can be expressed by the Modulation Transfer Function (MTF) which is its equivalent in the frequency domain. Several studies investigated the impact of the PSF/MTF on land cover classification (Huang et al., 2002b), sub-pixel landscape feature detection (Radoux et al., 2016), sub-pixel class proportion estimation (Huang et al., 2002b; Townshend et al., 2000).

\[ \text{IFOV} \]
\[ \text{GIFOV} \]
\[ \nu \]
\[ \text{GSI} \]
\[ \text{Observation centroid} \]

\( a \) Relationship between the data acquisition process and an arbitrary pixel grid

\( b \) A Gaussian-shaped 1-D Point Spread Function model

\( c \) The corresponding Modulation transfer Function

Figure 4.1: Schematic representation of the different factors involved in the image acquisition process: (a) the footprint of the observation is different that the footprint of grid cell it is stored in, (b) the blurring is characterized by the Point Spread Function and (c) the corresponding Modulation Transfer function.

Finding an optimal pixel size has been a recurring yet intense subject of interest in the remote sensing community (Woodcock and Strahler, 1987; Marceau et al., 1994b; Atkinson and Curran, 1995; Garrigues et al., 2006; McCloy and Bocher, 2007; Duveiller and Defourny, 2010; Löw and Duveiller, 2014). However, in past the question was not framed in a data volume reduction context, because up to now the main limitation was detailed imagery with enough revisit frequency. In some way, we have moved from a situation dominated by what Strahler et al. (1986) described as an L-resolution model to that of an H-resolution model. In the H-resolution, image pixels are smaller than the actual objects in the image (e.g. the crop specific fields for agricultural applications) whereas in the L-resolution, the pixel size exceeds the object size. This has a strong impact on image classification. When the image objects are smaller than the pixels (L-resolution), one faces the typical challenge of classifying mixed pixels. But classification can also be difficult when image objects are larger than the pixels (H-resolution), as the within-class variance is likely to be high and could thus decrease class separability and accuracy (Marceau et al., 1994b). This has progressively led to the emergence of the geographic object-based image analysis (GEOBIA) paradigm (Blaschke et al., 2014). Unlike pixel-based analysis, GEOBIA focuses on image-objects –groups of adjacent pixels– rather than
individual pixels. Consequently, the full structural parameters of the image such as color, tone, texture, pattern, shape, etc. could be jointly analyzed (Blaschke et al., 2014). GEOBIA also represents a viable alternative to cope with the modifiable areal unit problem (Openshaw, 1984) since the focus is on meaningful geographical objects rather than arbitrarily-defined pixels (Hay and Castilla, 2008). However, a major problem in GEOBIA is the additional computational cost and parametrization that are required for image segmentation. Furthermore, typical region-growing segmentation is generally not easily scalable and transposable. Therefore, a strong incentive remains in exploring how optimizing pixel size could reduce data size while retaining the squared structure of the pixels, such as with quad-tree compression algorithms, but maintaining the properties of the imagery necessary for a given application, such as land cover classification.

The objective of this study is quantify the feasibility of regionally adjusting the spatial resolution of 10 m imagery to reduce the data volume and thus optimize data processing and delivery for a given application. We have chosen to focus on cropland mapping over large areas (country scale), an application that has a strong requirements in terms of accuracy, timeliness, and frequency of product delivery. To achieve this, we develop a methodology to characterize classification resolution-dependent errors and explore their behavior when we move across different spatial scales and with different levels of landscape fragmentation, taking special care to include the effects of the PSF. The final outcomes are maps of the coarsest acceptable spatial resolution that maintains classification errors under a given threshold, along with associated gains in computing time and storage size reduction.

4.2 Data and study sites

Two nationwide sites were selected to provide a range of characteristic landscapes in contrasted agro-systems. The first country of interest, Belgium, is located in the northwest of Europe, between 51°30' and 49°30’N, and 2°33’ and 6°24'E (Figure 4.2a). In spite of its small size (30,528 km²), it covers a rather large diversity of landscapes. Agriculture occupies almost 60% of the land, forests about 20% and the rest is occupied by urban areas. The proportion of agricultural areas diminishes according to the North-South gradient whereas the grassland share increases. The second study site, South Africa, is located on the southern tip of the African continent and lies between latitudes 22° and 35°S, and longitudes 16° and 33°E spreading over 1,221,037 km² (Figure 4.2b). Only 12% of South Africa’s land is used for crop production, and only 22% of this is high-potential arable land. The main growing regions lie along the more fertile soils of the Western Cape valleys and the KwaZulu-Natal province in the West. The “maize quadrangle” in the North West Province and northwestern Free State produces 75% of the country’s maize, the most widely grown crop.

Nationwide field boundary datasets were available for both countries. Field boundary polygons describe the smallest management unit for crop production.
Chapter 4. Optimization of the spatial resolution

In general, a single crop is sown in each field except in mixed-cropping systems. These vector datasets were first rasterized to 3 m pixels, which corresponds approximately to the spatial resolution at which the fields were originally digitized. At regularly spaced locations, sample units corresponding to blocks of 5,000×5,000 pixels were extracted from the VHR field boundary map. The rationale for selecting the extent of the sample units was that they should be wider than the size of the spatial features of interest by several order of magnitude while remaining much smaller than the agroecological zone defining the agrosystem spatial patterns. Formally, each sample location is thus described by a binary mask, henceforth labeled $M$, that takes the value of 1 where it covers the target fields and 0 otherwise. Samples with less than 100 cropland pixels were discarded. The grid spacing was adjusted in a country-specific manner so that the number of sample units per country was larger than 100.
Therefore, the spatial extent of the sample units was identical for both sites but their spacing was different. In total, 380 VHR binary maps were available for South Africa, and 120 for Belgium.

4.3 Methodology

4.3.1 Characterizing the resolution-dependent error: $\Delta E$

The resolution-dependent error is here quantified based on the Pareto boundary approach of Boschetti et al. (2004). Its computation includes two steps. First, a VHR map is aggregated to a coarser resolution so that the low resolution pixels reflect the sub-pixel proportion of the class of interest. Second, the omission and commission errors that occur at different sub-pixel proportion thresholds are calculated. Once reported in the omission error/commission error bi-dimensional space, they delineate the Pareto boundary. Hypothetical Pareto boundaries for a higher and a coarser spatial resolutions are presented in red and blue on Figure 4.3. The resolution error $E$ is derived at the intersection of the 1:1 line with the Pareto boundary, i.e., omission and commission errors were given equal weights. The increase in resolution-dependent error ($\Delta E$) when moving from a fine scale to a coarser one is represented by the length of the green line on Figure 4.3 and can be written as:

$$\Delta E = E_{\text{low}} - E_{\text{high}}$$

(4.1)

where $E_{\text{high}}$ is the intersection of the Pareto boundary and the 1:1 line at the high resolution, and $E_{\text{low}}$ is the intersection of the Pareto boundary and the 1:1 line at the low resolution. It should be noted that the methodology does not require a threshold on the sub-pixel class proportion to define whether a pixel corresponds to one class or the other.

4.3.2 Three aggregation approaches

Different modeling approaches can be used to simulate how an instrument would “see” a given landscape at a given coarser spatial resolution based on the VHR binary maps. To illustrate the effects of properly taking into account the PSF, we consider three different approaches as illustrated in the graphical flowchart of figure 4.4.

The most straightforward approach consist in applying a simple non-overlapping average resampling algorithm, thereby converting the discrete information on class occurrence into a continuous scale of sub-pixel class proportions. This approach is here referred to as “naive”, since it disregards entirely the known effect of the PSF. It could alternatively be considered as “ideal”, as it reflects what is expected in an ideal world. Using the area from the VHR maps, the resolution depended error generated by such aggregation scheme can be calculated for different scales, such as the 10 m resolution of Sentinel-2 and the 30 m of
Chapter 4. Optimization of the spatial resolution

**Commission error**

**Omission error**

\[ \Delta E \]

**1:1 line**

**E\text{low}\)**

**E\text{high}\)**

Figure 4.3: Pareto boundaries for a high and a low spatial resolution in red and blue, respectively. The area below the boundary is a region of unreachable accuracy because resolution-dependent errors. The resolution error \( E \) corresponds to the intersection between the Pareto boundary and the 1:1 line. The distance between the resolution error for the low and the high resolution is the resolution-dependent error \( \Delta E \) (represented in green).

Landsat-8. Subtracting them results in a \( \Delta E \), representing the increase in error when aggregating Sentinel-2 data to 30 m in a world without PSF effect. \( \Delta E \) is calculated for each sample site across the two countries. The spatial variability of \( \Delta E \) can then be characterized by spatially interpolating the punctual values with ordinary kriging (Matheron, 1971; Oliver and Webster, 1990) using an exponential model with a nugget effect.

The second aggregation method models explicitly the spatial response that is neglected with the “naive” approach. It is called here the “convolved” approach in reference to the convolution operation used when applying the PSF model on the VHR maps. The PSF was modeled assuming a two-dimensional Gaussian function. Published on-orbit MTF measurements provided at the Nyquist frequency for Sentinel-2 and Landsat were used to calibrate the Gaussian kernels. The lowest MTFs in the visible and near-infrared domains were selected, keeping specific values for the across track and along track directions when available.

The PSF of an instrument that images the Earth can be modeled by a two-dimensional Gaussian function (Li, 2000):

\[
h(x, y) = e^{-\frac{x^2 + y^2}{2\sigma^2}}
\]

where \( x \) and \( y \) are the coordinates in the object space (with their origin at the centroid of the IFOV), \( \sigma \) is the standard deviation of the Gaussian curve. Similar approaches were implemented by Duveiller and Defourny (2010); Huang et al. (2002b); Waldner and Defourny (2017). The spatial response of an instrument
4.3. Methodology

Figure 4.4: Flowchart describing how the three resampling scenarios were modeled starting from a VHR binary mask. To highlight how sub-pixel proportions varies at each step and with each resampling scenario, a zoom on a block of 6x6 Landsat pixels is provided on the bottom right corner. On the left-hand side, the formulas to compute the error reduction ($\Delta E$) as well as the gain and loss of resampling detail which image (a-h) is needed for their computation.

with a ground sampling distance $\nu$ is given by the bi-dimensional convolution of the spatial response model over the binary map $M$ followed by a sub-sampling...
Chapter 4. Optimization of the spatial resolution

operation results in a simulated image at a given coarser pixel size $\nu$ (Duveiller and Defourny, 2010):

$$
\Pi^M_{\nu} = M(x, y) * h_{\nu}(x, y)
$$

(4.3)

The image $\Pi^M_{\nu}$ is a grid with the same extent and pixel size as $M(x, y)$. Every pixel of $\Pi^M_{\nu}$ has therefore the original size of $\nu_0$ but it displays the value that would be encoded by an instrument with a ground sampling distance equal to $\nu$ when the centroid of its instantaneous field of view falls at the $(x, y)$ coordinate. Since the output simulated image must store an information separated by a distance of $\nu$ and not $\nu_0$ (the native resolution of $M(x, y)$), $\Pi^M_{\nu}$ provides an excess of information and needs to sub-sampled. Sub-sampling is simply the operation of selecting one pixel value for every $\nu$ pixels in both the x and y directions. To calibrate the Gaussian kernel, the $\sigma$ parameters must be known. A plethora of methods have been developed to assess the PSF either pre-launch or on orbit from remote sensing imagery (see for instance Crespi and De Vendictis (2009); Radoux et al. (2016); Wenny et al. (2015) or Pagnutti et al. (2010) for a review). Instead of measuring the PSF, available MTF measurements provided at the Nyquist frequency for Landsat-7 Enhanced Thematic Mapper (ETM+) and Sentinel-2 Multi-Spectral Imager (MSI) available in the literature were used. The lowest MTFs in the visible and near-infrared domains were selected, keeping specific values for the across track and along track directions when available. Instead of measuring the PSF, available MTF measurements provided at the Nyquist frequency for Landsat-7 Enhanced Thematic Mapper (ETM+) and Sentinel-2 Multi-Spectral Imager (MSI) available in the literature were used (Storey, 2001; Gascon et al., 2016). The lowest MTFs in the visible and near-infrared domains were selected, keeping specific values for the across track and along track directions when available. From the MTF measured at the Nyquist frequency, corresponding $\sigma$ values were computed thanks to the analytic solution for the Fourier transform of a Gaussian function (Abramowitz and Stegun, 1972; Bracewell, 1989):

$$
\mathcal{F}_x \left[ e^{-\frac{x^2}{2\sigma^2}} \right] (k) = \sqrt{2\pi\sigma^2} e^{-\frac{\pi^2 k^2 \sigma^2}{2}}
$$

(4.4)

where $k$ is the spatial frequency and $\sigma$ is the variance of the Gaussian function. Through an iterative process, the corresponding $\sigma$ values were identified by minimizing the difference between the measured and modeled MTFs.

