



Delft University of Technology

Space-based SAR and optical remote sensing for productivity monitoring and mapping of sugarcane

Molijn, Ramses

DOI

[10.4233/uuid:d8fa4d5f-eb9a-47f5-a77d-355458683ea9](https://doi.org/10.4233/uuid:d8fa4d5f-eb9a-47f5-a77d-355458683ea9)

Publication date

2020

Document Version

Final published version

Citation (APA)

Molijn, R. (2020). *Space-based SAR and optical remote sensing for productivity monitoring and mapping of sugarcane*. [Dissertation (TU Delft), Delft University of Technology, University of Campinas]. Delft University Publishers. <https://doi.org/10.4233/uuid:d8fa4d5f-eb9a-47f5-a77d-355458683ea9>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

SPACE-BASED SAR AND OPTICAL REMOTE SENSING FOR PRODUCTIVITY MONITORING AND MAPPING OF SUGARCANE

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus Prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on Tuesday 4 February 2020 at 10:00 o'clock

by

Ramses Amadeus MOLIJN

Master of Science in Geomatics,
Delft University of Technology, the Netherlands
born in Puttershoek, the Netherlands

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. R.F. Hanssen,	Delft University of Technology
Prof. dr. J. Vieira Rocha,	University of Campinas, Brazil

Independent members:

Prof. dr. rer. Nat. C. Schmullius	Friedrich Schiller University Jena, Germany
Prof. dr. C.R. de Souza Filho	University of Campinas, Brazil
Prof. dr. I. Hajnsek	ETH Zürich, Switzerland
Prof. dr. ir. S.C. Steele-Dunne	Delft University of Technology
Prof. dr. P. Osseweijer	Delft University of Technology



Delft
University of
Technology



Keywords: SAR and optical remote sensing, sugarcane, growth monitoring, land cover mapping, ground reference measurements, data assimilation, saturation effects

Printed by: Ridderprint

Front & Back: Adapted design by Laura Eggens

Copyright © 2020 by Ramses A. Molijn

ISBN: 978-94-6366-252-9

An electronic version of this dissertation is available at
<http://repository.tudelft.nl/>.

The doctoral research has been carried out in the context of an agreement on joint doctoral supervision between University of Campinas (Unicamp), Brazil, and Delft University of Technology, the Netherlands.

This is a PhD thesis in the dual degree program as agreed between UNICAMP and TU Delft.

Love, science and the rest.

Ramses A. Molijn

CONTENTS

Summary	vii
Summary	vii
Samenvatting	viii
Resumo	x
Preface	xiii
1 Introduction	1
Background and motivation	1
Research framework	5
Research objective	5
Research approach	5
Research questions and outline	6
References	7
2 Ground reference data for sugarcane productivity monitoring	13
References	13
3 Space-based SAR and optical remote sensing for sugarcane productivity monitoring	35
References	36
4 Space-based SAR and optical remote sensing data for land cover monitoring	73
References	73
5 Space-based SAR remote sensing signals affected by precipitation for vegetation characterization	87
References	88
6 Conclusions and recommendations	107
Conclusions	107
Recommendations	113
References	115
Curriculum Vitæ	117
List of Publications	119

SUMMARY

SUMMARY

In this study, research on productivity and land cover monitoring is presented, with a focus on sugarcane, based on space-based remote sensing observations that were collected by Synthetic Aperture Radar (SAR) and optical sensors. The study aims to provide new insights into techniques and methodologies that allow for cost-efficient monitoring of sugarcane productivity and the wide-scale expansion of sugarcane over long time series. It is part of a wider framework of research projects, sharing the overarching goal to contribute to a biobased society, where the resources for the production of chemicals, materials and energy are based on biomass, produced in a competitive and sustainable manner. São Paulo state, Brazil, was selected as study area, primarily because it is one of the most prominent regions worldwide that hosts sugarcane, which is one of the most prominent crops for bio-energy production.

The study includes the following research.

First of all, results are presented that are based on a wide range of biophysical measurements that were collected during a year-long ground measurement campaign in several sugarcane fields. These results are accompanied by detailed quality assessments, illustrating their reliability when collecting such measurements through such campaigns. In addition, the methodology for setting up and carrying out the ground campaign is explained, which was designed to minimize biomass alterations in the field in light of the use of the measurements for validation of space-based SAR and optical remote sensing signals.

Secondly, remote sensing signals from various satellites are compared to the ground reference measurements in order to develop space-based sugarcane productivity monitoring techniques. It includes an analysis on the sensitivity of C-band and L-band SAR and optical observations to sugarcane biomass growth, to precipitation events and to SAR sensor configurations. In addition, the spatial features in satellite imagery from the various sensors are analyzed for their temporal consistencies in order to deduce time windows during which the satellite observations are most effective for productivity monitoring. It was found that especially saturation, precipitation and sensor configurations dictate the effectiveness, particularly for SAR. Furthermore, the highest spatial resolution optical imagery proved to perform best for mapping intra-field productivity differences that were measured in the field. In addition to this study, two related but smaller studies are presented. The first focuses on a specific remote sensing technique to identify patterns in a sugarcane field that occur persistently in time. The second demonstrates how plant gaps in a densely ground-measured sugarcane field affect signals from various SAR and how this effect is influenced by spatial averaging windows, precipitation events, sugarcane height and SAR sensor type.

Thirdly, the performance of a specific Bayesian land cover monitoring model that

combines SAR and optical observations is demonstrated. The model is an adaptation of the Hidden Markov Model, which allows for the temporally-consistent tracking of vegetation states regardless of gaps in satellite observations. Attention is paid to the effect of precipitation during SAR observations on the model's performance and to certain vegetation conditions that cause classification confusion between land cover types. The research finally provides detailed insights into when SAR-only observations outperform optical-only observations and vice versa, in addition to the advantages when combining them.

Finally, a technique is introduced that exploits SAR signal fluctuations caused by varying (ground and plant) surface wetness conditions in order to improve the characterization of vegetation. Three scenarios that define the selection of SAR observations were investigated for their effect on the classification performance: (i) no distinction between wetness conditions, (ii) distinction between wetness conditions at the time of the SAR acquisitions and (iii) distinction between wetness conditions between consecutive SAR acquisitions. Particularly when the wetness conditions differ under the last scenario, it was found that performances improve. When combining this information with a-priori knowledge on soil types, the accuracy of the classification further increases. For this, maps are used that are a result from applying the previously introduced Hidden Markov Model over the entire state of São Paulo.

The datasets that are used in these studies were mainly acquired by the SAR satellites Sentinel-1, Radarsat-2 and ALOS-2, and by the optical satellites Landsat-8 and Worldview-2. For the studies that are related to land cover monitoring and vegetation characterization, high performance computing was required due to the vast amount of observation data and the complexity of the applied techniques. These facilities were mainly provided by the Dutch national supercomputer of SURF and by Google Earth Engine.

SAMENVATTING

Deze studie bevat onderzoeken die zijn verricht naar het monitoren van productiviteit en landbedekking, met name in relatie tot suikerriet, op basis van aardobservatie afkomstig uit de ruimte van *Synthetic Aperture Radar* (SAR) en optische sensoren. Het doel van het onderzoek is om nieuwe inzichten in technieken en methodologieën te verkrijgen die bijdragen aan het kosten-efficiënt monitoren van de productiviteit van suikerriet en van grootschalige toename van suikerrietplantages over lange tijdreeksen. Het onderzoek is uitgevoerd in een breder verband van onderzoeken die gezamenlijk streven naar het bijdragen aan een *biobased society*, een samenleving waar de grondstoffen voor de productie van chemicaliën, materialen en energie afkomstig zijn van biomassa die gewonnen wordt op een concurrerende en duurzame manier. De Braziliaanse staat São Paulo is gekozen als onderzoeksgebied, voornamelijk omdat dit één van de meest prominente gebieden op aarde is waar suikerriet wordt verbouwd en omdat suikerriet wereldwijd één van de belangrijkste gewassen is voor de productie van bio-energie.

De studie bevat de volgende onderzoeken.

Allereerst worden de resultaten beschreven die komen uit metingen die zijn verworven gedurende een jaarlange veldcampagne in verscheidene suikerrietvelden. Deze resultaten gaan gepaard met gedetailleerde kwaliteitswaarderingen die een beeld geven van de betrouwbaarheid van de metingen wanneer deze worden verworven middels

soortgelijke veldcampagnes. Tevens wordt de methodologie voor het opzetten en uitvoeren van de veldcampagne uitgelegd. Deze is zodanig ontworpen dat veranderingen in de velden zo veel mogelijk worden beperkt, aangezien deze metingen voornamelijk dienen ter referentie voor de metingen afkomstig uit SAR- en optische satellietsensoren.

In het tweede onderzoek worden de aardobservaties vergeleken met de grondvalidatiemetingen ten behoeve van het ontwikkelen van technieken voor het monitoren van de productiviteit van suikerriet vanuit de ruimte. Het bevat een analyse van de gevoeligheid van C-band en L-band SAR- en optische observaties voor de biomassa-aangroei van suikerriet, neerslag en configuratie van de SAR-sensoren. Daarnaast worden de spatiële patronen in de satellietbeelden van de verschillende sensoren geanalyseerd op consistentie in tijd, teneinde tijdsriteria af te leiden die aangeven wanneer satellietmetingen het meest effectief kunnen worden ingezet voor het monitoren van de productie. Hieruit blijkt dat voornamelijk de verzadiging van de signalen, neerslag en de sensorconfiguratie invloed hebben op de effectiviteit, in het bijzonder voor SAR, en dat optische beelden met de hoogste spatiële resolutie het meest geschikt zijn voor het in kaart brengen van de gemeten productiviteitsverschillen in de suikerrietvelden. Naast dit onderzoek worden ook twee kleinere gerelateerde onderzoeken behandeld. Het eerste spitst zich toe op het gebruik van satellietobservatie om spatiële patronen in een suikerrietveld te ontdekken die continue in tijd aanwezig zijn. Het tweede toont aan welk effect de uitsparingen in een intensief gemeten suikerrietveld hebben op signalen uit SAR-sensoren en hoe dit effect vervolgens wordt beïnvloed door morfologische verrichtingen op de beelden, neerslag, hoogte van suikerriet en het type SAR-sensor.

Het derde hoofdonderzoek beschrijft de prestatie van een specifiek Bayesiaans model voor het monitoren van landbedekking dat SAR- en optische satellietmetingen combineert. Het gebruikte model is een aangepaste versie van het *Hidden Markov Model* die mogelijkheid biedt om het stadium waarin de vegetatie zich verkeert consistent in tijd te volgen ongeacht de dekking van satellietmetingen. Aandacht is gegeven aan de invloed van neerslag wanneer satellietmetingen worden verricht op de prestatie van het model en aan condities van vegetatie waardoor verwarring in classificatie ontstaat tussen verschillende typen landbedekking. Het onderzoek geeft ten slotte gedetailleerd inzicht in wanneer SAR-satellietmetingen beter presteren dan optische satellietmetingen en vice versa, en welke voordelen de combinatie van deze twee met zich meebrengt.

Het vierde en laatste hoofdonderzoek beschrijft een techniek die gebruik maakt van fluctuaties in SAR-metingen wat wordt veroorzaakt door vochtigheid van het (grond- en plant-)oppervlak om zodoende de classificatie van vegetatiecondities te verbeteren. Onderzoek is gedaan naar de invloed op het presteren van de classificatietechniek op basis van drie scenarios die de selectiecriteria van SAR-metingen bepalen: (i) geen onderscheid tussen vochtigheidscondities van het oppervlak, (ii) onderscheid tussen de vochtigheidscondities op het moment dat SAR-metingen zijn verricht en (iii) onderscheid tussen vochtigheidscondities tussen opeenvolgende SAR-metingen. Het onderzoek toont aan dat de beste resultaten kunnen worden verkregen wanneer verschillen in vochtigheid onder het laatstgenoemde scenario optreden. Als deze informatie wordt gecombineerd met a-priori kennis van grondtypen kunnen de resultaten verder worden verbeterd. Voor dit onderzoek is gebruik gemaakt van landbedekkingskaarten die zijn gemaakt met behulp van het *Hidden Markov Model* over heel São Paulo staat (nagenoeg zes keer

de oppervlakte van Nederland).

De datasets die zijn gebruikt voor deze onderzoeken zijn voornamelijk verworven door de SAR-satellieten Sentinel-1, Radarsat-2 en ALOS-2, en door de optische satellieten Landsat-8 en WorldView-2. Voor de onderzoeken die zijn uitgevoerd voor het monitoren van landbedekking en classificatie van vegetatie was *high performance computing* vereist vanwege de aanzienlijke omvang van observatiedata en de complexiteit van de toegepaste technieken. Hiervoor is voornamelijk gebruik gemaakt van de Nederlandse supercomputer van SURF en Google Earth Engine.

RESUMO

Este estudo apresenta uma pesquisa sobre o monitoramento de produtividade agrícola e do uso e cobertura da terra, com foco na cana de açúcar, utilizando dados de sensoriamento remoto, que foram coletados a partir de dados de Synthetic Aperture Radar (SAR) e sensores ópticos. A pesquisa tem por objetivo realizar uma discussão sobre novas técnicas e metodologias que permitem um monitoramento de baixo custo da produtividade da cana de açúcar e também de uma ampla cobertura de expansão sucroenergética por meio de séries temporais. Esta pesquisa é oriunda de grandes projetos de pesquisa, que possuem objetivo comum de contribuir para uma sociedade baseada em bioprodutos, onde os recursos para a produção de produtos químicos, materiais e energia são baseados em biomassa, produzida em sistemas competitivos e sustentáveis. O estado de São Paulo, Brasil, foi selecionado como área de estudo, principalmente por ser uma das principais regiões globais de produção de cana de açúcar, sendo uma das mais proeminentes na produção de bioenergia.