The third and last aggregation approach seeks to mitigate the effects of the PSF by upscaling the imagery in a combination of the previous two methods. First, the “convolved” approach is applied to the VHR binary masks to a 10 m image accounting for the PSF. Second, pixel class proportions were aggregated to 30 m by a simple “naive” non-overlapping average procedure. This procedure can be found in existing satellite systems such as PROBA-V, in which the central camera acquires nominally at a ground sampling distance of 100 m but delivers reflectance data at 300 m (Sterckx et al., 2014). Similarly, because the two side cameras of PROBA-V have a GSD of 300 m, the resulting imagery based on the combined use of all 3 cameras is delivered as a 1 km. In this
4.3. Methodology

“upsampling” approach, the change in resolution error $\Delta E$ can be defined by taking as reference either the 30 m convolved image or the 10 m convolved image. In the first case, the error will decrease and will be henceforth referred to as the gain of upscaling ($\Delta E_G$) while in the second it will increase. This increase will be referred to as the loss of upscaling ($\Delta E_L$).

4.3.3 Characterizing landscape fragmentation

$\Delta E$ is susceptible to vary according to the underlying spatial patterns and the fragmentation of the landscape. A large number of metrics have been developed to characterize landscape fragmentation, each highlighting a specific aspect of the landscape (see Uuemaa et al. (2013) for a review). Several studies have also shown how the magnitude of the Pareto boundary correlates with spatial fragmentation metrics (Boschetti et al., 2004; Vintrou et al., 2012a; Waldner et al., 2015b). To better understand the role of fragmentation, we use the Matheron Index ($MI$) (Matheron, 1970), which is calculated as follows:

$$MI = \frac{P_{\text{cropland}}}{\sqrt{A_{\text{cropland}}} \times \sqrt{A_{\text{total}}}}$$

(4.5)

where $A_{\text{total}}$ is the total area considered (the VHR boxes in this case), while $P_{\text{cropland}}$ and $A_{\text{cropland}}$ are the perimeter and the area of the cropland class found in $A_{\text{total}}$. For a given cropland area, the more fragmented it is the longer will be its perimeter, and hence the $MI$ will take a higher value. $MI$ is calculated over all VHR sites on both sites to explore its relationship with $\Delta E$ and see if it is transferable across landscapes. To do so, linear regressions between $MI$ and $\Delta E$ are calculated for each landscape, and their similarity is evaluated using Chow tests (Chow, 1960), which test for equality between sets of coefficients in two linear regressions. To compare two regression lines, the Chow test computes an $F$-test that combines the respective sums of squares of the two models and the residual sum of squares from a model obtained with pooled observations.

4.3.4 Mapping scale requirements while respecting minimal error targets

In a final methodological step to derive optimal resolution maps, the logic behind the calculation of $\Delta E$ is expanded to multiple scales. We took the deliberate choice to aggregated the VHR maps to the spatial resolutions of currents sensors, namely to 10 m, 30 m, 56 m, 100 m, 250 m, 375 m, corresponding to Sentinel-2, Landsat-8, AWiFS, PROBA-V, MODIS, and VIIRS (180 m was also used to have an intermediate scale between PROBA-V’s central camera and MODIS nadir resolution). In order to simulate how each sensor would actually see the scene, aggregation is done using the ‘convolved’ approach using a PSF model scaled to the pixel resolution of interest. A second order polynomial model is then used to interpolate $\Delta E$ between these specific pixel sizes at each VHR box. Given a maximum limit of resolution-dependent error that can be
tolerated, this relationship can be used to identify the coarsest pixel size at a given position (similarly to what is done in Duveiller and Defourny (2010) and Löw and Duveiller (2014)). This is considered here as the optimal spatial resolution, as it allows for maximal data reduction without compromising results due to errors above a target requirement, here set at 3%. This is illustrated for two real samples with high and low fragmentation (Figure 4.5). Again, kriging was used to interpolate the optimized resolution calculated for each VHR box to the whole country.

Figure 4.5: Application of the error reduction principles to a multi-scale analysis for a landscape exhibiting a high fragmentation and medium fragmentation. The graphs illustrate the error reduction that would be achieved by using 10 m data for resolutions ranging from 10 to 100 m. The horizontal red dashed line highlights corresponds to a user-defined acceptable increase of the resolution-dependent errors, here set at 3%. The intersection of the horizontal red dashed line with the polynomial model curve corresponds to the optimized resolution.
4.4 Results

4.4.1 The effect of aggregation on $\Delta E$

The aggregation framework developed in this work allows us to compare the three different strategies (naive, convolved, and upscaled) to aggregate the information from a 10 m spatial resolution (like that of Sentinel-2) to that of 30 m (like Landsat-8). The naive approach can be considered as representing the ideal sub-pixel proportions that are desired; it is therefore used here as a benchmark to which the other two are compared. Figure 4.6 illustrates the mean absolute deviations between all pixels values in all VHR maps in both studied landscapes for the upscaled and convolved cases with respect to the naive scenario at 30 m. In both landscapes, using the upscaled approach, i.e., a convolution at 10 m followed by an upscaling to 30 m, shows considerably lower errors than using only a convolution approach (reduction by half of the errors in Belgium and by two thirds in South Africa). This gain can be interpreted as the advantage of using a 10 m Sentinel image and upscaling it over simply using a Landsat-8 image. For completion, the comparison between naive and convolved at 10 m is also provided.

The plots in Figure 4.6 also serve to illustrate differences in the properties of the two landscapes. For instance, the fact that there is little error reduction between the 30 and the 10 m case in South Africa suggests the fields are larger and more compact, and can thus be mapped with more ease at coarser spatial resolutions. The higher spread in the distribution of the differences between upscaled and naive in Belgium are also probably due to a combination of smaller field sizes with a higher variability of field sizes among the different VHR sample sites.

4.4.2 The spatial distribution of $\Delta E$

When the VHR maps are aggregated using the convolved approach to 10 m and 30 m, we respectively obtain a simulation of how Sentinel-2 and Landsat-8 measure the landscape. The maps of $\Delta E$ using the convolved approach in Figures 4.7a and 4.7c therefore show the gain in error reduction for cropland classification when using Sentinel-2 at 10 m rather than Landsat-8 at 30 m. This gain is spatially variable for both countries. In Belgium, strong improvements are expected in the North/Northeast of the country as well as in the Southeastern part, areas where grasslands tend to dominate over cropland (Peeters, 2009), and the mean field size is smaller than the national average. In South Africa, $\Delta E$, and thus the expected gain in using Sentinel-2 over Landsat-8, is also smaller in the most agriculturally intensive areas (the Western Cape Province and the maize quadrangle covering areas in the Free State and North West primarily), but rises in landscapes less dominated by agriculture, e.g., in KwaZulu-Natal $\Delta E$ reached 10%, and in Northern Cape Province 20%.
Chapter 4. Optimization of the spatial resolution

Figure 4.6: Histograms of the mean absolute classification error between i) naive and convolved sub-pixel proportions at 10 m, ii) naive and convolved sub-pixel proportions at 30 m, and iii) naive and upscaled sub-pixel proportions at 30 m for Belgium and South Africa. On average, upscaling reduces resolution-dependent errors considerably.

In order to illustrate the effect of taking the PSF into account, Figures 4.7b and 4.7d show the spatial distribution of $\Delta E$ when limiting the aggregation to a naive approach. There is a systematic underestimation of the possible classification error reduction when moving from 30 m imagery to 10 m imagery if we only consider such naive approach.

As already shown in Figure 4.6, if imagery at 30 m is desired, upscaling 10 m imagery to 30 m provides a considerable gain in terms of reducing resolution dependent classification errors instead of using native 30 m imagery. Figure 4.8 extends this analysis in space for the two studied countries by comparing (1) the gain between upscaled, convolved 10 m imagery with convolved 30 m imagery, and (2) the loss with respect to not upscaling the 10 m imagery. In both countries, the gain was considerably larger than the loss; it ranged up to 20% while the loss was limited to 3%. While some spatial structures are visible for the gain, they do not appear as clearly for the loss. These points strengthen the added-value of having a 10 m instrument and support the idea that if a
4.4. Results

Coarser resolution can be tolerated, it can be aggregated to reduce the volume of the data and still have higher quality information that a natively coarser instrument.

4.4.3 The relationship between $\Delta E$ and landscape fragmentation

Landscape fragmentation, as expressed by the Matheron Index, appears to be a strong predictor of $\Delta E$ in both countries. $\Delta E$ increases linearly with

![Maps of spatial prediction of $\Delta E$ between 30 and 10 m. The maps on the left hand side illustrate this error reduction for the ‘convolved’ PSF-based resampling algorithm, represented the real situation when comparing imagery from two satellites such as Sentinel-2 and Landsat-8. On the right hand side, the maps shows the $\Delta E$ obtained under a “naive” aggregation procedure in which the PSF effects are neglected. The blue and orange lines represent the isolines at 5 and 10% of error reduction, respectively.]