Este estudo inclui os seguintes pesquisas.

Primeiramente, os resultados são baseados em uma ampla gama de medições de dados biofísicos coletados ao longo de um ano, em sucessivos trabalhos de campo e em diferentes ambientes de produção de cana de açúcar. Esses resultados são acompanhados por avaliações detalhadas de parâmetros de qualidade, ilustrando a confiabilidade de se coletar tais medidas em campo. Além disso, é detalhada toda metodologia de execução dos trabalhos de campo, que foi planejada para minimizar as alterações de biomassa no campo para que se pudesse validar os dados de SAR e de sensoriamento remoto óptico. Em segundo lugar, os dados de sensoriamento remoto dos diferentes satélites são comparados às medições de campo para desenvolver técnicas de monitoramento espacial da produtividade da cana-de-açúcar. Isso inclui uma análise sobre a sensibilidade das bandas C e L do SAR e das bandas dos sensores ópticos, em relação ao crescimento da biomassa de cana-de-açúcar, a precipitação e configurações do sensor SAR. Além disso, as características dos diversos sensores utilizados são analisadas quanto às suas consistências temporais, a fim de deduzir os melhores intervalos temporais para que as observações dos satélites fossem mais eficazes para o monitoramento da produtividade. Descobriu-se especialmente que a saturação, precipitação e as configurações do sensor determinam a eficácia, principalmente para SAR. Além disso, as imagens sensores ópticos de maior resolução espacial provaram ter um melhor desempenho para mapear a variação da produtividade dentro dos talhões. Além deste estudo, dois estudos relacionados, porém menores, são apresentados. O primeiro analisa uma técnica específica de sensoriamento remoto para identificar padrões em áreas de cana

de açúcar que ocorrem persistentemente no tempo. Já o segundo, demonstra como as falhas de plantio em áreas de cana densa afetam os sinais de diferentes SAR e como esse efeito é influenciado por intervalos espaciais, precipitação, altura da cana-de-açúcar e tipo de sensor SAR. Em terceiro lugar, o desempenho de um modelo específico de monitoramento da cobertura terrestre baseado em redes Bayesianas que combina dados SAR e ópticos foram demonstrados. O modelo é uma adaptação do Modelo oculto de Markov, que permite um consistente rastreamento temporal dos estágios da vegetação, independentemente de lacunas nas observações dos satélites. Cabe ressaltar o efeito da precipitação durante observações de SAR no desempenho do modelo e em certas condições vegetativas que ocasionam erros na classificação entre os tipos de cobertura da terra. Por fim, o estudo promove uma detalhada discussão sobre quando as observações de SAR superam as observações ópticas e vice-versa, com resultados melhores do que quando combinadas. Por último, apresentamos uma técnica que explora as oscilações no sinal do SAR causadas pelos diferentes níveis de umidade na planta e solo, com o intuito de melhorar a caracterização da vegetação. Ao todo, três cenários que definem a seleção das observações SAR foram investigados quanto ao seu efeito no desempenho da classificação: sem distinção entre condições de umidade; distinção entre as condições de umidade no momento da aquisição dos dados SAR; e distinção entre as condições de umidade entre aquisições consecutivas de SAR. Particularmente, foi considerado crucial para melhorar os desempenhos na classificação quando as condições de umidade se diferem no último cenário. Ao combinar esta informação com um conhecimento a priori dos tipos de solo, a acurácia da classificação aumenta ainda mais. Para tanto, são utilizados mapas resultantes da aplicação do já apresentado Modelo oculto de Markov em todo o estado de São Paulo. Os conjuntos de dados utilizados nestes estudos foram adquiridos principalmente pelos satélites SAR Sentinel-1, Radarsat-2 e ALOS-2 e pelos satélites ópticos Landsat-8 e Worldview-2. Para os estudos relacionados ao monitoramento da cobertura da terra e à caracterização da vegetação, um alto desempenho computacional foi necessário devido à grande quantidade de observações e da complexidade das técnicas aplicadas. Para tanto, foi necessário utilizar as instalações fornecidas pelo supercomputador nacional holandês de SURF e pelo Google Earth Engine.

PREFACE

This PhD track has offered a fantastic journey and I have never regretted the decision to take on the challenge. Back in 2009, however, during my final stages of writing my Masters thesis in Geomatics at Delft University of Technology, the sheer effort it took me to finalize the report practically diminished the desire to ever take on a PhD position. Even the *cum laude* classification of the MSc. track, the earlier taken study positions in Australia and Texas, and the truly fascinating BSc. Aerospace engineering track I had finished, were, back then, not convincing to continue working in academics. I needed new challenges. The subsequent traineeship at Fugro and Shell brought me a fancied cool off period and gave sensational insights into the world of business. I acquired irreplaceable operational experiences in a diverse environment with colleagues I admired greatly. And even though the work brought exciting opportunities nationally and internationally, it was not meant to be. It was not meant to be because it was too early to become a generalist, because my passion remained with outer space and because my underlying desire was to contribute to a greener world, not to contribute to oil and gas exploration. Also a small sidestep to the financial world did not quench the thirst, despite the beauty of witnessing the sensation that occurs when Japanese and Dutch business cultures meet. I needed the academic depth again.

Then came December 2012, when Prof. dr. ir. Ramon Hanssen, who I already knew from the studies Geomatics and Aerospace Engineering, turned up on my radar again by posting two PhD positions on 'microwave and optical remote sensing for a biobased society'. In short, the positions had everything I longed for: green, space, brains and abroad¹. I started in 2013, with a bare two-pager of the PhD project in one hand and a laptop in the other. This was entrepreneurship and academics in one. For both the Geoscience & Remote Sensing (GRS) department at the TU Delft and me, the research on biomass and land cover monitoring through the integrative use of radar and optical satellite imagery was new. It was uncharted territory for us. Fortunately, several international academic groups had published their work, allowing for literature reviews. And there was my academic hero, Dr. Lorenzo Iannini, who (back then) had recently started a postdoc position on the same subject, though he was also new to the subject. It was up to us then.

The project arose from a cooperation between BE-Basic², a public-private partnership that aims at building a sustainable biobased economy, and the TU Delft and Unicamp. The biobased economy is based on the principle that the resources for the production of chemicals, energy and materials come from biomass rather than conventional (fossil) resources. Founded in 2010, the foundation BE-Basic is funded by the Dutch Ministry of Economic Affairs, Agriculture and Innovation and by industrial part-

¹There was already a partnership established with the University of Campinas (Unicamp) in Brazil.

²visit be-basic.org for more information

ners and knowledge institutions. Within the BE-Basic framework, a dozen so-called Flagships were defined, each focusing on specific aspects of the biobased economy, ranging from R&D for production technology to societal embedding. One of these Flagships covers the socio-economic and sustainability impacts that are associated with the transition to a biobased economy. Among other research areas, it fosters the research on efficient and competitive biomass production in a sustainable manner, particularly focusing on productivity increase and wide-scale implementation of crops that can be converted into usable energy and the production of materials. In order to conduct cost-efficient monitoring of crop productivity and monitoring of wide-scale crop expansion over long time series, BE-Basic involved the GRS department of TU Delft, because of its knowledge position on developing and deploying satellite-based monitoring techniques. In addition, the foundation partnered with Unicamp's Geoprocessing Department of the Faculty of Agricultural Engineering (FEAGRI), because of its area of expertise on sustainable agriculture through the use of remote sensing data and geographic information systems in combination with agricultural modeling. Unicamp's involvement brought specific research opportunities to this research due to its scientific vantage point within the Brazilian state of São Paulo, a major international player in the production of bio-energy through sugarcane.

Back to the PhD process, the next years were amazing. The work switched between cracking data by endless programming and brute-force high performance computing, understanding the signals from radar and optical satellite sensors, learning how an extraordinary crop like sugarcane grows, setting up and maintaining international partner relations, disseminating results to international audiences and political figures, writing proposals and finally keeping the stakeholders satisfied. A quick glance at the scripts I produced, reveals that the total number of lines easily exceeds 100,000 lines of code (which may in fact be interpreted as a sign of hard labor as well as inefficient coding by the way). In any case, the results never came easily and the publications even less effortlessly. During the first year we explored the research framework, reviewed related literature and identified the research opportunities. The second year I took the step to move to Campinas, Brazil, to conduct collaborative research with fellow Brazilian PhD students, and to set up and carry out extensive ground measurement campaigns. Acquiring ground reference datasets over dozens of days, sometimes in 35°C while fully clothed against the razor sharp leaves and for twelve hours on duty, is tough. Nevertheless, the stay was the cherry on the pie; academically, socially and emotionally. The initial plan was to stay one year with two family members; it became 15 months and with three family members.

Back in the Netherlands, papers needed to be produced. However, making the remote sensing data work with the ground reference data was tedious and hard work. The (post-)processing also took an incredible amount of time, let alone crafting and tweaking the models, interpreting the results and writing down the matter. Inherent to many PhD projects, loads of results have never made it to the finish. These range from quick try-outs to scientifically sound outcomes; all were submissive to the wrath of time. After four years, I worked on the PhD thesis and carried out additional activities, including writing proposals for grants (of which one was granted). After five years I started working elsewhere and had to work on it in the evenings. In terms of academic output, the

result is (as first author) three published journal articles, one submitted journal article, four published conference proceedings and two published datasets.

The process of coming to this dissertation has been wonderful yet intense. Therefore I bow my head in gratitude to all who have managed to remain patient and confident. This holds particularly for Sanne, who I deeply admire for her acceptance to the obligations that emerged during the journey. You're the best. In addition, I am thankful to family and friends who kept inquiring me on the status of my research and kept looking forward to the big day. This holds particularly true for my late grandfather who missed the defence by one month. I dedicate this dissertation to my family and especially to him. He was fond of nuances that are made in science, particularly when introduced by the term 'however'. I am sure this dissertation meets this criterion.

In terms of scientific contribution, I am most grateful to Lorenzo Iannini, who relentlessly helped me in setting the proper statistical and mathematical frameworks of the research, while simultaneously making sure I would leave the discussion with more questions than I had started with. Lorenzo, over time you have become a dear friend. Thanks, pal. You have really contributed to the work. Thanks for the insane adventures we managed to position ourselves in and then survived (e.g. slums, Brotas, Saint-Ferréolles-Neiges), and thanks for the tears of laughter when needed most (e.g. perceived spiders, mosquitoes, jaguars, harvesters when it was only the wind or the tip of a sugarcane leaf). Talking about Brazil, *os meus chapas, galo!* I have had a splendid time in your fantastic country. Thanks for the warm welcome to us gringos, we will always remain close. Most of all, thank you for helping with the field measurements. I hope the moments at Rancho da Linguíça compensated for some of the pain in the field. Also I am indebted to my other university colleagues, including Dr. Paco López Dekker for his ever lasting scientific enthusiasm and valuable contribution to the SAR-precipitation paper, and Prof. dr. Jansle Vieira Rocha, Dr. Rubens Lamparelli and Prof. dr. Paulo Graziano for their expert opinion on sugarcane monitoring, their valuable connections with research partners and their hospitality at Unicamp. Also, thank you Jansle for the delicious feijoada vegetariana and Rubens for the Frango Assado. And, last but definitely not least, of course Ramon, for accepting me as PhD candidate in the first place, for the freedom you gave me to design and execute the research, for the opportunity to conduct research in Brazil, for relying on me to finish, for extending my contract and finally to help me making this dissertation right.

*Ramses Amadeus Molijn
Delft, January 2020*

1

INTRODUCTION

Tout est vivant, tout est lié. L'eau, l'air, le sol, les arbres. Toute la magie du monde se déroule sous nos yeux.

Les arbres respirent et renvoient l'eau du sol en brume légère dans l'atmosphère. Ils forment un manteau qui atténue la violence des pluies. Ils apportent l'humidité nécessaire à toute vie. Les forêts stockent le carbone. Elles en contiennent davantage que toute l'atmosphère de la terre. Elles sont l'une des pierres angulaires de l'équilibre climatique dont nous dépendons tous. Les arbres des forêts primaires abritent les trois-quarts de la biodiversité de la planète. C'est-à-dire de tout ce qui vit sur terre. Ces forêts apportent les remèdes qui nous guérissent. Les substances que secrètent ces plantes, notre corps peut les reconnaître. Nos cellules parlent le même langage. Nous sommes de la même famille.

Yann Arthus-Bertrand

BACKGROUND AND MOTIVATION

IN a biobased society, the resources for the production of chemicals, materials and energy are based on biobased renewable resources (i.e. biomass) instead of fossil fuels-based resources [1–4]. Biomass can be defined as raw materials of vegetative origin and generally includes all organic material coming from plants or plant-derived materials. Biomass types that are used for biobased products typically comprise agricultural crops (e.g. sugarcane, corn, grain), green biomass (grasses), wood biomass (e.g. pine, eucalyptus), aquatic biomass (algae, sea weed) and leftover biomass released during harvesting (e.g. hay, manure), during the production process (e.g. corn fiber, potato peels, wood branches) and after product use (e.g. sewage sludge, textile) [3, 5, 6]. Biomass as feedstock for biobased products has advantages over fossil resources (principally renewable and CO₂ less intensive) but also disadvantages (e.g. competing with food sources, deforestation, loss of biodiversity and increased local water use) [1, 2, 7–10].