Figure 4.7: Spatial prediction of $\Delta E$ between 30 and 10 m. The maps on the left hand side illustrate this error reduction for the ‘convolved’ PSF-based resampling algorithm, represented the real situation when comparing imagery from two satellites such as Sentinel-2 and Landsat-8. On the right hand side, the maps shows the $\Delta E$ obtained under a “naive” aggregation procedure in which the PSF effects are neglected. The blue and orange lines represent the isolines at 5 and 10% of error reduction, respectively.
Chapter 4. Optimization of the spatial resolution

Figure 4.8: Spatial prediction of the gain (a, c) and loss (b, d) due to upscaling from 10 to 30 m. The gain is on average much larger than the loss. Compared to $\Delta E$, spatial patterns are less marked.

fragmentation, confirming the reasons behind the spatial patterns seen in Figure 4.7. Figure 4.9a shows the regression lines for $\Delta E$ in both the convolved and the naive cases, all of which have an $R^2_{adj}$ above 0.93 and which are highly significant ($p$-value $< 2.2e-16$). The strength of these relationships suggests that $MI$ could be used as a proxy for error reduction without having to calculate $\Delta E$ explicitly, at least for the ranges of fragmentation explored here. However, even if the relationships of Belgium and South Africa appear quite similar, the Chow tests indicate that the regression coefficients are statistically different in all cases (F=138.12 and F=238.78 respectively for the naive and convolved case). This suggest the relationship between landscape fragmentation and the error reduction is site-dependent. The change in slope between the convolved and naive cases illustrates how the effect of taking the PSF into account is more important in more fragmented areas. It is also interesting to point out that, when the effect of the PSF is expressed in terms of proportion of $\Delta E$, the slope parameter was found statistically not different from zero ($p$-value=0.653 at the
4.4. Results

\( \alpha = 0.05 \) level).

\[
\begin{align*}
\text{Belgium} & : y = -3.36 + 230 \times x (R^2 = 0.93) \\
\text{South Africa} & : y = -1.31 + 139 \times x (R^2 = 0.95)
\end{align*}
\]

\[
\begin{align*}
\text{Belgium} & : y = -0.28 + 197 \times x (R^2 = 0.98) \\
\text{South Africa} & : y = -0.05 + 128 \times x (R^2 = 0.98)
\end{align*}
\]

\text{(a) } \Delta E \text{ versus fragmentation}

\text{(b) Loss and Gain versus fragmentation}

\textbf{Figure 4.9:} Relationships between landscape fragmentation and \( \Delta E \) under the convolved and the naive resolution as well as the gain and loss due to upsampling for Belgium and South Africa. \( \Delta E \) increases linearly with high landscape fragmentation for both scenarios and countries (\( R^2_{adj} > 0.83 \)). The gain exhibits a strong linear relationship with the Matheron Index (\( R^2_{adj} > 0.96 \)) while there seems to have no relationship with the loss (\( R^2_{adj} < 0.05 \)).
The comparison of data upscaled to 30 m with the 10 m convolved and 30 m convolved data along increasing degrees of landscape fragmentation (Figure 4.9b) further strengthens the recommendation of upsampling finer resolution data if the target is 30 m imagery. As it can be expected, the gain is proportional to the degree of fragmentation, since the consequences of the PSF effects will be stronger if there is more spatial detail. Interestingly, the loss of information remains constant for increasing fragmentation, at least for the range of MI present in these landscapes and for the change of scale considered (slope parameter not statistically different from zero according to a t-test, \( p\)-value=0.0774). Regardless of the fragmentation, the net positive impact of changing the observation scale is greater than the loss in upsampling. Here again, a Chow test on the regression coefficients strongly rejected the null hypothesis that the regression coefficients between the sites were equivalent (\( F=60.59, p\)-value<2.2e-16).

### 4.4.4 Data reduction using an optimized resolution

To exploit the possibility of reducing data volume by aggregating data, we generated maps showing the spatial resolution that keeps the resolution-dependent error below a threshold 3% compared to the 10 m imagery (Figure 4.10). Areas that are intensively cropped and characterized by large fields, e.g., the Free State and North West provinces in South Africa, can tolerate coarser spatial resolutions up to 50 m. Conversely, in fragmented areas such as the Northern Cape province in South Africa and most areas in Belgium, resolutions in the range of 10 to 30 m remain mandatory to keep the resolution error below the 3% criterion. Since the convolved aggregation approach was used, this 3% threshold is conservative. Indeed, as shown in the previous results, using an upscaled approach would result in lower \( \Delta E \) values for this optimal spatial resolution.

**Figure 4.10:** Maps of the optimized resolution for (a) Belgium and (b) South Africa. At this optimized scale, resolution-dependent errors have increased by no more than 3% compared to the acquisition resolution (10 m). The light and dark blue lines represent the isolines of 20 and 30 m, respectively.
4.5 Discussion

A simple quantification of reduction in terms of processing time and volume storage that such an spatial resolution optimization would allow is presented on Table 4.2. This analysis was based on the results presented on Figure 4.10 as well as on actual storage and processing performances of the “Sentinel-2 for Agriculture Toolbox” of the European Space Agency (Bontemps et al., 2015). Processing time estimations include atmospheric correction (Hagolle et al., 2008, 2010, 2015) and cropland mapping (Valero et al., 2016) when applied on a 2 x Intel(R) Xeon(R) CPU E5-2650 v3 @ 2.30GHz (20 cores total, 40 threads), 128 GB RAM (26 TB disk space). The simulation is provided for seven spectral bands only and for the period covering the main cropping season. All processes were assumed linearly scalable. For Belgium, 31% of the country could be processed at 20 m, which generates a reduction of 23% of the needed volume. For South Africa a reduction of 44% of data volume or processing time can be achieved by upscaling 59% of the area to 20 m. Since 9% of the area can be further upscaled to 30 m, this can extend the data volume reduction to 47%.

4.5 Discussion

The two nationwide exercises demonstrate the feasibility of characterizing classification resolution-dependent errors and generating maps of optimized spatial resolution at which data volume reduction is possible without compromising cropland mapping accuracy. Results showed that the error reduction is spatially heterogeneous and that it is strongly and linearly correlated with landscape fragmentation. We estimate that 31% of Belgium and 59% of South Africa could be processed at a reduced resolution of 20 m instead of 10 m without generating errors beyond 3%, which would reduce data volume requirements by 23% and 44% respectively. The methodology developed here is generic for any application requiring classification and should be transposable to different landscapes with ease.

Previous studies have shown how, particularly for agricultural applications, spatial resolution requirements vary in space according to the spatial patterns in the observed landscapes (Ozdogan and Woodcock, 2006b; Duveiller and Defourny, 2010; Löw and Duveiller, 2014). Here, this concept has been taken to an entirely different level by using a large amount of VHR sample maps and interpolating the results at country level, thus showing spatial gradients of changes in optimal resolution. The methodology developed here could be used at even larger scales in places where field boundaries are available, such as using the Land Parcel Identification System (LPIS) data from the European Union. Similar datasets are not likely to be available in many of the priority areas for cropland mapping (Waldrner et al., 2015c), but a systematic sample-based approach as implemented in this study could still be used to concentrate digitizing efforts and then derive large scale patterns using kriging. In the absence of strong changes in land use or agricultural policy, results of optimal resolution
Chapter 4. Optimization of the spatial resolution

for a given landscape should remain fairly stable in time.

The study has focused on Sentinel-2 because of the game-changing nature of this constellation, given its systematic acquisition of 10 m imagery with a 5-daily revisit frequency, global coverage and no cost for the end user. Far from suggesting that the spatial resolution of Sentinel-2 imagery is too high, results in this study show not only how processing time can be optimized but also how adopting an upscaling approach considerably improves classification accuracy compared to using a natively coarser spatial resolution. The main reason behind this phenomenon is the mixing of the signal described by the instrument’s PSF. While previous studies have highlighted how such effects can have considerable consequences for certain applications (Huang et al., 2002b; Duveiller et al., 2011), here we show how it still is relevant even when considering fine spatial resolutions in an H-resolution context, i.e., when image objects are larger than the pixels. Image restoration methods exist that attempt to remove PSF effects, such as deconvolution (Shen et al., 2012) or pan-sharpening (Thomas et al., 2008), but they require heavy computations and an accurate estimation of the PSF, which may not be available and even if it is, it may no longer be accurate as the PSF can even change in time (Storey, 2001). When the available spatial resolution is finer than the target image objects, upscaling is a straightforward technique that does not have such inconveniences. Up to a certain extent, upscaling could also improve classification accuracy because when the image objects, e.g., fields, are much smaller than the pixels, the within-class variance is likely to be high and could thus decrease class separability and accuracy (Marceau et al., 1994b). Upscaling Sentinel-2 imagery to an optimized resolution may thus not only reduce data size but also increase classification accuracy.

The methodology presented here could be further be improved in future studies. Results are currently based only on binary masks of cropland extent, meaning that intra-field variability is not considered explicitly. These effects could be explored using actual Sentinel-2 and Landsat-8 imagery to derive empirical $\Delta E$ values and comparing them with the theoretical ones obtained here. Comparing different algorithms may be particularly insightful as the performance of cropland mapping method is site-specific (Waldner et al., 2016). The study could be also easily be extended from mapping generic cropland to crop type mapping, in which resulting optimized resolution maps would probably vary from one crop to another depending to the complexity of their respective spatial patterns in the landscape and their intrinsic spectral separability. The definition of $\Delta E$ is currently based on a neutral assumption of compensation of omission and commission errors, but dedicated studies could envisage to alter this metric to favor the reduction of either commission or omission, depending on the specific requirements of certain applications. A composite metric could also be considered to deal with error reduction of multiples classes at the same time.
Table 4.2: Quantification of how data volume and processing time can be reduced by using an optimized resolution for cropland mapping for the different countries studied. Data volume and processing time are given for a time series of Sentinel-2 data covering the main growing season. The volume and time reduction are calculated assuming a linear reduction when passing from 10 to 20 m, and from 20 to 30 m.

<table>
<thead>
<tr>
<th>Country</th>
<th>Number of tiles</th>
<th>Storage size [Tb]</th>
<th>Time for Atmospheric corrections [h]</th>
<th>Time for classification [h]</th>
<th>Mappable area at 20 m [%]</th>
<th>Volume or time reduction [%]</th>
<th>Mappable area at 30 m [%]</th>
<th>Volume or time cum. reduction [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>6</td>
<td>0.299</td>
<td>219</td>
<td>4.5</td>
<td>31</td>
<td>23</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>South Africa</td>
<td>158</td>
<td>7.885</td>
<td>5767</td>
<td>119</td>
<td>59</td>
<td>44</td>
<td>9</td>
<td>47</td>
</tr>
</tbody>
</table>
As progress in satellite image acquisition continues, the evolution from an L-resolution to an H-resolution world should become increasingly more definitive for certain applications such as cropland mapping. If an optimal spatial resolution can clearly be established for most applications, the next generation of sensors could be designed to sample the Earth with a finer ground sampling interval than the pixel size at which the images are delivered. This would enhance the image quality without increasing the amount of data to be processed as this operation could occur on-board. This could alleviate the pressure from ground segments to deliver large data volumes on demand in near real time (Li et al., 2014a). However, resampling and upscaling on-board for the sake of timeliness alone would be severely shortsighted as the fine resolution data would be definitely lost. There are several reasons for this. First, a single optimal resolution would not exist (as shown here it varies in space, but would also vary for different applications). Second, unforeseen new applications of data can always occur at any time that may highly benefit from an increased spatial resolution. An example is how hyperspectral instruments such as GOSAT and GOME-2, designed for atmospheric measurements with very large pixels (> 40 km), were serendipitously used to retrieve sun-induced fluorescence of vegetation, a proxy for photosynthesis (Frankenberg et al., 2011; Joiner et al., 2013). However, this new signal is of high interest at much finer scale, spurrying the need for a dedicated downscaling protocol (Duveiller and Cescatti, 2016). Finally, in an H-resolution context, the natural continuation is to migrate away from pixels towards the object-oriented approach (GEOBIA) in which the contextual information can be exploited. A recurring problem that remains is optimizing region-growing algorithms when there is no prior knowledge of the expected size of the object. The methodology proposed here could be used to find the adequate scale at which to work and derive optimal parametrization to streamline a more operational segmentation of Sentinel-2 data.