A wide range of studies that are related to the development of the biobased sector have been published, from economic and technological feasibility analyses [5, 11–13]

to social and sustainability impact assessments [1, 10, 14, 15]. Part of the sustainability issues associated with biobased solutions relates to the environmental impacts from feedstock production, for example for bio-energy and biochemicals, and focus on direct or indirect land use change (LUC) [9, 11, 14]. As the need for certain crops increases and their acreages expand over time, other land cover types are replaced and moved out to other regions. As an example, the plantation of suitable crops as feedstock for biofuels can displace grassland to other regions, indirectly leading to stress on natural vegetation [9, 16, 17].

Bio-energy and biofuels form one of the most prominent sectors within a biobased economy from a research and market perspective [1, 3, 5, 9, 18]. Worldwide, sugarcane is the number one crop in terms of production quantity and one of the main crops used for bio-energy production [19, 20] due to its highly competitive characteristics. Illustratively, the ethanol production efficiency per unit area and fossil energy ratio (output biofuel energy over fossil fuel input energy) are (indicatively) 8000 liter per hectare and 10 for sugarcane, respectively, as compared to 4000 and 1.5 for maize [20, 21], which is another major bio-energy crop in terms of worldwide biofuel production quantity [22]. Brazil is the largest sugarcane-growing nation and a global leader in biofuel production [22]. São Paulo state, roughly six times larger than the Netherlands, is responsible for 60% of Brazil's sugarcane growth [17, 23–25], while only covering 3% of Brazil's territory. During the last decades, this crop expanded by more than 100% in this region and now approximates 6 million hectares [23, 26], which is almost 1.5 times the area of the Netherlands. Sugar and bio-ethanol are the most dominant products from sugarcane and the latter allows Brazil to substitute roughly half of the automotive's gasoline consumption [20]. Due to its dynamic land cover change over the last decades, mainly caused by a rapid expansion of sugarcane acreage, and its position as bio-energy producer, São Paulo state has been a subject of a myriad of studies [8, 9, 17, 25, 27, 28].

In relation to regular food crops, sugarcane is substantially different for several reasons. First of all, sugarcane is a perennial crop, whereby the growth period is approximately one year and the plant regrows after harvest from the remaining root systems. After five to seven of these so-called ratoon cycles, the decrease in yield makes it more economical to plant new shoots [29]. Furthermore, the crop is densely planted and grows to approximately four meters high, making it a crop with one of the highest biomass values per hectare (>100 tons per hectare are not uncommon in Brazil [21]). In comparison, annual crops, including most vegetables, grains and grasses, grow and die within several months and produce tens of tons per hectare. To give an impression of the characteristics of sugarcane, Figure 1.1 illustrates several growth stages, accompanied by photos that were collected during the acquisition of field measurements for this research.



Figure 1.1: Visual representation of sugarcane growth and collection of ground measurements: (1) marking measurement location and taking the geolocation for revisits; (2) and (3) state of sugarcane after two months (approximately 0.5 meter high) and typical row spacing; (4) state of sugarcane after four months (approximately 1.5 meter high); (5) state of sugarcane after 7 months (approximately 2.5 meter high); (6) canopy of sugarcane after 7 months; (7) state of sugarcane after 10 months (approximately 3.5 meter high); (8) measurement of sugarcane stalk thickness (several centimeter); (9) weighing of sugarcane biomass; (10) collection of sugarcane biomass from the field; (11) sugarcane before grinding; (12) sugarcane after grinding and before drying in the oven.

In order to monitor crops for their productivity and wide-scale expansion over long periods of time, satellite-based techniques have proven to be effective and cost-efficient [30–36]. The reasons for selecting such monitoring techniques are multifaceted. For decades, various international space agencies and private companies have acquired observations of the Earth's surface by using various kinds of space-borne sensors and platforms, providing extensive sets of global observations over long time spans [37–44]. Consequently, a mixture of datasets with satellite imagery is available, differentiated mainly by acquisition technique, by their spatial and temporal resolution and by their accessibility for public use. The type of acquisition technique depends mainly on the sensor type and determines the physical principles of the remote sensing signals. As an apt example, observations from optical and synthetic aperture radar (SAR) sensors differ fundamentally since their distinct wavelengths interact differently with the sensed objects [31, 36, 45, 46]. Optical sensors passively record radiation in the (near-)visible spectrum, whereby the signal's sensitivity to atmospheric disturbances (like clouds and aerosols) causes inconsistent temporal availability of observations of ground features. In addition, the short wavelength causes the signal to be mainly sensitive to the planar features of objects, such as the upper canopy layer of plants [32, 36, 47–49]. In contrary, SAR sensors emit radiation actively in the microwave spectrum, allowing the signals to penetrate clouds and hence temporally consistent observations of ground features can be acquired. In addition, the signal's longer wavelength increases its sensitivity to the volumetric and dielectric features of the sensed objects, such as the leaves, branches, stem and water content of plants. Consequently, such characteristics influence the availability and diversity of observations, allowing for a system-of-systems approach [50–53]. In such a system, the characteristics of the available sensors can be exploited such that the various remote sensing products can complement each other and can be integrated for improving the final results.

In light of such a system, at the start of this study, one particular game-changing satellite mission was introduced to the remote sensing community: Sentinel-1 [31, 54, 55]. Developed and funded by the European Space Agency (ESA), the first mission's satellite, Sentinel-1A, became operational in October 2014 (one and a half year after the start of this study). It is a C-band SAR satellite that collects imagery practically continuously, revisiting virtually each (land) location on Earth at least once every 12 days initially and at least once every 6 days after its duplicate satellite, Sentinel-1B, became operational two years later. The data products are made available based on a complete, open and free-of-charge basis. This marked a turning point for the SAR community, since such SAR data products were commonly limited in time and space (as was offered by ESA's ERS and Envisat missions until 2012 [54]) or limited in data availability, whereby users either needed to pay or apply for research grants with quota on ordering scenery (as was offered by the Canadian Space Agency's Radarsat missions). Hence, for the first time in history, a set of SAR remote sensing data emerged over large areas, based on consistent and relatively high sampling frequency, and freely available. For space-borne observations from optical sensors, such availability was already provided for by NASA's Landsat missions, offering free global data products that were (and continue to be) acquired over several decades [30, 56, 57]. Its most recent satellite, Landsat-8, observes virtually each (land) location on Earth at least every 16 days, since 2013. Data from these two satellites

(Sentinel-1 and Landsat-8) form the bulk of the observations that were used in this study. Data from other satellites were analyzed as well, but to a lesser extent due to their more limited temporal and spatial availability.

The availability and variety of space-based remote sensing platforms over the last decades have resulted in a handful global (and regional) land cover maps featuring several land cover types, including (generally) crops, natural vegetation, water and urban, and featuring spatial resolutions ranging between tens of meters to a kilometer and update frequencies between single stationary maps to near-real-time maps [26, 37–44]. These and crop productivity products were, at least at the commencement of this study, usually based on optical-only remote sensing data rather than based on a combination of optical-data and SAR-data [37, 58] and offered limited classification taxonomy. As will be discussed in more detail in the next chapters, it is the recent increase in availability of frequent SAR measurements, as provided by Sentinel-1, that was essential for the investigation into contribution of these measurements to the quality and features of land cover and productivity maps, whereby aspects such as spatial and temporal resolution (e.g. related to insensitivity to cloud coverage), classification accuracy and classification taxonomy (e.g. related to the sugarcane class), and biophysical feature retrieval (e.g. related to biomass) are focused on.

The characteristics of sugarcane (in particular its long growth period and high biomass) introduces opportunities as well as challenges when mapping and monitoring its growth through SAR and optical remote sensing data. As will be outlined further in the following chapters, the opportunities are mainly related to the availability of frequent remote sensing observations and the distinct growth and harvest patterns. The challenges are mainly related to the saturation of the remote sensing signals and the relatively high homogeneity of biomass in the field.

RESEARCH FRAMEWORK

RESEARCH OBJECTIVE

In this study, rather than focusing on the assessment of environmental impacts from sugarcane expansion and rather than developing global land cover products, the emphasis was put on the design and application of techniques that combine imagery from space-based SAR and optical sensors for monitoring the sugarcane production and expansion, whereby São Paulo state in Brazil is the main study and application area. Hence, the study was narrowed down to the following research objective:

*To advance the state-of-the-art on **crop productivity monitoring and land cover monitoring**, with a focus on **sugarcane in São Paulo state, Brazil**, through the **integrative use of remote sensing data from space-based optical and SAR sensors**, acknowledging the underlying mission to contribute to competitive biomass production in a sustainable manner for a **biobased society**.*

RESEARCH APPROACH

In order to fulfill the objective, first of all, the state of current techniques and their applications were examined. Without going into detail on the literature reviews, as these are presented in each of the following chapters, the use of SAR remote sensing data for pro-

ductivity and land cover monitoring is particularly dominated by vegetation types such as annual crops, grasslands and forests rather than sugarcane. This can be explained by the limited availability of SAR data until the introduction of Sentinel-1 in combination with the challenging characteristics of sugarcane, as were both touched upon in Section [Background and motivation](#). With the introduction of Sentinel-1 and the availability of optical data opening up possibilities for intense monitoring of sugarcane and land cover mapping over large areas based on dense time series, the processing and analyzing of such extensive datasets required high performance computing (HPC) resources. In this study, these were mostly facilitated by the Dutch supercomputer owned by the SURF co-operative¹ and by Google Earth Engine², which is a computing platform for geospatial analysis on Google's infrastructure.

Based on the literature reviews and the access to suitable remote sensing data and to computing resources, we identified the techniques that needed to be developed. In particular, the focus was put on analyzing the similarities and differences between optical and SAR remote sensing signals and on how these could be effectively combined for the areas of applications. Without going into detail on the technique selection and explanation, as these are presented in each of the following chapters, appropriate ground reference datasets were required in order to train and validate the techniques. However, when assessing the published datasets and datasets that were available at the involved and partnering institutions, we concluded that none was sufficiently adequate for both the growth monitoring and the land cover mapping studies we envisioned. This was based on the main requirements the dataset would need to fulfill: (i) the dataset should span the upcoming observations by Sentinel-1, (ii) the need for specific types of plant biophysical measurements (for growth monitoring) and for specific types of land covers (for land cover monitoring), and (iii) the presence of these measurements in São Paulo state, Brazil. As a consequence, ground reference data needed to be collected, which allowed us to shape the measurement campaigns in line with the remote sensing acquisition schemes and to conduct quality assessments on the datasets. Finally, although the study focuses primarily on the combination of both optical and SAR data, one chapter is dedicated to a research on SAR-only data in order to provide a detailed analysis of the potential of the spatially and temporally dense Sentinel-1 measurements for advanced land cover mapping.

RESEARCH QUESTIONS AND OUTLINE

Based on the described background, motivation, research objective and research approach, the following research questions were formulated. Each of them is related to a chapter in this study:

1. Chapter 2: How can ground reference measurements that describe the biophysical state of sugarcane be acquired to support the analysis of space-based SAR and optical measurements for productivity monitoring?
2. Chapter 3: How can measurements from space-based SAR and optical sensors be exploited to effectively monitor sugarcane productivity?

¹visit [SURFsara](http://SURFsara.nl)

²visit earthengine.google.com

3. Chapter 4: How can the combined use of space-based SAR and optical measurements improve their singular use for land cover monitoring with a focus on sugarcane classification?
4. Chapter 5: How can a-priori knowledge on precipitation and soil type be used advantageously for the characterization of vegetation states through space-based SAR measurements?

Each of the following chapters consists of research analyses that were publicly disclosed by means of three published journal articles, one submitted journal article, two published conference proceedings and two published datasets. In order to place each of these publications into context within this study, each will be introduced at the beginning of the respective chapter. Some articles that were published during this study are not included in this study, see [List of Publications](#).

REFERENCES

- [1] P. Rafiaani, T. Kuppens, M. V. Dael, H. Azadi, P. Lebailly, and S. V. Passel, *Social sustainability assessments in the biobased economy: Towards a systemic approach*, [Renewable and Sustainable Energy Reviews](#) **82**, 1839 (2018).
- [2] I. Muñoz, K. Flury, N. Jungbluth, G. Rigarlsford, L. M. Canals, and H. King, *Life cycle assessment of bio-based ethanol produced from different agricultural feedstocks*, [The International Journal of Life Cycle Assessment](#) **19**, 109 (2013).
- [3] J. E. van Dam, B. de Klerk-Engels, P. C. Struik, and R. Rabbinge, *Securing renewable resource supplies for changing market demands in a bio-based economy*, [Industrial Crops and Products](#) **21**, 129 (2005).
- [4] BE-Basic, [BE-Basic Foundation - Annual report 2015](#), Tech. Rep. (BE-Basic Foundation, Delft, 2015).
- [5] C. Panoutsou, S. Manfredi, and B. Kavalov, *Biomass resource efficiency for the biobased industries*, Tech. Rep. (European Commission - Joint Research Centre, Ispra, 2013).
- [6] C. Fritsch, A. Stäbler, A. Happel, M. A. Cubero, G. Belotti, I. Aguiló, M. Abadias, M. Gallur, I. M. Cigognini, A. Montanari, M. J. Lopez, F. Suarez-Estrella, N. Brunton, L. Sisti, and M. Ferri, *Agri and food waste valorisation co-ops based on flexible multi-feedstocks biorefinery processing technologies for new high added value applications*, Tech. Rep. (European Communities, Madrid, 2017).
- [7] H. Chum, A. Faaij, J. Moreira, G. Berndes, P. Dhamija, H. Dong, B. Gabrielle, A. G. Eng, W. Lucht, M. Mapako, O. M. Cerutti, T. McIntyre, T. Minowa, K. Pingoud, R. Bain, R. Chiang, D. Dawe, G. Heath, M. Junginger, M. Patel, J. Yang, E. Warner, D. Paré, S. K. Ribeiro, K. Seyboth, P. Matschoss, S. Kadner, T. Zwickel, P. Eickemeier, and G. Hansen, *Bioenergy*, in [Renewable Energy Sources and Climate Change Mitigation: Special Report of the Intergovernmental Panel on Climate Change](#), In IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation, edited by

- C. von Stechow, G. Hansen, K. Seyboth, O. Edenhofer, P. Eickemeier, P. Matschoss, R. Pichs-Madruga, S. Schlömer, S. Kadner, T. Zwickel, and Y. Sokona (Cambridge University Press, Cambridge, 2011) pp. 209–332.
- [8] A. Walter, M. V. Galdos, F. V. Scarpore, M. R. L. V. Leal, J. E. A. Seabra, M. P. da Cunha, M. C. A. Picoli, and C. O. F. de Oliveira, *Brazilian sugarcane ethanol: developments so far and challenges for the future*, [Wiley Interdisciplinary Reviews: Energy and Environment](#) **3**, 70 (2014).
- [9] F. van der Hilst, J. A. Versteegen, G. Woltjer, E. M. Smeets, and A. P. Faaij, *Mapping land use changes resulting from biofuel production and the effect of mitigation measures*, [GCB Bioenergy](#) **10**, 804 (2018).
- [10] S. Sleenhoff and P. Osseweijer, *How people feel their engagement can have efficacy for a bio-based society*, [Public Understanding of Science](#) **25**, 719 (2016).
- [11] K. H. Zhuang and M. J. Herrgård, *Multi-scale exploration of the technical, economic, and environmental dimensions of bio-based chemical production*, [Metabolic Engineering](#) **31**, 1 (2015).
- [12] C. I. Santos, C. C. Silva, S. I. Mussatto, P. Osseweijer, L. A. van der Wielen, and J. A. Posada, *Integrated 1st and 2nd generation sugarcane bio-refinery for jet fuel production in Brazil: Techno-economic and greenhouse gas emissions assessment*, [Renewable Energy](#) **129**, 733 (2018).
- [13] C. M. Alves, M. Valk, S. de Jong, A. Bonomi, L. A. van der Wielen, and S. I. Mussatto, *Techno-economic assessment of biorefinery technologies for aviation biofuels supply chains in Brazil*, [Biofuels, Bioproducts and Biorefining](#) **11**, 67 (2017).
- [14] G. M. Souza, M. V. R. Ballester, C. H. de Brito Cruz, H. Chum, B. Dale, V. H. Dale, E. C. Fernandes, T. Foust, A. Karp, L. Lynd, R. Maciel Filho, A. Milanez, F. Nigro, P. Osseweijer, L. M. Verdade, R. L. Victoria, and L. Van der Wielen, *The role of bioenergy in a climate-changing world*, [Environmental Development](#) **23**, 57 (2017).
- [15] M. P. Parada, L. Asveld, P. Osseweijer, and J. A. Posada, *Setting the design space of biorefineries through sustainability values, a practical approach*, [Biofuels, Bioproducts and Biorefining](#) **12**, 29 (2018).
- [16] A. Alkimim and K. C. Clarke, *Land use change and the carbon debt for sugarcane ethanol production in Brazil*, [Land Use Policy](#) **72**, 65 (2018).
- [17] B. F. T. Rudorff, D. A. de Aguiar, W. F. da Silva, L. M. Sugawara, M. Adami, and M. A. Moreira, *Studies on the Rapid Expansion of Sugarcane for Ethanol Production in São Paulo State (Brazil) Using Landsat Data*, [Remote Sensing](#) **2**, 1057 (2010).
- [18] H. Chum, A. Faaij, J. Moreira, G. Berndes, P. Dhamija, H. Dong, B. Gabrielle, A. G. Eng, M. Mapako, W. Lucht, O. M. Cerutti, T. McIntyre, T. Minowa, and K. Pingoud, *2011: Bioenergy*, in *Bioenergy*, In IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation, edited by O. Edenhofer, R. Pichs-Madruga,

- Y. Sokona, K. Seyboth, P. Matschoss, S. Kadner, T. Zwickel, P. Eickemeier, G. Hansen, S. Schlömer, and C. Van Stechow (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2011) pp. 209–332.
- [19] Food and Agriculture Organization of the United Nations, *Food and agriculture data (FAOSTAT)*, (2017).
- [20] P. Zuurbier and J. van de Vooren, *Sugarcane Ethanol - Contributions to climate change migration and the environment* (Wageningen Academic Publishers, Wageningen, 2008).
- [21] A. J. Wacławovsky, P. M. Sato, C. G. Lembke, P. H. Moore, and G. M. Souza, *Sugarcane for bioenergy production: An assessment of yield and regulation of sucrose content*, *Plant Biotechnology Journal* **8**, 263 (2010).
- [22] OECD/FAO, *OECD-FAO Agricultural Outlook 2018-2027*, Tech. Rep. (Food and Agriculture Organization of the United Nations, Rome, 2018).
- [23] L. G. T. Carpio and F. Simone de Souza, *Optimal allocation of sugarcane bagasse for producing bioelectricity and second generation ethanol in Brazil: Scenarios of cost reductions*, *Renewable Energy* **111**, 771 (2017).
- [24] P. G. Machado, N. A. M. Rampazo, M. C. A. Picoli, C. G. Miranda, D. G. Duft, and K. R. E. de Jesus, *Analysis of socioeconomic and environmental sensitivity of sugarcane cultivation using a Geographic Information System*, *Land Use Policy* **69**, 64 (2017).
- [25] A. W. A. Palludeto, T. S. Telles, R. F. Souza, and F. R. de Moura, *Sugarcane expansion and farmland prices in São Paulo State, Brazil*, *Agriculture and Food Security* **7**, 1 (2018).
- [26] INPE, *Canasat*, (2017).
- [27] J. Goldemberg, S. T. Coelho, and P. Guardabassi, *The sustainability of ethanol production from sugarcane*, *Energy Policy* **36**, 2086 (2008).
- [28] C. E. Caldarelli and L. Gilio, *Expansion of the sugarcane industry and its effects on land use in São Paulo: Analysis from 2000 through 2015*, *Land Use Policy* **76**, 264 (2018).
- [29] M. A. Vieira, A. R. Formaggio, C. D. Rennó, C. Atzberger, D. A. Aguiar, and M. P. Mello, *Object Based Image Analysis and Data Mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas*, *Remote Sensing of Environment* **123**, 553 (2012).
- [30] M. C. Hansen and T. R. Loveland, *A review of large area monitoring of land cover change using Landsat data*, *Remote Sensing of Environment* **122**, 66 (2012).
- [31] S. C. Steele-Dunne, H. McNairn, A. Monsivais-Huertero, J. Judge, P. W. Liu, and K. Papathanassiou, *Radar Remote Sensing of Agricultural Canopies: A Review*, (2017).

- [32] G. Zheng and L. M. Moskal, *Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors*, [Sensors](#) **9**, 2719 (2009).
- [33] M. Herold, L. See, N. E. Tsendbazar, and S. Fritz, *Towards an integrated global land cover monitoring and mapping system*, [Remote Sensing](#) **8**, 1 (2016).
- [34] A. Pérez-Hoyos, F. Rembold, H. Kerdiles, and J. Gallego, *Comparison of global land cover datasets for cropland monitoring*, [Remote Sensing](#) **9** (2017), 10.3390/rs9111118.
- [35] B. Wu, R. Gommers, M. Zhang, H. Zeng, N. Yan, W. Zou, Y. Zheng, N. Zhang, S. Chang, Q. Xing, and A. van Heijden, *Global crop monitoring: A satellite-based hierarchical approach*, [Remote Sensing](#) **7**, 3907 (2015).
- [36] S. Sinha, C. Jeganathan, L. K. Sharma, and M. S. Nathawat, *A review of radar remote sensing for biomass estimation*, [International Journal of Environmental Science and Technology](#) **12**, 1779 (2015).
- [37] G. Grekousis, G. Mountrakis, and M. Kavouras, *An overview of 21 global and 43 regional land-cover mapping products*, [International Journal of Remote Sensing](#) **36**, 5309 (2015).
- [38] J. Xiong, P. S. Thenkabail, J. C. Tilton, M. K. Gumma, P. Teluguntla, A. Oliphant, R. G. Congalton, K. Yadav, and N. Gorelick, *Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on google earth engine*, [Remote Sensing](#) **9**, 1 (2017).
- [39] J. Chen, J. Chen, A. Liao, X. Cao, L. Chen, X. Chen, C. He, G. Han, S. Peng, M. Lu, W. Zhang, X. Tong, and J. Mills, *Global land cover mapping at 30 m resolution: A POK-based operational approach*, [ISPRS Journal of Photogrammetry and Remote Sensing](#) **103**, 7 (2014).
- [40] J. Jokar Arsanjani, A. Tayyebi, and E. Vaz, *GlobeLand30 as an alternative fine-scale global land cover map: Challenges, possibilities, and implications for developing countries*, [Habitat International](#) **55**, 25 (2016).
- [41] E. Bartholomé and A. S. Belward, *GLC2000: a new approach to global land cover mapping from Earth observation data*, [International Journal of Remote Sensing](#) **26**, 1959 (2005).
- [42] M. A. Friedl, D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang, *MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets*, [Remote Sensing of Environment](#) **114**, 168 (2010).
- [43] P. Defourny, L. Schouten, S. Bartalev, S. Bontemps, P. Caccetta, C. D. Bella, V. Gond, G. W. Hazeu, A. Heinimann, M. Herold, J. Knoop, G. Jaffrain, R. Latifovic, H. Lin, A. Nonguierma, E. V. Bogaert, C. Vancutsem, P. Bicheron, M. Leroy, and O. Arino, *Accuracy Assessment of a 300 m Global Land Cover Map: The GlobCover Experience*, 33rd International Symposium on Remote Sensing of Environment, Sustaining the Millennium Development Goals, 1 (2009).

- [44] S. Bontemps, P. Defourny, E. V. Bogaert, V. Kalogirou, and J. R. Perez, *GLOBCOVER 2009 Products Description and Validation Report*, *ESA Bulletin* **136**, 53 (2011).
- [45] M. S. Moran, L. Alonso, J. F. Moreno, M. P. Cendrero Mateo, D. F. de la Cruz, and A. Montoro, *A RADARSAT-2 Quad-Polarized Time Series for Monitoring Crop and Soil Conditions in Barrax, Spain*, *IEEE Transactions on Geoscience and Remote Sensing* **50**, 1057 (2012).
- [46] H. McNairn and B. Brisco, *The application of C-band polarimetric SAR for agriculture: a review*, *Canadian Journal of Remote Sensing* **30**, 525 (2004).
- [47] V. Guissard, C. Lucau-Danila, and P. Defourny, *Crop specific LAI retrieval using optical and radar satellite data for regional crop growth monitoring and modelling*, *Remote Sensing for Agriculture* **5976**, 59760S (2005).
- [48] Z. Malenovský, K. B. Mishra, F. Zemek, U. Rascher, and L. Nedbal, *Scientific and technical challenges in remote sensing of plant canopy reflectance and fluorescence*, *Journal of Experimental Botany* **60**, 2987 (2009).
- [49] D. Lu, *The potential and challenge of remote sensing-based biomass estimation*, *International Journal of Remote Sensing* **27**, 1297 (2006).
- [50] A. Orynbaikyzy, U. Gessner, and C. Conrad, *Crop type classification using a combination of optical and radar remote sensing data: a review*, *International Journal of Remote Sensing* **40**, 6553 (2019).
- [51] C. Gómez, J. C. White, and M. A. Wulder, *Optical remotely sensed time series data for land cover classification: A review*, *ISPRS Journal of Photogrammetry and Remote Sensing* **116**, 55 (2016).
- [52] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, *Google Earth Engine: Planetary-scale geospatial analysis for everyone*, *Remote Sensing of Environment* **202**, 18 (2017).
- [53] M. E. Hajj, N. Baghdadi, M. Zribi, and H. Bazzi, *Synergic use of Sentinel-1 and Sentinel-2 images for operational soil moisture mapping at high spatial resolution over agricultural areas*, *Remote Sensing* **9**, 1 (2017).
- [54] R. Torres, P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Ö. Rommen, N. Floury, M. Brown, I. N. Traver, P. Deghaye, B. Duesmann, B. Rosich, N. Miranda, C. Bruno, M. L'Abbate, R. Croci, A. Pietropaolo, M. Huchler, and F. Ros-tan, *GMES Sentinel-1 mission*, *Remote Sensing of Environment* **120**, 9 (2012).
- [55] Z. Malenovský, H. Rott, J. Cihlar, M. E. Schaepman, G. García-Santos, R. Fernandes, and M. Berger, *Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land*, *Remote Sensing of Environment* **120**, 91 (2012).