4.6 Conclusion

The unprecedented availability of global 10 m time series with 5-daily repeat frequency from the Sentinel-2 constellation is lifting the long-lasting constraint of data availability that has weighed on applications such as crop mapping and monitoring. However, the sheer quantity of data poses serious challenges to deliver products to users in a timely manner. This has led us to explore whether such data load can be reduced by finding an optimized resolution to which the high spatial resolution can be aggregated to, thereby reducing the cost and time of storage, processing and distribution while maintaining the output accuracy. To do so, we established a dedicated methodology to characterize resolution-based classification errors and explore how these propagate under different aggregation scenarios. Results showed how blurring effects caused by the sensor’s PSF may reduce considerably the classification performance of a Landsat-8 30 m image with respect to a Sentinel-2 10 m upscaled image to 30 m. Two large scale demonstration exercises led to nationwide maps of
4.6. Conclusion

optimized spatial resolution that ensure cropland classification errors remain below 3% that were used to calculate to potential to reduce data load. We estimate that 31% of Belgium and 59% of South Africa could be processed at a reduced resolution of 20 m instead of 10 m, which would reduce data volume requirements by 23% and 44% respectively. The methodologies developed here are generic and thus not restrictive to agriculture applications nor to Sentinel-2. They could easily be extended to multi-class problems and remains relevant regardless of the scale.
Chapter 5

Where can pixel counting area estimates meet user-defined accuracy requirements?*

Highlights

- A framework is proposed to separate the area estimation bias into a classifier bias and resolution bias.
- The resolution bias can be anticipated before any actual classification by modeling the relationship between spatial resolution and landscape fragmentation.
- Given user-defined accuracy requirements on the area estimates, the maximum classifier bias that can tolerated is computable.
- Concrete guidelines can be derived to define the requirements in terms of spatial resolution or to identify the applicability domain area estimation by pixel counting.
- Modeling explicitly the sensor’s point spread function further reduces the applicability domain because more mixed pixels are introduced.

Abstract. Pixel counting is probably the most popular way to estimate areas from satellite-derived maps. It involves determining the number of pixels allocated to a specific thematic class and multiplying it by the pixel area. In the presence of asymmetric classification errors, the pixel counting estimator is biased. The overarching objective of this article is to define the applicability conditions of pixel counting so that the estimates are below a user-defined accuracy target. By reasoning in terms of landscape fragmentation and spatial resolution, the proposed framework decouples the overall classification bias between a resolution

bias and a classifier bias. The consequence is that prior to any classification, part of the tolerated bias is already committed due to the choice of the spatial resolution of the imagery. How much classification bias is affordable depends on the joint interaction of spatial resolution and fragmentation. The method was implemented over South Africa for cropland mapping, demonstrating its operational applicability. Particular attention was paid to modeling a realistic sensor’s spatial response by explicitly accounting for the effect of its point spread function. The diagnostic capabilities offered by this framework have multiple potential domains of application such as guiding users in their choice of imagery and providing guidelines for space agencies to elaborate the design specifications of future instruments.

5.1 Introduction

Land mapping and area estimation are among the most common applications of remote sensing. Even though they are complementary, they answer to different needs – area estimation having a more direct economic impact and more stringent accuracy requirements defined by statistical standards (Gallego, 2004). Agriculture and forestry are at the forefront of these efforts (for instance, see Soares et al. (2008); Chen et al. (2016b); Mayaux and Lambin (1995)). Gallego (2004) categorized the use of remote sensing for area estimation in three groups. In the first group, remote sensing plays the essential part in the estimation. The role of ground data is confined to the calibration of the classification algorithm, or to sub-pixel analysis. The second group comprises methods that integrate inaccurate satellite-derived information with accurate samples often collected in situ. They include for instance regression (Gonzalez-Alonso et al., 1997), calibration based on the confusion matrix (Hay, 1988; Conese and Maselli, 1992; Wall et al., 1984; Chhikara et al., 1986; Deppe, 1998; González-Alonso and Cuevas, 1993; Lewis and Brown, 2001) and small area estimators. In the third and last group, the satellite imagery supports the design of an area frame sampling, e.g., see Tsiligirides (1998). The focus of this article rests on the first group.

The most common approach to estimate areas – referred to as pixel counting – is to count the number of pixels belonging to a specific class. This approach seems to have been accepted in the early days of remote sensing (Gallego, 2004). Despite criticism on its lack of statistical justification, it is still widely adopted (see Wardlow and Egbert (2010); Shao et al. (2010); Vinciková et al. (2010); Potgieter et al. (2013); Yang et al. (2007c); Immitzer et al. (2016); Müller et al. (2015); Bartalev et al. (2016)). Nonetheless, the accuracy of these estimates is known to be related to that of the maps from which they are derived. Even with very accurate maps, errors in area estimates may occur (Moody and Woodcock, 1994; Pax-Lenney and Woodcock, 1997). In fact, the pixel counting estimator is known to be biased (Hay, 1988; Card, 1992; Czaplewski and Catts, 1992) because there is no guarantee that the omission error and the commission error will counterbalance one another. Empirical assessments have shown frequent
5.1. Introduction

Asymmetry in the error distribution. In his paper, Czaplewski (1992) presented an informal method to anticipate the magnitude misclassification bias in areal estimates. However, it is extremely challenging to foresee the accuracy of a classification method as, even when using the same input data, it depends on many factors such as the choice of the classifier (Löw et al., 2015a; Waldner et al., 2016), the location (Waldner et al., 2016), the typology, the proportion of the classes and the number of training pixels (Zhu et al., 2016). Nonetheless, the ever-increasing spectral content and revisit frequency of recent and upcoming satellites (and their combination) is expected to progressively reduce the classifier error in the future. Stehman (2005) formulated a model to compare area estimates derived from wall-to-wall mapping and statistical sampling through their respective mean square error. For both approaches, the mean square error can be partitioned into two components, bias and variance, the former being attributable to classification error. Unlike wall-to-wall mapping, statistical sampling is also affected by a sampling variance because different sampling realizations will produce different area estimates. Such a model is instrumental to identify which approach is most appropriate approach for area estimation.

The influence of spatial resolution on classification accuracy and subsequently, on area estimation is a question that has long interested the remote sensing community. A comprehensive study of the effect of spatial resolution on classification accuracy was undertaken by Markham and Townshend (1981) which concluded that classification accuracy reflects a trade-off between two factors: i) the within-class variability and ii) the boundary effect (Toll, 1985; Latty et al., 1985). On the one hand, the increased spectral variance of land cover types associated with finer spatial resolution may decrease the spectral separability of classes and result in lower classification accuracy. Empirical studies (Landgrebe et al., 1977; Latty et al., 1985; Williams et al., 1984; Toll, 1985) have observed that an increased spatial resolution does not necessarily improve the classification accuracy. A finer spatial resolution could lead to larger within-class variances, yielding higher classification errors (Cushnie, 1987; Treitz et al., 1992). This indicates that highly-separable classes (low variance) are less likely to suffer from the deterioration of classification performance due to an increased spatial resolution (Hsieh et al., 2001). On the other hand, the fact that some pixels in an image at a certain resolution are mixed—composed of multiple land cover classes—can also introduce uncertainty into the area estimates. Mixed pixels and the problems they cause have long been reported (Atkinson and Curran, 1995; Turner et al., 1989). Boschetti et al. (2004) introduced the concept of the Pareto boundary to quantify and isolate the effect of the spatial resolution (mixed pixels) on the accuracy of a map. They illustrated how, for a given resolution, mixed pixels introduce a bias acting as a conflicting objective when trying to minimize the omission or the commission error. This approach has been successfully applied in several contexts such as cropland mapping (Vintrou et al., 2012a; Waldner et al., 2016), desert locust habitat monitoring (Waldner et al., 2015b), and burned area mapping (Mallinis and Koutitas, 2012). As retrieving area estimates from coarse spatial resolution is inaccurate due to the effect of spatial aggregation on class proportions, Mayaux and Lambin

Chapter 5. Requirements for reliable pixel counting area estimates

(1995) implemented an inverse calibration model (Czaplewski and Catts, 1992) by integrating a fragmentation metric in the double sampling approach. The approach was further improved by integrating texture information (Mayaux and Lambin, 1997).

When scaling up landscape data, the magnitude of the errors in the estimation of land cover proportion depends on the spatial resolution of the map, the initial proportion of the landscape in the different land cover types, and their spatial arrangement at the initial resolution (Turner et al., 1989; Moody and Woodcock, 1994). At a given resolution, the number of pixels with mixed land cover is linked to the intrinsic characteristic of the features on the ground and it is a function of their shape, size and fragmentation (Eva and Lambin, 1998; Mayaux and Lambin, 1995; Woodcock and Strahler, 1987). By extension, some landscapes exhibit larger errors in area estimates than others when mapped at the same resolution because of their respective fragmentation (Ozdogan and Woodcock, 2006b).

The spatial resolution of an instrument is a concept that is more complex to apprehend than what it might appear at first blush. The image is never an exact reproduction of a landscape because small details are blurred. This blurring can be characterized by the net sensor point spread function (PSF) which expresses the sensor spatial responsivity of the sensor (Schowengerdt, 2006). The net PSF has three components that are related to the optics, the detectors and the motion of the sensor. The optical PSF refers to the spatial distribution of the signal in the image of a point source because an optical instrument never perfectly forms a point image of a point source. The PSF of the detectors describes the spatial blurring caused by the non-zero area of each detector while the motion PSF relates to the blurring occurring if the image moves across detectors during the time taken to integrated the signal for a pixel. Alternatively, the PSF can be expressed by its Fourier transform, the Modulation Transfer Function (MTF) (Williams and Becklund, 1989). Several studies have investigated the impact of the PSF/MTF on land cover classification (Huang et al., 2002b), sub-pixel landscape feature detection (Radoux et al., 2016), sub-pixel class proportion estimation (Huang et al., 2002b; Townshend et al., 2000; Wang and Atkinson, 2017).

Pixel counting estimators would be competitive for area estimation if one could ensure that their estimates would always remain below a certain user-defined target. This target defines the maximum classification bias in area that can be tolerated. With this as backdrop, the present paper addresses the question of the joint effect of pixel size and spatial pattern for area estimation. A framework is proposed to decouple the classification bias into two components: 1) the resolution bias (the bias due to the spatial resolution itself) and 2) the classifier bias (the bias due to classification errors of the classifier). To demonstrate is applicability, the framework is then applied over South Africa for cropland area estimation. It should be noted that the assessment method described in this paper has been developed for binary cases, e.g., a class of
interest (the foreground) versus all other land cover classes (the background). However, it can be extended to multi-class problems by successively grouping the classes in the background class.

### 5.2 Concepts and methods

The conceptual framework is based on simulating how the maximum tolerable resolution bias varies across landscapes and spatial scales. It is driven by the following rationale. The proportion of mixed boundary pixels results from the combined effect of the spatial resolution and the spatial patterns of a class in the landscape. Regardless of the spatial patterns of a class, the number of mixed pixels will decrease as the spatial resolution increases. Eventually, and for any spatial pattern, there is a spatial resolution at which the proportion of boundary pixels becomes marginal compared to that of pure pixels. At this point, the area estimate is very close to the true area value. So whether or not the omitted and committed areas offset one another becomes irrelevant as, even if extremely unbalanced, they do not affect the area estimation significantly. Conversely, as the resolution coarsens, sub-pixel proportions are no longer concentrated at the extremes of the class patches and the area estimation error can be large depending on landscape spatial structure. This rationale is illustrated for four actual landscapes on Figure 5.1.