- [56] D. P. Roy, M. A. Wulder, T. R. Loveland, W. C.E., R. G. Allen, M. C. Anderson, D. Helder, J. R. Irons, D. M. Johnson, R. Kennedy, T. A. Scambos, C. B. Schaaf, J. R. Schott, Y. Sheng, E. F. Vermote, A. S. Belward, R. Bindenschadler, W. B. Cohen, F. Gao, J. D. Hipple, P. Hostert, J. Huntington, C. O. Justice, A. Kilic, V. Kovalsky, Z. P. Lee, L. Lymburner, J. G. Masek, J. McCorkel, Y. Shuai, R. Trezza, J. Vogelmann, R. H. Wynne, and Z. Zhu, *Landsat-8: Science and product vision for terrestrial global change research*, *Remote Sensing of Environment* **145**, 154 (2014).
- [57] M. A. Wulder, J. G. Masek, W. B. Cohen, T. R. Loveland, and C. E. Woodcock, *Opening the archive: How free data has enabled the science and monitoring promise of Landsat*, *Remote Sensing of Environment* **122**, 2 (2012).
- [58] P. S. Thenkabail, *Land Resources Monitoring, Modeling, and Mapping with Remote Sensing* (CRC Press, 2015) pp. 791–831.

2

GROUND REFERENCE DATA FOR SUGARCANE PRODUCTIVITY MONITORING

The fear of appearances is the first symptom of impotence.

Fyodor Dostoyevsky through Rodion Raskolnikov

THIS chapter has been published in Scientific Data, Nature Research Journal, in 2018 [1]. The dataset itself was published in 2015 [2]. The aim of the research was to collect a ground reference dataset that allows for the detailed analysis of optical and SAR satellite measurements for productivity monitoring. The article is a dataset descriptor based on a set of ground reference measurements, consisting of various biophysical measurements, that were collected over the growth of several sugarcane fields in São Paulo state by TU Delft and Unicamp. The measurement acquisition plan introduces a methodology that is aimed at estimating biomass of sugarcane plants through non-destructive biophysical measurements in combination with allometric models. These models were calibrated based on destructive measurements, which were collected at locations that were excluded from detailed analyses of satellite measurements as presented in Chapter 3. The results are presented and explained in detail, together with several uncertainty measures that are related to the acquisition campaign. Following the publication, conversion errors in the final PDF document were observed for which an author correction was published. This correction can be found immediately following the original publication.

REFERENCES

- [1] R. A. Molijn, L. Iannini, J. V. Rocha, and R. F. Hanssen, *Ground reference data for sugarcane biomass estimation in São Paulo state, Brazil*, [Scientific Data 5, 1 \(2018\)](#).

- [2] R. A. Molijn, *Sugarcane ground reference data over four fields in São Paulo state, Brazil, taken in 2014 and 2015*, (2015), dataset, doi.org/10.4121/uuid:37112e18-f794-4d66-a8cd-7f1e92af09fc.

SCIENTIFIC DATA

OPEN

Data Descriptor: Ground reference data for sugarcane biomass estimation in São Paulo state, Brazil

Received: 4 January 2018

Accepted: 17 May 2018

Published: 7 August 2018

Ramses A. Molijn¹, Lorenzo Iannini¹, Jansle Vieira Rocha² & Ramon F. Hanssen¹

In order to make effective decisions on sustainable development, it is essential for sugarcane-producing countries to take into account sugarcane acreage and sugarcane production dynamics. The availability of sugarcane biophysical data along the growth season is key to an effective mapping of such dynamics, especially to tune agronomic models and to cross-validate indirect satellite measurements. Here, we introduce a dataset comprising 3,500 sugarcane observations collected from October 2014 until October 2015 at four fields in the São Paulo state (Brazil). The campaign included both non-destructive measurements of plant biometrics and destructive biomass weighing procedures. The acquisition plan was designed to maximize cost-effectiveness and minimize field-invasiveness, hence the non-destructive measurements outnumber the destructive ones. To compensate for such imbalance, a method to convert the measured biometrics into biomass estimates, based on the empirical adjustment of allometric models, is proposed. In addition, the paper addresses the precisions associated to the ground measurements and derived metrics. The presented growth dynamics and associated precisions can be adopted when designing new sugarcane measurement campaigns.

Design Type(s)	observation design • time series design
Measurement Type(s)	plant structure • leaf area index • plant matter • water-based rainfall
Technology Type(s)	data collection method
Factor Type(s)	
Sample Characteristic(s)	Saccharum hybrid cultivar RB867515 • Piracicaba Mesoregion • cropland biome • Saccharum hybrid cultivar SP80-3280

¹Geoscience and Remote Sensing, Delft University of Technology, Delft 2628CN, The Netherlands. ²FEAGRI, Unicamp, Campinas 13083-875, Brazil. Correspondence and requests for materials should be addressed to R.M. (email: r.a.molijn@tudelft.nl).

Background & Summary

Sugarcane is the number one crop worldwide in terms of production quantity. It provides for more than 40% of the car fuel in the largest sugarcane producing nation, Brazil. São Paulo state is responsible for more than half of the national bio-ethanol production from sugarcane, which expanded more than 100%, 2 million hectares, over the last 15 years. The crop is usually grown for five to seven years from the same ratoon¹, i.e. the plant sprouts from the same root system after each harvest event. The cycle of sugarcane can vary from 8 to 24 months depending on the cultivars and planting date, with the 12 month and 18 month types dominating the Brazilian plantations². The growth stages after planting consist of the germination phase, when the first leaves and stems that are part of the root system emerge, the tillering and canopy development phase, indicatively from the second to the fifth month when secondary stems arise above ground, the grand growth phase, between the fifth and eighth month, and finally the maturing and ripening phase until harvest, see the left part of Fig. 1. The sucrose content, from which the bio-ethanol and sugar is produced, is concentrated in the stem, also referred to as stalk, of the sugarcane. Sugarcane yield is commonly measured in tons cane per hectare (TCH)^{3–6} and ranges up to 160 TCH^{6,7}.

In order to improve the understanding of the plant itself and of the related impacts, several sugarcane measurement campaigns have been conducted in Brazil for a wide range of studies, most notably including biological analysis of plant processes^{6,8–13}, agronomic analysis on the effects of different fertilization and soil practices^{4,5,14–16} and bio-physical parameter extraction from space-based remote sensing signals^{7,10,17–22}. Due to the plant morphology and its growing environment, sugarcane measurements are labour-intensive and easily affected by significant errors. Quantifying uncertainty metrics in the collected and processed measures is extremely valuable in data modeling and data integration frameworks. However, so far, only a few contributions^{5,6,19,23} explicitly accounted for the quality of their reference ground data.

With such premise, we planned a ground surveying and data elaboration strategy that could entail an in-depth uncertainty analysis. The purpose for conducting the measurement campaign was to investigate the sensitivity of optical and radar satellite imagery to sugarcane biophysical features along the whole growth cycle. Nevertheless, the produced dataset can valuably fit a variety of other sugarcane studies. The biophysical parameters of interest include: cane height, cane density, leaf area index (LAI), cane biomass and leaves biomass. The measurements were carried out at four fields in the São Paulo state, selected for their heterogenous conditions in terms of crop health and ratoon cycle. The fields were surveyed multiple times at different development stages. The campaign addressed two additional needs: that of minimizing the measurements' effects on the remote sensing signals and that of maximizing efficiency in terms of costs. Both demands led to the limitation of the destructive biomass procedures. As a result, the majority of the surveying locations, denominated Elementary Sampling Units (ESU), comprise only biometric measurements. The biomass weighting operations were then added to a selected number of locations, hence recalled Elementary Sampling Units with Biomass (ESUB). Consequently, a method to retrieve biomass from non-destructive biometric measurements at ESUs was adopted. In this method, the destructive measurements at ESUBs are used to calibrate the allometric models and to cross-validate their outcome. The correlation coefficient between the directly measured and estimated biomass values is 0.89. The selection of the surveying locations and the data collection and processing procedures will be explained in the sec:Methods section.

The uncertainty analyses (expressed in precisions, see the sec:Technical-Validation section) are based on spatially-intensive repetitive measurements and on the error propagation laws applied to the biomass equations. For cane biomass, it was expected and shown that the instrument operation precision is

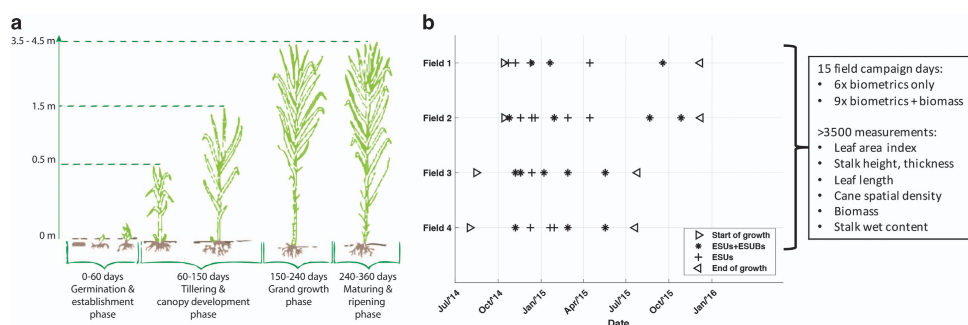


Figure 1. Sugarcane phenological development and data acquisition scheme. (a) sugarcane growth stages with indicative heights and plant geometries, adapted and modified from ref. 27; (b) ground reference data acquisition scheme, accompanied by a summary of the data collection.

smaller than the precision associated to the local plant variations. The precision related to model fitting and data interpolation procedures contributes significantly to the total precision, but it reduces over time. The relative precision on indirect biomass measures is hence lower for mature cane. As expected, the leaf biomass precisions dominate the total precision at an early stage but are quickly taken over by the stalk-related precisions.

Finally, based on the findings, we like to make some recommendations on setting up novel ground measurement campaigns. First of all, planning the ESU and ESUB locations for taking ground measurements requires thorough preparation beforehand, predominantly counting for the costs of the repetitive measurements needed to derive the associated precisions. Based on high-resolution optical images one should select those ESU locations for which the difference in biophysical parameter is expected to be statistically significant within the fields of interest. The proposed techniques as explained in this paper can be used accordingly.

Methods

The relationships between the methods and the corresponding figures are schematized in Fig. 2. Details on the practices and additional information on the linkages will be given in the following sections with references to the numbers in the figure.

Selection of study area and fields

The sugarcane fields of interest were proposed by the fields' owner, one of the largest sugarcane producers and energy companies in Brazil, based on their availability, accessibility and quality. The fields are located in an area densely covered by sugarcane plantations, spread around the Bom Retiro sugarcane mill, twenty kilometers from Piracicaba, São Paulo state, Brazil, see Fig. 3. They are mostly flat and contain inter-field and intra-field roads of two to ten meters wide. The fields' characteristics, see Table 1, show that there are differences in ratoon cycles, area and start and end of growth. All fields are situated on clay grounds.

In order to select the locations to measure, we visually analyzed optical images from Landsat-8 from previous years and identified consistent anomalies in NDVI values, which are indicative for sugarcane biomass differences^{7,24}. Subsequently, we split the locations into elementary sampling units (ESUs) for the regular biophysical parameter measurements and elementary sampling units for biomass (ESUBs). Three to five ESUs and one or two ESUBs per field were selected, see Fig. 4. These ESUs and ESUBs were visited multiple times during one growth season, see Fig. 1, using a GPS location device featuring approximately five meters location accuracy.

Ground reference data acquisition

The measurements in each ESU were taken systematically following a strict procedure. Fig. 5 illustrates schematically the methodology for taking the LAI, biometrics and biomass measurements, as explained in the following subsections. Distinction is made between plant, cane, stalk and leaves, whereby the plant is the collection of canes emerging from the same root system and cane is the integral of the stalk and leaves. The temporal developments of the measured parameters are illustrated in Fig. 6. The different behaviors in time, as indicated by the illustrative trend lines, are addressed at each parameter description.

Biometrics

At all four corner points the following biometric measurements were taken:

- stalk heights of highest canes. The stalk height is here defined as the length from the stalk at ground level to the crown from which the upper leaves emerge,
- stalk thicknesses of the same canes, taken at the stalk on ground level,
- longest leaf lengths of the same canes,
- number of canes and plants per meter by counting the number of canes and plants over five meters distance for two adjacent rows, as a measure for cane spatial density.

The highest canes were selected by eye. As will be explained in subsequent sections, the variations in modeled intra-field biomass values are dictated by the measured cane spatial density at the ESUs. Hence, the selection of the highest canes mainly serves to establish the general biomass curve and subsequently the field-wide cane biomass interpolators (see Section subsec:Cane-biomass-interpolators). This addresses the potential issue that the selection of the measured (highest) canes may not be representative for estimating the cane biomass per unit area (expressed in tons cane per hectare-TCH) at the measurement locations. The dataset also includes the measurements of smaller canes at each measurement location. The smallest canes that are unrepresentative for the sugarcane at the measurement location were not taken into account, for example sprouting canes amongst mature canes. These, however, are not used in any further modeling or estimation.

In addition to the biometrics, photos embedded with the geo-location were taken for location verification, for checking the leaf angle distribution and for future referencing. The indicative trend lines come from a fitted linear function for height and a logistic function, showing stabilization after a certain moment in time, for stalk thickness (increasing in time). For cane spatial density, a logistic function as a

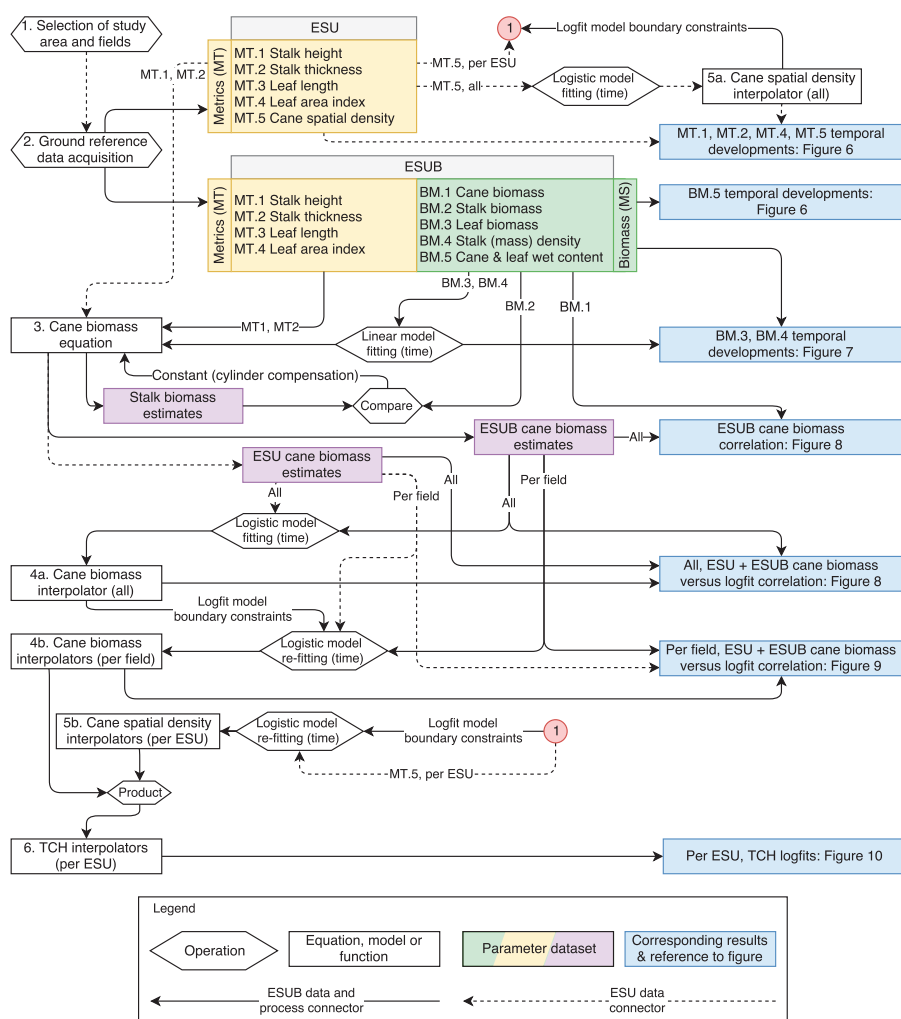


Figure 2. Workflow indicating the operations, relations between the methods and references to the corresponding resulting figures.

function of time was fitted as well, which is used in a later stage to constrain the re-fitted functions at ESU-level, see “4a” and “4b” in Fig. 2. This will be explained later in this section in more detail. The shapes of the resulting curves are in line with the reported trends in^{11,13,19}, although the cane spatial density in these studies first shows a slight increase in the first 150 days before the decrease. For the cane height and stalk thickness the functions are constrained to pass through the zero origin.

Leaf Area Index (LAI)

For LAI measurements, a LICOR LAI-2000 Plant Canopy Analyzer was used with the default calibration values as reported in the LAI manual²⁵. The manual and²⁶ were consulted for setting up the following procedure. During the first growth stages of the sugarcane, until approximately two meters stalk height, the LAI measurements are taken from the middle of the ESU in four different directions with one above-canopy measurement followed by three below-canopy measurements; two times in cross-row direction and two times in along-row direction, see Fig. 5. After this point in time, the measurements showed

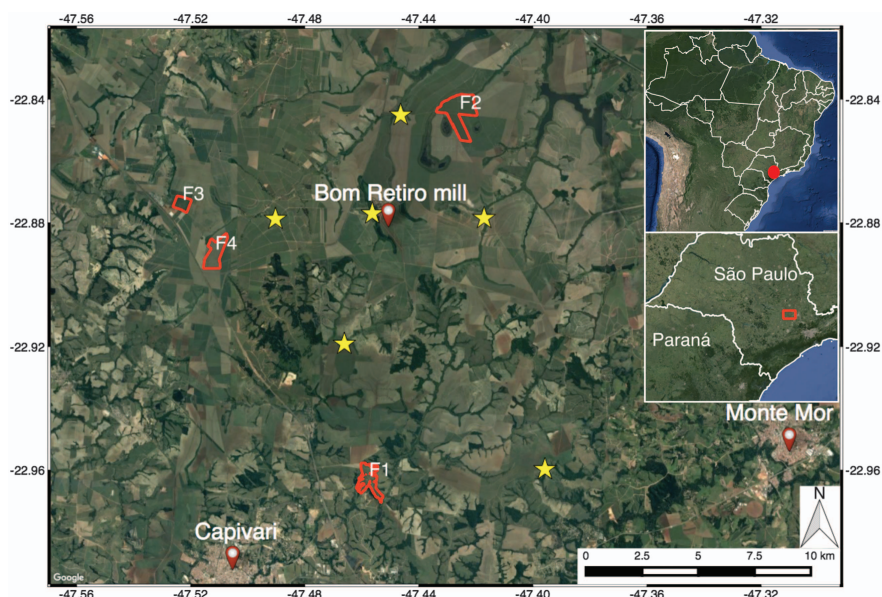


Figure 3. Study area, indicating the four fields of interest with red polygons and the weather stations with yellow stars.

negligible differences between the directions and we assumed the canopy coverage became planar isotropic; the coverage of the leaves is uniform in both horizontal directions. Consequently, the LAI measurements are taken with one above-canopy measurement followed by nine below-canopy, all of them within ten meters in row direction divided over three rows. As for the precise placement of the sensor for all below-canopy measurements, the distance to the plants was taken into account, i.e. per set of three measurements one measurement was taken adjacent to the cane, one at one third of the row spacing and the last at two third of the row spacing. LAI measurements in the direction of the road were avoided to minimize the possible road effects. Also the climatic conditions were taken into account by retaking above canopy measurements when clouds were moving. The resulting measured LAI can also be referred to as effective LAI since it is not corrected for the leaf clumping effect.

The correction from effective LAI to actual LAI was carried out as follows. Along two meters and for two rows wide ($1.5 + 0.9$ meters) the LAI was measured in four different directions and three measurements per direction with spacing of one meter (instead of three meters as for the ESU measurements). Within this area, around the center point all plants were cut and the green leaves were stripped from the canes, which were then placed on a white sheet. Through supervised classification with standard image analysis software, the area of the green leaves was divided by the total area, i.e. 2.4 by 2 meters. The average of the differences between the measured and observed LAI taken at two dates was subtracted from the LAI measurements. The correction results in a few LAI values to become slightly below zero.

The trend line for actual LAI in Fig. 5 is a second order polynomial fit, which is supported by results presented in ref. 11,13,19,23, and is forced through zero origin. The few LAI values smaller than zero are plotted as zero.

Biomass

In order to minimize destructive measurements at the ESUs, which would bias the remote sensing signals, biomass ESUs (ESUBs) were selected. Four whole plants (containing generally between ten and twenty canes) were cut at ground level and the number of canes per plant was counted. The stalk heights, stalk thicknesses and leaf lengths were measured per plant similarly as described above. The mass of the entire plant was measured separately and by dividing by the number of canes its average and variation in cane weight was obtained. This will be referred to in the sec:Technical-Validation section. After the weighing, per plant two representative canes were selected and combined, creating a sample of eight canes of the four plants. Leaves were then separated from the stalks and weighed separately, giving the leaf biomass per cane. The mass of these stalks divided by the volume, approximated by the cylindrical volume as a function of the measured stalk thickness and height (see Equation 1), gives the

Field name	Ratoon	Area [ha]	Start of growth	Harvest	Center coordinate [latitude, longitude]
F1	1st cycle	58	30/10/2014	07/10/2015	~22.9607, -47.4578
F2	1st cycle	115	14/10/2014	07/12/2015	~22.8558, -47.4228
F3	2nd cycle	25	15/08/2014	26/07/2015	~22.8738, -47.5230
F4	9th cycle	59	01/08/2014	21/07/2015	~22.8895, -47.5108

Table 1. Sugarcane fields’ characteristics with growth and harvest dates, which are for F1 and F2 not applicable to the entire field.

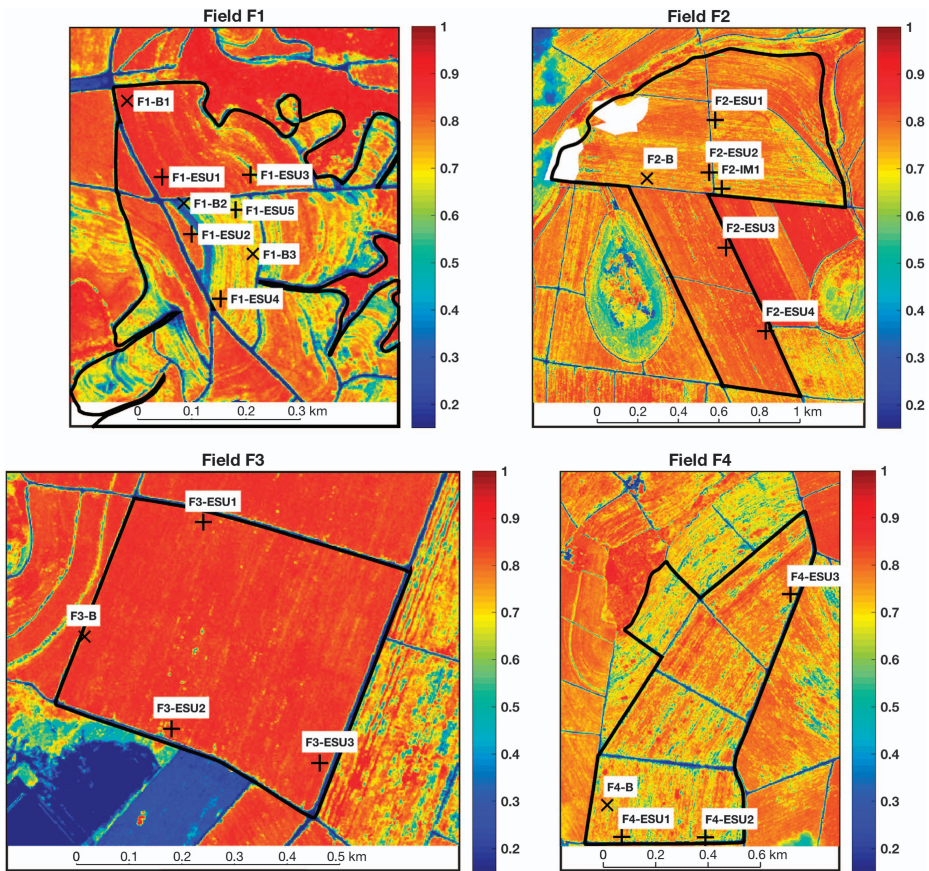


Figure 4. Delineation of the four studied sugarcane fields and ESU and ESUB (b) locations. The background map shows the NDVI from WorldView-2 on January 25, 2015. The plus signs (+) and crosses (x) show, respectively, the location of the elementary sampling units (ESUs) and elementary sampling units for biomass (ESUBs). The white patches in field F2 are masked clouds.

stalk (mass) density. Note that the minor compensation for the tapered geometry of the cylinder, referred to the scaling factor C in the equation, is not yet taken into account in Fig. 6. The remaining canes of the four plants, the stripped leaves and stalks were separately ground into centimeter-sized pieces, a sample was taken, weighed and dried in the oven at 65 °C for 72 h. The complement of the ratio between the mass of the dried and the mass of the wet samples gives the wet matter content, i.e. computed individually for the leaves, stalks and four different plants.

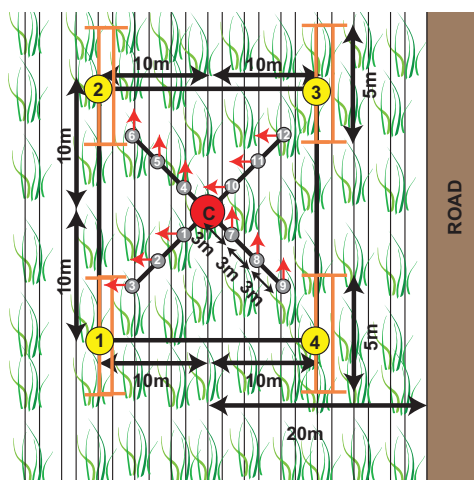


Figure 5. Schematic overview of the ESU measurements. The LAI measurements are taken at 12 different locations (grey circles) in four different directions (red arrows) around the center point (C), which lies 20 meters from the road. The biometric measurements are taken at points 1 to 4 (yellow circles), including the number of plants and canes along two rows spanning 5 meters (orange lines).