The rationale as explained above details the case of a theoretical map not affected by classification errors. Of course, it is very unlikely for maps derived from remote sensing to be unaffected by classification errors. Thanks to the Pareto boundary validation method (Boschetti et al., 2004), classification errors can be separated into the error due to the spatial resolution itself and the error due to the classifier. By transposing this concept to area estimation bias, one would be able to characterize, at resolution $L$ and for a landscape fragmentation $F$, the two constituent terms of classification bias ($\beta^F$), i.e., classifier bias ($\epsilon^F$) and resolution bias ($\rho^F$):

$$\epsilon^F + \rho^F = \beta^F \tag{5.1}$$

For pixel counting area estimates to be competitive, the classification bias must be less than or equal to a user-defined target $\tau$ that defines the maximum tolerable bias, that is:

$$\epsilon^F + \rho^F \leq \tau \tag{5.2}$$

While the classifier bias is only measurable \textit{a posteriori}, the resolution bias could be modeled \textit{a priori} in a similar way that the Pareto boundary models the omission and commission error. The equation can be rewritten as to define the maximum tolerable classifier error as the difference between the accuracy target and the resolution bias:

$$\epsilon^F \leq \tau - \rho^F \tag{5.3}$$
It follows that before any classification the affordable bias defined by the end-users is actually lower than the user-defined target because the spatial resolution itself is already a source of bias. The magnitude of the maximum classifier bias that is affordable will vary as a function of spatial resolution and landscape fragmentation jointly. Both left-hand side terms can be known independently of classification. Therefore, the challenge becomes the definition of the relationship between the spatial resolution, the landscape fragmentation to predict the resolution bias and then derive the affordable classifier bias.

To quantify the resolution bias, we propose to use a set of very high resolution (VHR) binary masks to simulate masks at coarser pixel sizes and to derive information on the spatial patterns. Typically, pixels in the VHR binary masks take the value “1” when covering a foreground pixel and “0” elsewhere. The characterization of the landscape is presented in section 5.2.1 and the calculation of the area estimation error for different spatial resolutions in section 5.2.2. Section 5.2.3 combines the outputs of sections 5.2.2 and 5.2.1 to define the maximum tolerated classifier bias. For the sake of mathematical generality, the foreground, and the background classes are henceforth referred to as $\omega_1$ and $\omega_2$, respectively. Finally, it must be noted that the proposed methodology applies to hard classifiers only; a soft classification scheme could indeed theoretically achieve 100% accuracy even with mixed pixels.

### 5.2.1 Landscape characterization

The landscape of each sample unit was characterized by the Matheron Index ($MI$) (Eva and Lambin, 1998) which indicated the level of fragmentation. In a binary case, it can be computed as follows:

$$MI(\omega_1) = \frac{P_{\omega_1}}{\sqrt{A_{\omega_1}} \sqrt{A_{total}}}$$

(5.4)

where $P_{\omega_1}$ is the total perimeter of class $\omega_1$, $A_{\omega_1}$, the area of the foreground, and $A_{total}$, the total area. The numerator measures the number of pairs of adjacent pixels classified as foreground and background. This can be assimilated to a measure of the length of the perimeter line of foreground pixels. The denominator normalizes this count by the size of the foreground and the size of the entire area. The Matheron Index has already been successfully implemented to relate landscape patterns to cropland classification accuracy (Lambert et al., 2016), as well as to resolution bias for cropland mapping (Vintrou et al., 2012a). For the needs of the present study, the Matheron Index needs to be computed at the native resolution of the VHR masks.
Figure 5.1: Example of how mixed pixels occur at different spatial resolutions and across several landscape fragmentation. The top row presents four very high resolution binary maps depicting actual landscapes. Fragmentation of those landscapes increases from left to right. Rows below illustrate the same landscape at increasingly coarser resolutions. The resolution bias is given in the top-left corner.
5.2.2 Evaluating the area estimation bias

To study the effects of the spatial scale, the VHR binary masks must be scaled to coarser pixel sizes $L$. Starting from the original resolution, the spatial resolution is progressively degraded to coarser resolutions, assigning to each new coarser pixel a value corresponding to the proportion of class $\omega_1$ it includes. Ideal hard classification maps can be obtained by defining a threshold $t$ on the proportion of foreground within a low-resolution pixel (Boschetti et al., 2004). The classification is said to be ideal because it is only possible to reduce the commission error by raising the threshold $t$ which in turns increases the omission error, and conversely.

The Pareto boundary approach focuses on quantifying the error due to the spatial resolution in terms of omission and commission errors. For the purpose of area estimation, this approach needs to be adapted to account for a potential offset of the omitted and committed areas. Indeed, a particularity of area estimation is that errors can compensate one another. Such behavior makes the direct use of the commission and omission error inappropriate. Instead, another criterion – the absolute area estimation bias ($B_L(t)$) – was evaluated instead.

$$B_L(t) = \frac{\|O_L(t) - C_L(t)\|}{A} \quad (5.5)$$

with $A$ the actual area of class $\omega_1$; $O_L(t)$ and $C_L(t)$ are respectively the area omitted and committed by the optimal low-resolution classification obtained with threshold $t$. The double straight lines denote the absolute value. For the sake of conciseness, the absolute area estimation bias is henceforth referred to as the resolution bias. However, the reader should be aware that the resolution bias as defined in this paper holds a different meaning than that of Boschetti et al. (2004). For a given threshold $t$ and a given resolution $L$, the omitted and committed areas can be computed as follows:

$$O_L(t) = \int_{0+}^{t} iN_L \, di \quad (5.6)$$

$$C_L(t) = \int_{t}^{1} (1-i)N_L \, di \quad (5.7)$$

Furthermore, the total foreground area $A$ is given by:

$$A = \int_{0+}^{1} iN_L \, di \quad (5.8)$$

where $N_L$ is the number of cells with fraction $i$ ($0 < i < 1$) of class $\omega_1$ from the native resolution of the VHR masks. In practice, the three previous equations are discrete because sub-pixel proportion values are constrained by the precision with which they are stored, i.e., by the bit depth. In the discrete case, they become:
5.2. Concepts and methods

\[ O_L(t) = \sum_{i>0}^t iN_{Li} \quad (5.9) \]

\[ C_L(t) = \sum_{i=t}^1 (1 - i)N_{Li} \quad (5.10) \]

\[ A = \sum_{i=t}^1 iN_{Li} \quad (5.11) \]

At each resolution, the area estimation error was then characterized by a single metric: the average distance from the origin to the boundary. Thus, for each initial VHR binary mask, the resolution bias at a given resolution was computed as

\[ \rho^F_L = \frac{1}{n} \sum_{t=1}^n B_L(t) \]

where \( n \) is the number of times the resampled reference map was discretized to derive the Pareto boundary.

5.2.3 Defining the maximum tolerable classifier bias

The user-defined target defines the upper-bound bias that can be tolerated. In the previous sections, the set of selected VHR binary masks was characterized in terms of scale, spatial fragmentation, and associated bias. Information on these populations can be summarized and mapped along the resolution and fragmentation dimensions. By reporting all triplets in a 3D space, one could approximate a surface that models the relationship between these three variables. One could typically expect that the resolution bias increases with both pixel size and landscape fragmentation (Figure 5.2).

Finally, the modeled surface is sliced so that the resolution-dependent bias is less than or equal to the accuracy target. Any point below that line and above the surface meets the accuracy requirements. Also, the larger the distance between the modeled surface and the requirement plan, the larger the classification errors one can afford.

5.2.4 Modeling the sensor’s spatial response

The method described so far is valid for a theoretical instrument with a ideal (and unrealistic) square wave point spread function. This section takes it a step further by explicitly modeling the sensor’s point spread function. The PSF of an Earth imaging instrument can be modeled by a bell-shaped two-dimensional Gaussian function (Li, 2000):

\[ h(x, y) = e^{\left(-\frac{x^2 + y^2}{2\sigma^2}\right)} \quad (5.12) \]

where \( x \) and \( y \) are the coordinates in the object space (with their origin at the centroid of the instantaneous field of view), \( \sigma \) is the standard deviation of the Gaussian curve. A similar approach was adopted by Duveiller and Defourny (2010) and Huang et al. (2002b). The modulation transfer function at Nyquist frequency is generally provided to users as measure of the point spread function.
Chapter 5. Requirements for reliable pixel counting area estimates

Figure 5.2: Example of relationship between landscape fragmentation, spatial resolution, and the resolution bias. The red dashed line correspond to the accuracy target.

Given an MTF value, the corresponding $\sigma$ value can be obtained thanks to the analytic solution for the Fourier transform of a Gaussian function (Abramowitz and Stegun, 1972; Bracewell, 1989):

$$F_x \left[ e^{-\frac{x^2}{2\sigma^2}} \right](k) = \sqrt{\frac{2}{\pi\sigma^2}} e^{-2\pi^2 k^2 \sigma^2}$$  \hspace{1cm} (5.13)

where $k$ is the spatial frequency and $\sigma$ is the variance of the Gaussian function. The spatial response of an instrument with a ground sampling distance $\nu$ is given by the bi-dimensional convolution of the spatial response model over the VHR binary map $M$ followed by a sub-sampling operation results in a simulated image at a given coarser pixel size $\nu$ (Duveiller and Defourny, 2010):

$$\Pi_M^\nu = M(x, y) * h_\nu(x, y)$$  \hspace{1cm} (5.14)

The image $\Pi_M^\nu$ is a grid with the same extent and pixel size as $M(x, y)$. Every pixel of $\Pi_M^\nu$ has therefore the original pixel size $\nu_0$ but it displays the value that would be encoded by an instrument with a ground sampling distance equal to $\nu$ when the centroid of its instantaneous field of view falls at the $(x, y)$ coordinate. Since the output simulated image must store an information separated by a distance of $\nu$ and not $\nu_0$ (the native resolution of $M(x, y)$), $\Pi_M^\nu$ provides an excess of information and needs to be sub-sampled. Sub-sampling is simply the operation of selecting one pixel value for every $\nu$ pixels in both the $x$ and $y$ directions.

The MTF is specific to each spectral bands of a sensor and may vary over time. Hence, to remain generic the MTF values at the Nyquist frequency
that are used henceforth correspond to the upper and lower bounds of the MTF specifications of the MultiSpectral Imager (MSI) onboard of Sentinel-2 (MTF = 0.15-0.3 for the 10-m bands (European Space Agency, Last consulted: 2016-12-15)).

5.3 Application to cropland mapping in South Africa

One land cover class of particular interest is the cropland as area estimates of cultivated lands are essential for purposes such as estimating production or irrigation needs (Ozdogan and Woodcock, 2006b). The framework developed above was implemented in realistic conditions to evaluate the accuracy requirements for cropland assessment over the whole South African country. Attention was focused on identifying the maximum tolerable classifier error for cropland area estimation at different resolutions and spatial scales.

5.3.1 Study site

South Africa is located on the southern tip of the African continent and lies between latitudes 22° and 35°S, and longitudes 16° and 33°E spreading over 1,221,037 km². It has a dual agricultural economy, with both well-developed commercial farming and more subsistence-based production in the deep rural areas. While 12% of South Africa’s land can be used for crop production, only 22% of this is high-potential arable land. The greatest limitation is the availability of water, with uneven and unreliable rainfall. Around 1.3-million hectares are under irrigation, and around 50% of South Africa’s water is used for agriculture. The main growing regions lie along the most fertile soils of the Western Cape valleys and the KwaZulu-Natal province in the East. The “maize quadrangle” in the North West Province and northwestern Free State produces 75% of the country’s maize. Maize is most widely grown, followed by wheat, sugarcane, and sunflower.