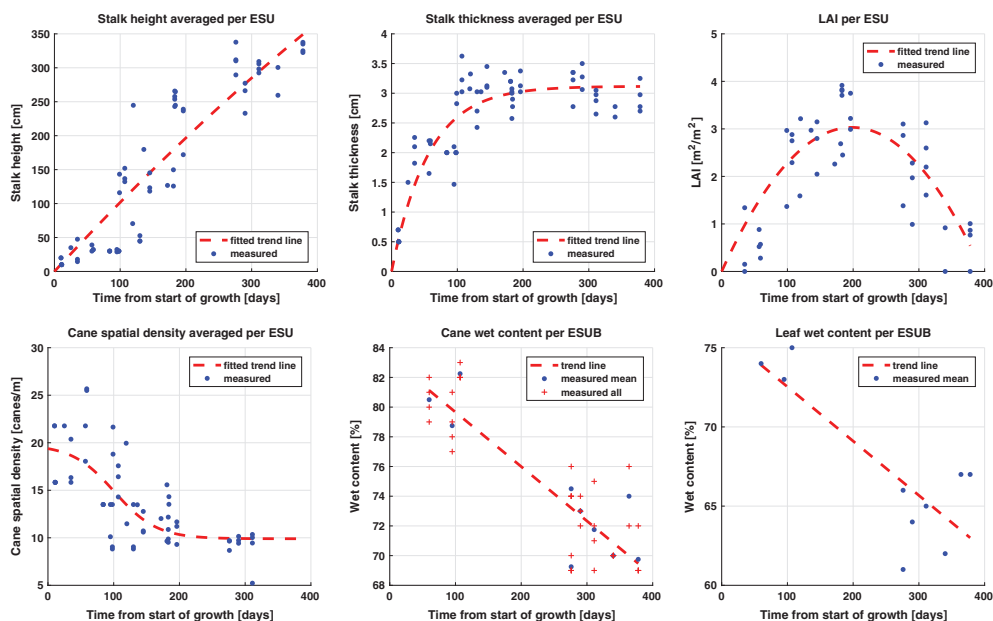


Figure 6. Temporal developments of measured stalk height and thickness, (actual) LAI and cane spatial density, averaged per ESU, and cane wet content and leaf wet content, averaged per ESUB. The fitted trend lines are indicative and only for illustrative purposes.

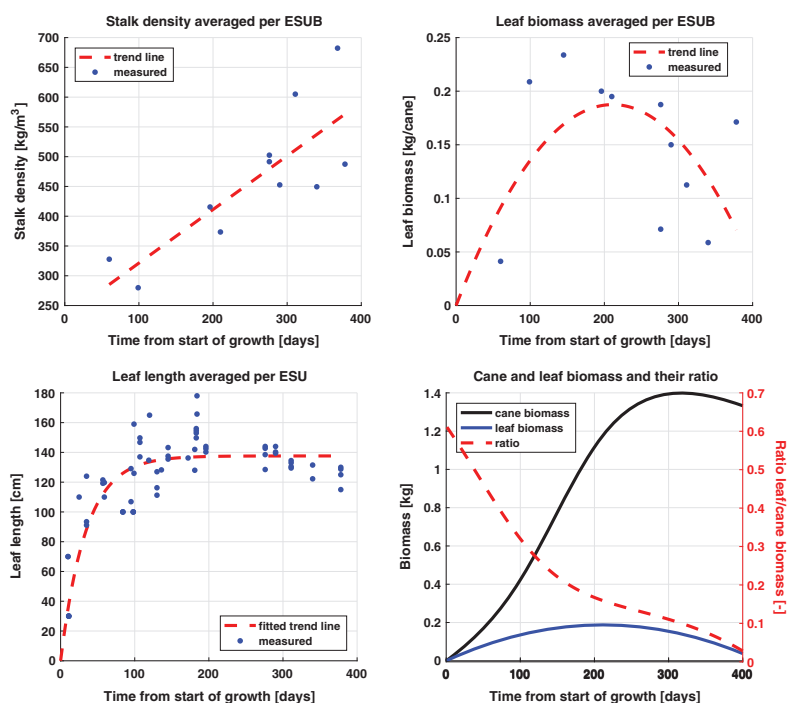


Figure 7. Temporal developments of stalk (mass) density (ρ_s), leaf biomass per cane (BM_L), averaged per ESUB, leaf length, averaged per ESU, and cane and leaf biomass and their ratio. The fitted trend lines for BM_L and ρ_s are used in the biomass equation.

The cane wet content and leaf wet biomass show a decrease over time, see Fig. 6, which was assumed to be linear over time (despite the time gap). The general observed tendency that cane and leaves become drier is supported by^{11,19}. The former shows similar water content values for stalks, though slightly higher and more constant during the first 100 days. These fitted linear functions are included only for indicative purposes.

The only measured parameters for which functions are fitted to be used in the biomass equation, see Equation 1, are stalk (mass) density and leaf biomass. This operation is indicated in Fig. 2 by “Linear model fitting (time)” with “BM.3” and “BM.4” as inputs. Stalk density shows a linear increasing trend, which can be attributed to the increase in sucrose content as reported by^{6,8,24}, which has higher mass density than water (approximately 1.6 g/cm³ and 1.0 g/cm³, respectively).

Similar to LAI, a second order polynomial fit was selected as the trend line for leaf biomass per cane, see Fig. 7. This is based on expected correlation with LAI and findings in the above mentioned LAI-related literature, in particular^{11,19}. Due to the low number of measurements, the precision of the function fit is non-optimal. To put the importance of leaf biomass into perspective, we added the approximated development of the ratio of leaf biomass over total cane biomass in Fig. 7. It shows that leaf biomass is dominant to the cane biomass during the first 50 days, after which it rapidly drops to less than 20% of the cane biomass after 200 days and decreases to 5% at the end of the cycle. Hence, the precision of the leaf biomass function has more impact at early state than at later states. The validity of the function is supported however by Fig. 8, showing an acceptable correlation of 0.89 between estimated cane biomass (i.e. as a result from the biomass equation) and measured cane biomass. Acceptable agreement is also found for individual estimations during early state.

Intensive measurements

In order to assess the reliability of the field measurements, three types of intensive measurements were carried out at three different dates. The details on how these precisions were measured, what their values are and how they are used for further analysis can be found in the sec:Technical-Validation section.

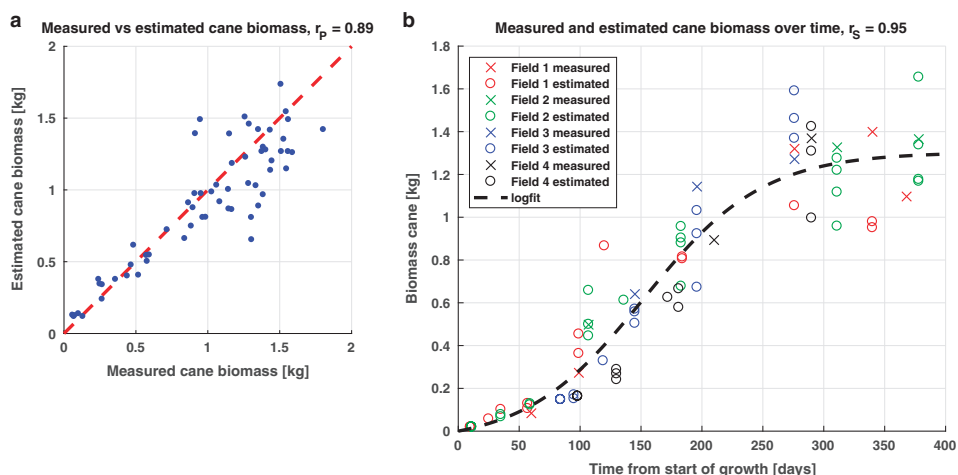


Figure 8. Performance of biomass estimation and performance of modeled growth curve. (a) relationship between estimated cane measured and measured cane biomass, accompanied by the (Pearson's) correlation (r_p); (b) general biomass growth curve showing the measured and estimated biomass over time from start of growth from all fields, accompanied by the fitted logistic function and the Spearman's correlation (r_s).

Precipitation

Over the course of the sugarcane growth cycles, seven weather stations acquired the daily cumulative precipitation, see Fig. 3. The rate of agreement between the weather stations of detecting precipitation was 0.96, which leads to our assumption that all fields are subject to the same condition of precipitation.

Cane biomass equation

In order to estimate the biomass at ESU we propose a biomass estimation equation, which is based on the sum of stalk biomass (the biomass of the main stem itself) and leaf biomass:

$$BM_C = BM_S + BM_L$$

$$= \left(\frac{(\pi D)^2}{4} \cdot H \cdot \rho_S(t) \cdot F \right) + BM_L(t) \quad (1)$$

BM_C and BM_S are respectively the cane biomass and stalk biomass, BM_L is the biomass leaf per cane, all in kilogram, D is the sugarcane stalk diameter in meter, H is the sugarcane stalk height in meter, ρ_S is the stalk (mass) density in kg/m³ as a function of time, and F . The latter, equal to 0.977, is assumed to be a scaling factor that is constant in time and compensates for the stalk as a tapered cylinder. This operation is indicated as "Compare" in Fig. 2, it is computed by dividing the ESUB-measured stalk biomass by the estimated stalk biomass from the biophysical parameters not corrected by the constant.

When applying Equation 1 with measured biophysical parameters at the ESUBs, the resulting estimated cane biomass is (linearly) related to the measured biomass with a (Pearson's) correlation of 0.89, see the left part of Fig. 8. The biomass growth curve on the right shows the fitted logistic function in time with the measured and estimated biomass. Since this curve is monotonic but not linear, we express the goodness of fit between the function and the (measured and estimated) biomass through the Spearman's correlation, equal to 0.95.

Cane biomass interpolators

Since the times of remote sensing image acquisitions rarely coincide with the times of ground measurements, temporal interpolation between field measurements is required. However, the temporal frequency of measurements for each individual field is not sufficient to do this directly for each individual field due to extrapolation and measurement precisions. As such, we combine the biomass measurements (from ESUBs) and biomass estimations (from ESUs) from all fields and create a biomass interpolator as a function of time from start of growth. A logistic function is selected in order to force the sigmoid curve, which is typical for the behavior of growth over time^{3-5,21,23} and was also applied to stalk thickness and cane spatial density:

$$f(t) = \frac{\Delta y}{1 + e^{-k(t-t_0)}} \quad (2)$$

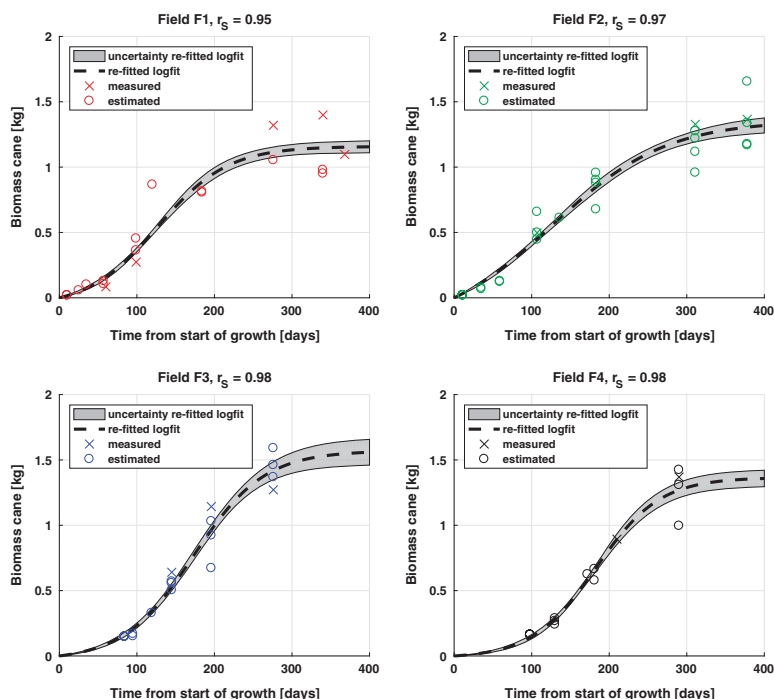


Figure 9. Measured and estimated cane biomass over time from start of growth and the re-fitted logistic fits, per ESU, accompanied by the Spearman's correlation (r_s) and the associated uncertainty bars (one standard deviation).

The coefficients consist of Δy , the limiting value defining the horizontal asymptote, k , the growth rate of the curve, and t_0 , the midpoint of the sigmoid. The model coefficients are estimated by the non-linear least squares method, without a-priori confidence bounds, using the measured and estimated biomass values, averaged per ESUB and ESU respectively, as input data and the function forced through zero origin. This is indicated in Fig. 2 by operation “Logistic model fitting (time)” based on the dataset containing all “ESU cane biomass estimates” and all “ESUB cane biomass estimates”, resulting in “4a”. The result is considered as the general biomass growth curve, see the right part of Fig. 8.

Subsequently, for each individual field new logistic functions are re-fitted, see Fig. 9. This re-fitting is constrained by using the estimated coefficients and their 95% confidence bounds from the general biomass growth function, respectively as initial values for the coefficients and as lower and upper bounds on the coefficients to be fitted. The latter bounds were expanded by 10% to allow more flexibility. This is indicated in Fig. 2 as operation “Logistic model re-fitting (time)” based on the dataset containing “ESU cane biomass estimates” per field and “ESUB cane biomass estimates” per field, resulting in “4b”. The resulting Spearman's correlations are all equal to or higher than 0.95, which is an improvement with respect to the general biomass growth curve of Fig. 7. The uncertainty bars (one s.d.) are computed with the methodology as depicted in the sec:Technical-Validation section.

Cane spatial density interpolators

For the cane spatial density, we apply the same logic as for the cane biomass; utilizing the boundary conditions of the general logistic fit, indicated as “5a. Cane spatial density interpolator (all)” in Fig. 2, illustrated as the trend line in Fig. 6, to re-fit logistic functions per ESU (i.e. not per field), indicated as “5b. Cane spatial density interpolator (per ESU)”.