5.3.2 Data

A nationwide field boundary dataset was at hand. It provides the area and location of each agricultural field where herbaceous crops are cultivated, regardless of their crop type. This vector dataset was first rasterized to 3 m which corresponds approximately to the spatial resolution at which the fields were digitized. Then, 15×15km² sample units were extracted at regularly spaced locations; the grid had spacing intervals of about 70 km (Figure 5.3). Sample units with a population of cropland pixels less than 100 were discarded.

5.3.3 Calibration and results

First, the relationship between scale and bias was calculated for all sample units. The spatial resolution of each block was degraded to 10 m, 30 m, 56 m, 100
m, 250 m, 375 m, 500 m, and 1000 m with an average resampling algorithm. The Matheron Index was also computed based on the VHR binary masks. Then, a 3D surface was modeled using Local Polynomial Regression Fitting (LOESS; Jacoby (2000)). The residual standard error of the LOESS model was 0.13, which is reasonable given that dependent variable varies from 0 to 4.69.

Across the landscapes observed in South Africa, the resolution bias ranged from 0.4% to 419%, with a mean of 143% (Figure 5.4). As expected, the effect of resolution was more dramatic in fragmented landscapes. For fragmented landscapes with a Matheron Index of 0.2, the error peaked from 20% to 380%. This has to be related to the more dramatic increase of the mixed pixel proportion as the pixel size coarsens, leading to significant over/underestimation of the area. The end-user target was set as follows: the classification bias must be within 10% of the true area. This value has been chosen to be realistic in an operational perspective (Duveiller and Defourny, 2010; Defourny et al., 2007; Stehman, 2005). This implies that the space was sliced so that the resolution bias was lower than or equal to the maximum tolerable bias (red dashed line on Figure 5.4). All points above the red dashed line do not meet the user target.

Figure 5.5 provides another view of the requirements in terms of resolution and fragmentation and is to be interpreted as follows. The area in light gray is the region where satellite-derived maps are not competitive because regardless of the classifier accuracy, the resolution bias is already exceeds the user tolerance. The colored area corresponds to the configurations where pixel counting could be competitive if the resolution bias remains equal to or less than the value given by the color code. For a landscape with a fragmentation of 0.3, Figure 5.5a
5.3. **Application to cropland mapping in South Africa**

Figure 5.4: Area estimation bias properties for all sample units plotted across the pixel size and landscape fragmentation dimensions.

shows that the requirement on area estimates can be met for instance with 50 m imagery and a classifier bias of about 3% or with 10 m imagery and a classifier bias of about 9%.

The explicit modeling of the sensor’s spatial response increased the number of mixed pixels in the image which subsequently reduced the affordable classifier bias and shrank the applicability domain (dark grey areas on Figures 5.5b and 5.5c). The difference between the upper bound and the lower bound of the MTF requirements (respectively, MTF$_{0.3}$ and MTF$_{0.15}$) is not striking.

Interestingly, the model can be re-applied to derive the spatial distribution of the maximum tolerable classifier bias given a user-defined accuracy target and a specific spatial resolution. For example, one could derive the maximum classifier bias that would be tolerable for a resolution bias of 10% as previously defined and a spatial resolution of 10 m or 30 m which corresponds to those of Sentinel-2 and Landsat-8 (Figure 5.6). At 10 m, the applicability domain covers over 97% of the country with an ideal sensor. It decreases to 92 and 84% when modeling the PSF (Table 5.1). Even for the widest PSF, the area of 97% of the cropland can be estimated. The same simulation was also derived from a spatial resolution corresponding to that of Landsat-8 (30 m) (see Figures 5.6b, 5.6d and 5.6f). In that case, the proportion of the country that meets the requirements drops to 58%. The difference between the two PSFs/MTFs is now remarkable: the applicability domain shrinks to 17 or 7% of the total area. Still, intensively cropped areas with large average field sizes (23-36% of the country) such as the Western Cape and areas in the Free State and North West provinces remain within the requirements.
Chapter 5. Requirements for reliable pixel counting area estimates

![Figure 5.5](image)

**Figure 5.5:** Estimated boundaries in the pixel resolution-landscape fragmentation space used to define the maximum tolerable classifier bias for crop area estimation. Light grey areas correspond to combinations of resolution and fragmentation for which the spatial resolution bias already exceeds the accuracy target. Dark grey areas meet the accuracy target with an ideal sensor but do not when the sensor spatial response is modeled.

| Table 5.1: Proportion of the country area and cropland area for which the accuracy target can be met at 10 or 30 m. |
|---------------------------------|---------------------------------|---------------------------------|
| Total area (%) | Cropland area (%) |        |
| Naive | MTF0.3 | MTF0.15 | Naive | MTF0.3 | MTF0.15 |
| 10 m | 97 | 92 | 84 | 99 | 98 | 97 |
| 30 m | 58 | 17 | 7 | 92 | 36 | 23 |
Figure 5.6: Spatialization of the maximum tolerable classifier error given an accuracy target of 10\% and a spatial resolution of 10 m and 30 m.
5.4 Discussion

The implementation of the proposed framework as presented the case study demonstrates its operational applicability. Concrete guidelines can be derived to define the requirements in terms of classifier bias for pixel counting to meet user-defined accuracy targets. Based on the simulation, 97% of the South African territory could be mapped with 10 m data, and 69% with 30 m data. Those figures dropped to 84% and 7% when the most pessimistic PSF/MTF is considered. Highly productive areas such as those in the Western Cape, the Free State and North West Provinces, tolerate a higher classifier error because of their low fragmentation. The method presented here was purposefully designed to be user-oriented. The choice of the parameters provides an attractive flexibility where the user can adjust the trade-off between spatial resolution and accuracy target. These values might need better fine-tuning with respect to end-user needs. A more informed choice could easily replace the arbitrary threshold of 10%. A trend analysis of crop area statistics such as that defined by Defourny et al. (2007) could, for example, be implemented to define the user target and hence calculate the maximum tolerable classifier bias or spatial resolution.

This framework requires a VHR map of the feature being mapped. In the study case, a nationwide field boundary dataset allowed to assess how the area of cropland at the very high resolution would change as a function of the Matheron Index. Field boundary datasets such as the Land Parcel Identification System in the European Union are increasingly becoming available thanks to open data policies. In some cases however VHR maps of the features of interest would not be readily available. One would thus need to digitize the feature of interest based on VHR imagery to obtain a sample of block units similar to the one used in the case study. The sample-based approach proposed in the framework was intentionally proposed to minimize the digitization effort. Several considerations impacting the choice of size, number and spatial distribution of the sample units must be taken into account. First, the choice of size of the sample unit depends on fragmentation and should be large enough to capture the structuring landscape patterns. Smaller extent than the one proposed can be used because fragmentation metrics tend to reach a plateau and stabilize with increasing extent—though the scaling relationship is index-specific (Wu et al., 2002; Wu, 2004). Second, the spatial distribution of the block should ensure that the whole diversity of landscapes is sampled. Third, increasing the number of sample units would lead to better surface fitting and more accurate spatial representation but implies a higher digitization effort.

The methodological framework presented is not limited to finding the maximum tolerable classifier bias. It can also serve to evaluate the adequacy of existing remote sensing systems for estimating areas in various regions of the world. In fact, the proposed framework provides at least three points of entry to the question of delivering sufficiently accurate area statistics. First, and most straightforward from an application perspective, it provides support to make an informed choice on the type of imagery to be used for a specific
application. Second, it gives the expected accuracy of area estimates given a
specific spatial resolution. Third, it allows defining the domain of applicability
of area estimates given an accuracy target as demonstrated for South Africa.
Such a diagnostic tool can be of great utility to provide specifications to space
agencies to support the design of future imaging instruments dedicated to the
monitoring of agriculture in complex agrosystems.

The framework provides a pessimistic estimate of the maximum tolerable
error. Indeed, throughout this paper, it assumed that the classifier is biased in
the same direction as the resolution bias, i.e., overestimation or underestimation,
and that the errors do not compensate each other at all. Let us consider a case
in which the maximum tolerable classifier bias is 5% because the resolution bias
already overestimates 5% of the foreground class. Under specific conditions, a
classifier that underestimates the foreground class by 15% would still fall within
the 10% target. Nonetheless, one cannot afford to bet on fortuitous compensa-
tions when designing a monitoring system that provides official area estimates.

The application presented in this paper was kept simple by design to draw
the attention to the methodology itself. A more comprehensive exercise could
be considered using more landscapes with a wider range of field sizes and spatial
patterns. The approach seems indeed mature enough for such a distributed
exercise as shown by the successful large-scale demonstration over South Africa.
In fact, the extent of the application is already considerably broader than
previous experiments on scale such as Duveiller and Defourny (2010); Löw and
Duveiller (2014); Ozdogan and Woodcock (2006b). Although the approach
focused on cropland mapping, it could accommodate any crop type or land
cover where VHR feature maps can be provided.

Several studies looked at identifying the optimal spatial resolution directly
from satellite imagery (Yang et al., 2007c; Atkinson and Curran, 1997; Woodcock
and Strahler, 1987; Pax-Lenney and Woodcock, 1997; Löw and Duveiller, 2014).
However, these approaches are tightly linked with the classifier itself. The
respective requirements might artificially evolve (even slightly) from year
to year, with a different number of calibration/validation data or images in the
time series. For instance, statistics such as the local variance and scale variance
have been previously defined by Woodcock and Strahler (1987). Nevertheless,
where such statistics vary locally over the region of interest, their use in selecting
a single spatial resolution may be compromised (Atkinson and Aplin, 2004).
The results obtained here are solely based on VHR binary masks and are thus
independent of any imagery which ensures a broad relevance of the findings for
they are not tied to its nature. This ensures a broad relevance of the results.
As a general conclusion of studies that looked at the question of the optimal
spatial resolution (see Marceau et al. (1994a); Hay et al. (1997) for instance),
it resulted that the optimal spatial resolution in the sense of minimizing the
intra-class variance depends on the class of interest and is primarily affected by
its spatial and structural parameters. As a result, the paradigm of geographic
object-image analysis emerged (Blaschke, 2010). The present study has a slightly
different perspective as it provides the maximum tolerable classifier error to be competitive with area statistics standards. Consequently, it provides a neutral assessment and the ability to meet the requirements will depend on the accuracy of the selected classification method. In practice, it would still be incumbent upon the user to conduct an accuracy assessment to confirm that bias is below the required threshold and then to estimate areas according to the best practices developed by Olofsson et al. (2013, 2014). As illustrated in the case study, large areas within the applicability domain may only tolerate very limited classifier biases. Whether the maximum tolerable classifier bias is achievable in practice or not is beyond the scope of the framework. Lu and Weng (2007) reviewed image classification methods and techniques for improving classification performance. Wider application of new advanced classification methods such as incremental import vector machines (Roscher et al., 2012), oblique random forests (Breiman, 2001), color spaces (Tokarczyk et al., 2015), rotation forest (Chen et al., 2016a), SVM ensemble approaches (Huang and Zhang, 2013), copula (Voisin et al., 2014), and kernel-based extreme machine learning, (Sonobe et al., 2016) is also expected to increase the classifier accuracy.

5.5 Conclusion

This paper proposes an end-to-end approach to defining requirements on the maximum tolerable classifier bias for pixel counting area estimates to meet user-defined accuracy targets. By reasoning in terms of landscape fragmentation and spatial resolution, the applicability domain of area estimation by pixel counting can be anticipated. The rationale driving the approach is that the bias affordable by the classifier is actually much less than the target requirements because the spatial resolution already commits part of the bias. The magnitude of this resolution bias is dictated by the joint effect of landscape fragmentation and the spatial resolution of the imagery. The case study described within this study underlined that the framework is consistently applicable over a large territory such as South Africa. This framework can be used both to guide users in choosing the appropriate imagery to enhance the quality of their area estimates and as a diagnostic tool to evaluate the adequacy of existing satellite-based area estimation systems in different agrosystems. While explored in the context of estimation of cropland areas, the framework presented remains applicable to generic area estimation questions.
Conclusions and perspectives

Synthesis

Detailed, timely and dependable maps the world’s cropland are one of the first building blocks towards improved global agriculture monitoring. In spite of recent developments in data mining, big data processing and Earth Observation systems, there is a disconnect between the state-of-the-art in crop mapping and monitoring methods and what is currently implemented in operational systems. For cropland mapping, it translates into high uncertainty of the current maps. This thesis assumed that the reasons for the absence of reliable operational cropland mapping system are tied to historical constraints and mainly the limited availability of ground truth data and remotely sensed imagery. Due to recent successful launches –such as the Sentinels– and the adoption of open data policies, the scarcity of remote sensing data is progressively being lifted. Instead, the unprecedented volumes that are being acquired pose new challenges to ensure timeliness of the product delivery. In this context, an optimization strategy was proposed to bridge the historical gap between cutting-edge scientific methods and operational monitoring.

The overall objective of the thesis was thus to develop methods for large-scale cropland mapping with satellite remote sensing towards regular updates of the global cropland map and reliable area estimates. Table 7 summarizes the work completed to reach this objective. It highlights the specific achievements and findings, and recalls the characteristics of the data used in each chapter. The overarching question was further divided into two main research questions which structure this synthesis. Along with the main lessons learned, perspectives for further improving the methods or the data exploitation are discussed.
Table 7: Thesis achievements and main findings by chapter along with details on the data used.

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Achievements</th>
<th>Main findings</th>
<th>RS data</th>
<th>Ancillary data</th>
<th>Spatial resolution</th>
<th>Extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Priority analysis, Unified Cropland Layer</td>
<td>Where to focus cropland mapping efforts, Legend definition is a large source of disagreement</td>
<td>/</td>
<td>Existing land cover and cropland maps, Validation datasets, Field size map</td>
<td>250 m</td>
<td>Global 60 km x 60 km sites in Argentina, Belgium, China, and Ukraine</td>
</tr>
<tr>
<td>2</td>
<td>Cropland mapping in different landscapes, Calibration with existing maps, spectral-temporal knowledge-based features</td>
<td>Accuracy of &gt;80%, Stability of the features over time, Impact of the reference map on the classification uncertainty</td>
<td>5-day SPOT-4 data (20 m), 16-day Landsat-8 (30 m)</td>
<td>Existing land cover maps, Validation datasets</td>
<td>20 m</td>
<td>South Africa</td>
</tr>
<tr>
<td>3</td>
<td>National-scale cropland mapping, Calibration with existing map, 3 multi-year features, Error analysis</td>
<td>Accuracy of &gt;92%, Intensive cropping systems better characterized, Weighted majority filtering</td>
<td>&lt;16-day Landsat-5, -7 and -8 (30 m)</td>
<td>Field boundary dataset</td>
<td>30 m</td>
<td>South Africa and Belgium</td>
</tr>
<tr>
<td>4</td>
<td>Modeling three sensors’ spatial responses, Quantification of the reduction of resolution-dependent errors</td>
<td>No single optimized resolution, Relationship between fragmentation and resolution error, Improvement through upscaling</td>
<td>/</td>
<td>Field boundary dataset</td>
<td>/</td>
<td>South Africa</td>
</tr>
<tr>
<td>5</td>
<td>Modeling sensors’ spatial responses, Quantification of the resolution bias</td>
<td>Prediction of the maximum tolerable classifier bias, Applicability domain of area estimation</td>
<td>/</td>
<td>Field boundary dataset</td>
<td>/</td>
<td>South Africa</td>
</tr>
</tbody>
</table>
How to take advantage of diverse ancillary data sources of variable quality to support the continuous update of cropland map?

Recognizing the value of existing cropland and land cover maps, the first research question seeks to capitalize on this wealth of information—yet sometimes strongly affected by uncertainty—rather than to ignore it. Three answers to this question were proposed.

By identifying where to focus cropland mapping efforts in priority.

In chapter 1, existing cropland maps were evaluated in terms of timeliness, resolution adequacy, thematic definition and confidence in order to rank the regions according to their need for update of their cropland map. Following a comprehensive identification and collection of national to global land cover maps, the proposed multi-criteria analysis is a very simple and yet straightforward approach to identifying key areas to be updated. As a result, priority regions such as Western Africa, Ethiopia and Madagascar and South East Asia were highlighted for the remote sensing community to focus its efforts on. In the future, new criteria could be easily added or data sources could be replaced by newer and more suitable ones, e.g., improved global field boundaries or validation data.

By combining the best cropland maps.

Still in chapter 1, a unified cropland layer at 250 m for the year 2014 was produced by combining the fittest national cropland products. The accuracy of this layer varies between 82 to 94% depending on the reference used for its validation. It was further improved by masking cropland areas with a global forest map which reduced the commission errors from 46% to 26%. Such layer could be updated continuously as new cropland maps are made available.

By extracting training data to calibrate new classifiers.

The third utility that was investigated comprises methods seeking to extract reliable pixels from land cover maps for generic classifier calibration. The lack of calibration data required for the supervised classification of satellite image time series is often a constraint that hinders the generation of up-to-date land cover maps. Given the dynamic nature of cropland, the cost of in situ data collection and the timeliness requirements of annual cropland mapping, methods calibrated with in situ data often appear as impractical operational solutions. Unlike in situ data, land cover maps are available globally at medium (250-300 m) or high resolution (30 m). Baseline land cover maps are a valuable generic source of calibration data for large-scale routine cropland classification as shown in chapters 2 and 3.
Chapter 2 demonstrated how this alternative can yield satisfactory results in agrosystems as contrasted as those observed in Argentina, Belgium, China and Ukraine. Having shown that this approach provides suitable results at the local scale, chapter 3 took it to the next level and adapted it for large-scale mapping. As a demonstration, it presented an update of the South African cropland map based on a 16-year old land cover map and multi-year spectral-temporal features derived from Landsat time series. To achieve the spatial consistency required in large-area mapping, several adaptations were needed. First, the temporal sampling was lower than in chapter 2 and required to extend the time series to multiple years. The spectral-temporal features were thus modified to fit the multi-year context, keeping nonetheless the same philosophy, i.e., spectral-temporal features targeting the cropland’s salient characteristics. Second, a stratification was applied to diminish the diversity of spectral signatures to be handled by a single classifier. The differential feature importances in the stratum-specific classifiers illustrate well this diversity. Third, a post-classification filtering step was implemented to remove speckle. Capitalizing on a wall-to-wall field boundary validation data set, the overall accuracy reached 92% but a stratum-level accuracy assessment as well as a validation using spatially constrained confusion matrices revealed large variation across landscapes.

Overall, the proposed approach has potential for routine large-scale and near-real time applications. Yet, it is no silver bullet for addressing the complexity of global cropland mapping. In chapter 2, it was shown that the confidence level of the baseline map impacted more strongly the pixel-level uncertainty than the accuracy. One might speculate that decline in uncertainty is a precursor of decline in accuracy. If the quality of the baseline is poor or if land cover change is widespread, it is highly unlikely that the cleaning algorithm will manage to select a representative and reliable calibration set. A finer analysis of the how the algorithm responds to a controlled increase of error could further highlight the limitations of such an approach (see Matton et al. (2015); Pelletier et al. (2017)). Also, the difference in resolution between the baseline and the spatial resolution used for the update might matter, especially when the image objects (fields) are relatively small. In chapter 3, evidence suggests that the method succeeded in mapping intensively cropped areas whereas smallholder farming systems were more challenging to map accurately. Traditional smallholder farming systems dominate the savanna range countries of sub-Saharan Africa and provide the foundation for food security. Future efforts should address smallholder farming system specifically. The accuracies obtained by extracting calibration data from existing maps are somewhat comparable to those obtained with in situ data reported in the literature (Inglada et al., 2015; Xiong et al., 2017). However, findings in chapter 5 suggest they might still be too low to allow area estimation by pixel counting. Area estimate corrections such as those proposed by Card (1992) or Conese and Maselli (1992) remain thus fully valid to extract reliable area estimates.
How to optimize the remote sensing input to map cropland over large areas?

The second research question contributes to defining how remote sensing data should feed cropland classification methods. Two leads were investigated: on the one hand, the definition of spectral-temporal features based on the knowledge of the expected cropland signature and on the other hand, the identification of an optimized spatial resolution for specific cropland mapping applications.

By using spectral-temporal features.

In chapter 2, the knowledge-based features that were used showed a relative stability when extracted in a rolling window as well as between years. These features have a straightforward interpretation that are consistent throughout the globe even if subject to local variations. To further increase their stability, they were extracted at interpolated dates rather than at acquisition dates –as dates of cloud-free acquisition changes from one year to another. For large scale mapping, such features also offer a simple and comprehensive framework to integrate images from different orbits without losing temporal details. Indeed, as the method operates the compositing of the features at the pixel level, it tolerates time series of different lengths which would increase temporal resolution and consequently the feature extraction. There are two main drawbacks of using annual spectral-temporal features. First, features are extremely sensitive to cloud/shadow contamination. Failure to correctly mask them or smooth them would inevitably result in noisy features. This could, in turn, lead to detrimental effects on the classification accuracy. Second, they are only meaningful when sufficient data are acquired over the course of the season. An alternative approach to increase the number of the data is to include multiple years as done in Chapter 3. This has strong implication on the legend definition as the annual variations of the cropland extent are blurred. Nonetheless, the data availability is only going to increase and the impacts of this issue are going to be strongly diminished in a near future.

By optimizing the spatial resolution.

Besides optimizing the spectral-temporal classifier input, substantial effort was devoted to the increasingly relevant question: at which resolution should the imagery be processed? The spatial resolution is indeed critical it commits part of the classification error as shown by Boschetti et al. (2004). A strong focus was put on measuring and predicting resolution-dependent errors and their related biases. This approach is particularly interesting because classifier-dependent errors are extremely challenging to anticipate. For instance, Waldner et al. (2016) established different cropland classification methods applied on the same input time series lead to significantly different outputs even when calibrated and validated with the same data, and calibration data.
Conclusions and perspectives

The question of an optimized spatial resolution was first investigated for cropland mapping in the general sense and second for area estimation. After clarifying the complexities involved in the notion of spatial resolution, chapter 4 established a framework to quantify \textit{a priori} the magnitude of the reduction of resolution-dependent errors between different spatial resolution. To that aim, three spatial responses were modeled: naive, convolved and upscaled (aggregation). Results showed that modeling the PSF led to sub-pixel proportion changes of about 6% that are in turn responsible for 20-35% of the total error reduction. A strong relationship ($R^2 > 0.93$) between landscape fragmentation and the error reduction was observed: the larger the fragmentation, the larger the error reduction. Upscaling appeared as an attractive solution to mitigate the blurring effect of the point spread function. Depending on the landscape fragmentation, it can reduce the error by 10 to 35% compared to a direct acquisition at a lower resolution. The marked spatial variability of the error reduction suggests that not only there is no one-size-fits-all spatial resolution for a given task but that the mapping strategy should be adapted accordingly with, for instance, pixel-based approaches in fragmented areas and object-based methods elsewhere to reduce the within-class variance and benefit from additional descriptors such as texture and context.

By identifying the applicability domain of area estimation by pixel counting.

The notion of resolution-dependent errors was further investigated for area estimation by pixel counting. Typically, pixel counting estimators tend to be biased as there is no guarantee that the omission and the commission errors will counterbalance one another. In chapter 5, the problem was reasoned in terms of landscape fragmentation and spatial resolution, a framework was proposed to decouple the overall bias between a resolution bias and a classifier bias. Given accuracy requirements of the estimators, the maximum tolerable classifier bias can be defined for a range of spatial resolution and landscape fragmentation. The method was implemented over South Africa, demonstrating its operational applicability. The diagnostic capabilities of this framework have multiple potential domains of application such as (i) guiding users in their choice of imagery, (ii) evaluating the adequacy of existing monitoring systems area estimations throughout the world, and (iii) providing guidelines for space agencies to design future instruments specifications.

Perspectives and implications for operational cropland mapping

Potential improvements and evolutions

There are several possible sources of improvement in the proposed classification approach. First, the random selection of pixels could be optimized to account for
Potential improvements and evolutions

the inherent characteristics of the classifier as proposed by Mathur and Foody (2008). Second, calibration and fusion of different base classifiers could increase the accuracy as the classification output is tightly related to the classifier inherent characteristics. For instance, Löw et al. (2015a) reported that despite good performances of the base classifiers, there was still up to 50% disagreement in the maps produced by the two single best classifiers. Fusing different base classifiers or different calibrations of the same base classifiers could be a source of improvement. Pixel-level uncertainty measures could be instrumental in such an approach, e.g., see Hao et al. (2015). Third, the use of existing land cover maps could also be expanded to constrain spatially the occurrence of classes as done in Friedl et al. (2010). More opportunities for methodological developments will emerge once cropland maps are produced on a regular basis (monthly or annually). In that regard, the Sen2Agri toolbox (Bontemps et al., 2015) is particularly promising. Chapters 4 and 5 provided insights on the appropriate resolution to select from user point of view. This might have implications on the support of the classification – for instance, pixel-based approaches in fragmented areas and object-based methods elsewhere to reduce the within-class variance and benefit from additional descriptors such as texture and context. Besides the implementation of recent promising algorithms such as incremental import vector machines (Roscher et al., 2012), oblique random forests (Breiman, 2001), color spaces (Tokarczyk et al., 2015), or rotation forest (Chen et al., 2016a), SVM ensemble approaches (Huang and Zhang, 2013), copula (Voisin et al., 2014) is expected to increase the accuracy of the cropland maps to reach the accuracy requirements for robust area estimation.

In a context of increasing availability of Earth Observation data, the number of land cover and cropland maps is expected to grow alongside. Certainly, combining the best cropland maps in a unified product or fusing multiple maps will remain a sound approach on the long term. More sophisticated fusion approaches that rely on additional data sources such as reference data certainly ought to be considered (see Gengler and Bogaert (2016)). However, if no broader consensus is reached on the legend, inconsistencies will inevitably remain. Compared to the GLC-Share and the IIASA-IFPRI cropland maps, significant spatial disagreements were found in the unified cropland layer which might be attributed to discrepancies in the definition. This advocates for a shared definition of cropland. FAO’s Land Cover Modeling Language provides a solid foundation to reach to such a consensus.

The expected proliferation of land cover/cropland maps also entails a finer assessment of their accuracy. Given the large discrepancies between current cropland maps (Lambert et al., 2016; Adhikari and de Beurs, 2016), another critical issue to be addressed is the current inability to robustly compare different cropland maps. Several shortcomings affect the currently available validation data sets such as the GLC 2000, GlobCover 2005, STEP, VIIRS and GLC/CMO 2008 datasets. First, they were designed to validate general land cover maps and therefore cropland samples are too scarce for a proper assessment of the cropland-specific evaluation. Second, assuming that the total number of samples
Conclusions and perspectives

is relevant for global-scale assessments, it does not hold for smaller scales. It is therefore critical for the agriculture community to share a common reference data base to systematically assess the accuracy of the future and existing products.

Field boundaries were instrumental to quantify the resolution-dependent error in both chapter 4 and 5. Several methods were already proposed in the literature to extract field boundaries either based on raster data analysis (Yan and Roy, 2014) or on segmentation (Su et al., 2015) as well as to update them for one season to another (Janssen, 1993; Löcherbach, 1998; Torre and Radeva, 2000; Butenuth and Heipke, 2005; Heipke and Straub, 1999; Oesterle and Hahn, 2004; Matikainen et al., 2012). Yet, to date, such datasets remain scarce and future works should foster their extraction.

Repeated cloud cover jeopardizes the ability to derived robust knowledge-based spectral-temporal features. Early and mid-agricultural growing season, which are important periods for crop type area identification and crop yield forecasting, are characterized by both frequent and pervasive cloud extent (Whitcraft et al., 2015b). Multi-sensor time series analysis is an alternative to increase the amount of data available, yet the fusion process has its own complexity Vermote et al. (2016). Due to their all-weather and day-and-night capabilities, synergies with radar missions should also be investigated, especially as Sentinel-1 images the Earth every >2 days in C band. An alternative approach for annual cropland mapping would be to formalize the question in terms of change detection. Typically, yearly changes are related to fallow practices, land conversion to cultivation or fields abandonment. One could thus achieve yearly cropland mapping by tracking those phenomena specifically within a multi-annual maximum cropland extent (or arable land map, see Bartalev et al. (2016)) or by detecting fallows on annual or monthly basis (see Wu et al. (2014b)). Multi-annual cropland mapping alone does not satisfy the needs of downstream applications such as cropland use intensity, which support the assessment of intensification potential (Estel et al., 2016). Such studies require information such as cropping frequency (the number of cropped years), the fallow land extent (Kuemmerle et al., 2013; Li et al., 2014b; Portmann et al., 2010; Siebert et al., 2010), or the cropland abandonment (Löw et al., 2016a; Kuenmerle et al., 2009; Lakes et al., 2009) which could be directly derived from annual cropland maps. They are also essential for the extraction of other variables such as cropping patterns (temporal dynamics of cropland use), crop rotations, multi-cropping (number of harvests per year), crop duration (fraction of the year in which the cropland is covered with crops), or to investigate the carbon sequestration associated with changes in the cropland extent (Schierhorn et al., 2013).

The future of cropland mapping with Earth Observation data is going to be data rich. The generation of high resolution cropland maps for large areas will require processing of large amount of multi-sensor satellite images acquired. The need to store, manage and seamlessly process large amount of data (big data issues) should thus be addressed. Moving the algorithms to the data rather
the opposite (cloud-based systems) will become commonplace. This trend has already emerged with initiatives such as Google Earth Engine (Shelestov et al., 2017; Xiong et al., 2017) or new Mission Exploitation Platforms such as the one described in Goor et al. (2016).

Implication for operational cropland mapping

Each of the five chapters is an important and necessary piece of the puzzle to achieve routine yearly cropland mapping over large areas with high resolution image time series. Putting all the pieces together, one could draft the main characteristics of a global cropland mapping system (Figure 10). This system would have three modules. The first module would allow identifying the optimized resolutions based on field boundary samples as shown in chapters 4 and 5. The second module would be dedicated the cropland mapping itself. Spectral-temporal features would be extracted from the input time series—at the upscaled resolution when possible—and a classifier would be calibrated with existing map or in situ data when available. Finally, the third module would evaluate new maps based on the multi-criteria analysis and would update dynamically the priority map and the unified cropland layer. In that system, the priority map would serve as dashboard for the community to track and document the improvements of the global cropland map.

The bigger picture

Tackling hunger is without any doubt one of the greatest challenges of our time. Global food crises worsened significantly in 2016, driven by armed conflicts (Yemen, Afghanistan, Nigeria, Syria, South Sudan) and natural disasters such as El Niño and other climatic shocks (Ethiopia, Malawi, Zimbabwe, Haiti, Mozambique). Conditions look set to deteriorate further in 2017, with an increasing risk of famine in some areas (Food Security Information Network, 2017). This thesis was introduced in the context of improving large-scale agricultural monitoring to provide timely and dependable information on food production. Improved agricultural intelligence is needed to tackle food security, e.g., by limiting the asymmetry of information to reduce speculation on commodities, thereby stabilizing prices are reducing artificial price spikes. It could also highlight areas suffering from food deficits and raise awareness of the international community. Better cropland maps form the basis of improved crop production monitoring but remain a first—but necessary—step. Food security is a multifaceted concept as illustrated by its four constituent pillars: (i) the availability of sufficient quantities of food of appropriate quality, supplied though domestic production or import; (ii) access by individuals to adequate resources for acquiring appropriate foods for a nutritious diet; (iii) utilization of food through adequate diet, clean water, sanitation, and health care to reach a state of nutritional well-being where all physiological needs are met; and (iv) stability, because to be food secure, a population a household or and individual must have access to adequate food at all times. While cropland mapping mostly contributes to the first pillar, Earth Observation systems have
Conclusions and perspectives

Figure 10: Proposed global cropland mapping system. All chapters contribute to the design of the systems, either for the selection the resolution of the input data, for the production of new cropland maps or for providing a dashboard to the user community.

an important role to play in the three other dimensions as well. For instance, it can contribute to econometric studies on market integration (Jacques et al., 2015). Nonetheless, more research and initiatives related to the other pillars are needed, especially as the access, utilization, and stability dimensions of food security are represented by only 11.9, 13.9, and 4.2% of the total publications on food security and climate change, respectively (Wheeler and Von Braun, 2013). More research and resources are also needed to increase the quality of products, coordinate definitions of requirements for future earth observation products, reinforce ground data collection for calibration and/or validation, support national capacities and ensure and improve link to decision makers. Most questions addressed in this thesis (timeliness, big earth observation data, and lack of calibration data) are tightly related to the constraints of operational crop monitoring. These constraints remain relevant across the value chain of satellite-driven agricultural products.
At a finer scale, e.g., the farm level, remote sensing by satellites, drones, or other sensors is also slowly changing the traditional image of a farmer standing in a field, squinting anxiously at the sky for signs of rain. Precision farming or satellite farming is a farming management concept based on observing and responding to intra-field variations. Today, precision agriculture is about whole farm management with the goal of optimizing returns on inputs (machinery, labor, fertilizer, chemicals, seeds, water, energy, etc.) while preserving resources, leading to both cost savings and environmental benefits. Harnessing precision farming technology is also part of the solution to feed the World while preserving the environment for future generations.
References


References


157


References


References


Field, C. and Van Aalst, M. 2014. Climate change 2014: impacts, adaptation, and vulnerability, volume I. IPCC.


Food and Agricultural Organization. 1999. World Food Summit, UN Coordination and MDG Follow-up.
References


GOFC-GOLD. 2011. STEP reference dataset.


References


References


Markham, B. and Townshend, J. 1981. Land cover classification accuracy as a function of sensor spatial resolution.


References


References


References


References


References


References


References


References


References

World Bank. 2015. Food Price Index.


List of Publications

Publications as first author


Publications as co-author