TCH interpolators

Finally, in order to obtain the cane biomass per unit area we compute the tons cane per hectare (TCH):

$$TCH = BM_C \cdot C(t) \cdot 10/S \quad (3)$$

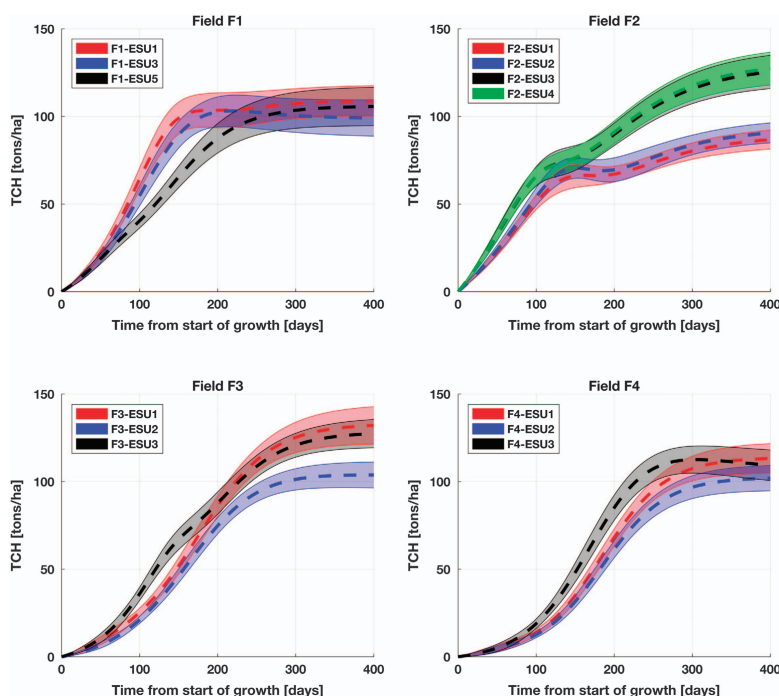


Figure 10. Estimated average TCH over time of start of growth per ESU, along with the associated uncertainty bars (one standard deviation).

BM_C is the cane biomass in kg, $C(t)$ is the measured cane spatial density in canes/m² as a function of time and S is the correction factor for spacing between the rows of canes equal to 1.2 (i.e. the average of 0.9 and 1.5). The multiplication by 10 converts kg/m² to tons/hectare. Multiplying the field-dependent BM_C with the ESU-dependent cane spatial density functions results in the TCH graph, see Fig. 10. Hence, we assume that the sugarcane plant growth within each field is constant and the cane spatial density varies per ESU, with the latter consequently dictating the intra-field differences from the ground reference data. These steps are indicated in Fig. 2 by operation “Product” based on “Cane biomass interpolators (per field)” and “Cane spatial density interpolators (per ESU)”. In the sec:Technical-Validation section the methodology for computing the one s.d. uncertainty bars, as depicted in the figure, is described.

Code availability

For pre-processing, visualizing and analyzing the presented data, MATLAB scripts are written, which are available upon request.

Data Records

The ground measurements, published for public use, can be found on the 4TU Centre for Research Data repository, see (Data Citation 1). The dataset contains readme files, which can be consulted for explanations on the location of the data records and supplementary material.

Technical Validation

For the technical validation of the measured and estimated parameters, three sets of uncertainty metrics are presented. The first set contains the absolute precisions of the measured parameters in the fields. The second set contains the precisions of the computed cane biomass parameter, based on error propagation laws applied on the biomass equation. The third set contains the precisions of the computed TCH parameter, which gives an enhanced view on the reliability of the TCH estimates and the significance of the intra-field differences.

Parameter	Instrument operation			Idealization		
	Early	Mid	Late	Early	Mid	Late
Stalk thickness [cm]	0.2	0.2	0.3	0.2	0.3	0.4
Stalk height [cm]	0.8	10.2	11.8	5.6	32.9	25.9
Cane spatial density [stalks/row/m]	0.8	0.7	0.5	6.3	4.0	2.1
LAI [m ² /m ²]	0.2	0.2	0.3	0.2	0.4	0.4
Stalk (mass) density [kg/m ³]	73	60	84	84	121	143
Leaf biomass per cane [g]	6.1	6.1	6.1	14.6	31.2	15.3

Table 2. Absolute instrument operation and idealization precisions, mainly based on the three intensive measurement campaigns in three growth stages.

Absolute precisions of field-acquired parameters

For this, we distinguish between three precision metrics:

- Instrument operation precision: dispersion of the measurement values when measuring the same parameter with the same instrument of the same plant. This precision is estimated through repeatedly measuring the same parameters of the same plant.
- Idealization precision: dispersion of the measurement values when measuring the same parameter with the same instrument at the same intended location but by chance another (neighboring) plant or cane was selected. The diameter of the area in which these plants were selected was five meters, approximately the accuracy of the GPS receiver. As a result, it was simulated that the measurements were taken at independent and uncorrelated visits at the same designated GPS location.
- Function fitting precision: dispersion of stalk (mass density), ρ_S and leaf biomass per cane, BM_L , fitted functions, approximated by the overall root-mean-square-error (RMSE) based on the measured data and estimations from these fitted functions.

The instrument operation precision is part of the idealization precision and hence the two cannot be treated as independent. The idealization precision can also be interpreted as a small scale local variability. For both types, the measurements were taken five times consecutively and conducted once at early state (until 150 days of growth), once at mid state (between 150 and 300 days of growth) and once at late state (more than 300 days of growth). It is assumed that the measurements at these single dates are representative for the corresponding growth states. In case of cane spatial density, the instrument operation precision is the dispersion of the number of canes when counting the same part of the row five times. For cane spatial density idealization precision, neighboring rows were counted. The counting was carried out over five meters, typically consisting of 60 to 100 canes, and subsequently converted the cane density per meter. The instrument operation precision for leaf biomass per cane is approximated by the average of the s.d.'s from repetitive weight measurements of cane biomass, and assumed to be constant over growth states. The idealization precision of leaf biomass per cane is approximated by the average of s.d.'s of the cane biomass measured per ESUB, per growth state, scaled by the ratio between leaf biomass and cane biomass. Hence, for both cases, the assumption is made that variability in cane biomass is representative for variability in leaf biomass. For stalk (mass) density, the instrument operation precision and idealization precision are computed through error propagation of the biomass equation (Equation 1), resulting in:

$$\sigma_{\rho_S} = \rho_S \sqrt{(\sigma_{BM_S}/BM_S)^2 + (2\sigma_D/D)^2 + (\sigma_H/H)^2} \tag{4}$$

Here, σ_{ρ_S} , σ_{BM_S} , σ_D and σ_H are the precisions of stalk (mass) density (ρ_S), stalk biomass (BM_S), cane diameter (D) and cane height (H), respectively. The ratio σ_{BM_S}/BM_S is approximated by the equivalent of cane biomass, because at the ESUB locations only the variation in cane biomass was measured (see the subsec:Biomass section). The precisions of the parameters used in the biomass equation and LAI are shown in Table 2, expressed as one s.d. and valid for single observations. The idealization precision for cane spatial density at late stage is approximated by following the trend of precisions during the previous stages.

The function fitting precisions of stalk (mass) density and leaf biomass are not completely independent from the idealization precisions, because the functions are fitted through the measured parameters that are affected by the idealization precision (and consequently also by the instrument operation precision). The function fitting precisions of these parameters are shown in Table 3.

Parameter	Fitting precision
Stalk (mass) density [kg/m ³]	65
Leaf biomass per cane [g]	58

Table 3. Absolute fitting precisions, approximated by the RMSE of the fitted functions, which are based on the measurements in the ESUBs. It is assumed these are constant over the growth states.

Relative precisions of cane biomass and cane spatial density

Since the precisions of cane biomass are not directly measured, we apply error propagation laws on Equation 1, resulting in:

$$\sigma_{BM_C} = \sqrt{\left(\sqrt{\vartheta_{BM_S} + \chi_{BM_S}} \cdot BM_S\right)^2 + (\sigma_{BM_L})^2} \quad (5)$$

$$\begin{aligned} \vartheta_{BM_S} &= \left(\frac{2\sigma_D}{D}\right)^2 + \left(\frac{\sigma_H}{H}\right)^2 + \left(\frac{\sigma_{\rho_S}}{\rho_S}\right)^2 \\ \chi_{BM_S} &= \left(\frac{\sigma_D}{D}\right)^2 + \left(\frac{2\sigma_D\sigma_H}{DH}\right)^2 + \left(\frac{2\sigma_D\sigma_{\rho_S}}{D\rho_S}\right)^2 + \left(\frac{\sigma_H\sigma_{\rho_S}}{H\rho_S}\right)^2 \end{aligned} \quad (6)$$

Here, σ_{BM_C} , σ_D , σ_H , σ_{ρ_S} and σ_{BM_L} are the precisions associated to cane biomass (BM_C), cane diameter (D), cane height (H), stalk (mass) density (ρ_S) and biomass leaf per cane (BM_L), respectively, whereby Equation 4 is used for σ_{ρ_S}/ρ_S and χ_{BM_S} is the second order term for biomass stalk per cane (BM_S). For this error propagation, given the limited measurement data availability, it is assumed that the errors of the variables are uncorrelated and normal distributed. These equations have been applied with the instrument operation and idealization precisions from Table 2 for computing the corresponding precisions of cane biomass, σ_{BM_C} . These are expressed as percentages relative to the average BM_C estimates and shown in Table 4 for the corresponding stages. It shows that the instrument operation precision, as may be expected, is always smaller than the idealization precision.

Also incorporated in Table 4 is the function fitting precision that is a combination of the function fitting precisions of cane (mass) density and leaf biomass, both individually approximated by their function's RMSE (see Table 3). Since the absolute function fitting precision is taken as constant over time, the relative precision (expressed as percentages) decreases over increasing cane biomass. From the idealization precisions of D and H and the function fitting precisions of ρ_S and BM_L , the total precision at location is computed. The idealization precisions of ρ_S and BM_L are not taken into account for the total precision at location, because these are not independent from the corresponding function fitting precisions. Consequently, the total precision at location is not a simple summation of these two precisions and results in smaller values than the idealization precisions for mid and late state. The difference between idealization precision and total precision can hence be interpreted as the effect of applying function fitting of ρ_S and BM_L .

In addition, from the measurements taken at each corner point of the ESUs and ESUBs, the ESU precision per stage is computed and shown in Table 3. These precisions are based on the averaged s.d. and the function fitting precisions, per stage. The table shows that the estimated variability within five meters (i.e. total precision at location) is similar yet slightly smaller than the variability within the ESU dimension of 20 by 20 meters (i.e. total precision within ESU).

The total precisions at location of Table 3 are used for perturbing the logistic fits of cane biomass of each field, as shown in Fig. 9. Explained in more detail, the uncertainty bars in Fig. 9, represented as one s.d. on either side of the logistic fit, were computed by taking the s.d. of hundred individual re-fitted logistic functions fitted through the measurements that were perturbed (normal) randomly based on the total precisions at location as input s.d.'s.

The same logic of precision estimation is applied to cane spatial density precisions (σ_C), see Table 5, again expressed as percentages relative to average cane spatial density. Fitting precisions are absent since no estimated parameters are involved. Although the absolute instrument operation precisions (i.e. counting error) decrease over the stages, see Table 2, the corresponding relative precision remains constant because of a similar decrease in cane spatial density. The relative idealization precisions decrease significantly over the stages, demonstrating the decrease in local variability. The ESU-related precisions are smaller than the location-related precisions, which could be explained by the apparent bias of the observer towards selecting and counting more dense sugarcane rows rather than randomly selecting sugarcane rows. Nevertheless, to remain consistent with the cane biomass case, for perturbing the logistic fits of cane spatial density again the total precisions at location were used. This is explained in more detail in the next section.

Relative precisions of TCH

As briefly explained before, the estimated average tons cane per hectare (TCH) profiles of Fig. 10 are based on the multiplication of the logistic fits of cane biomass of each field with the logistic fits of cane

Cane biomass (BM_C)	Early state	Mid state	Late state
Idealization precision	21%	30%	32%
(Instrument operation precision)	(19%)	(16%)	(20%)
Function fitting precision	29%	14%	12%
Total precision at location	32%	23%	23%
Total precision within ESU	36%	24%	27%

Table 4. Relative precisions with respect to cane biomass (BM_C).

Cane spatial density (C)	Early state	Mid state	Late state
Idealization precision	45%	29%	19%
(Instrument operation precision)	(5%)	(5%)	(5%)
Total precision at location	45%	29%	19%
Total precision within ESU	23%	16%	10%

Table 5. Relative precisions with respect to cane spatial density (C).

spatial density of each ESU. For estimating the uncertainty of these profiles for each ESU, the field-dependent perturbed cane biomass profile is multiplied with the ESU-dependent perturbed cane spatial density profile. Both the perturbation of cane biomass measurements as well as the perturbation of cane spatial density measurements were carried out one hundred times, which means that the resulting uncertainty is based on ten thousand combinations of perturbed profiles. The uncertainty bars in Fig. 10 consequently represent the s.d. of the perturbed profiles on each side of the mean profile. Table 6 shows these precisions as percentages of the TCH per state averaged over all ESUs per field.

These relative TCH precisions show a significant reduction over time due to the increased TCH values over time and due to fitting of the logistic function for BM_C and C based on the perturbed measurements. This yields a significant difference compared to what would be expected when directly propagating the cane biomass precisions of Table 4 and the cane spatial precisions of Table 5. To quantify this difference (i.e. between the precisions associated to TCH based on the fitted functions and the precisions when TCH would be directly computed through the measurements), we apply error propagation based on Equation 1 in combination with Equation 3, resulting in:

$$\sigma_{TCH} = \left[\left(\sqrt{\vartheta_{TSH} + \chi_{TSH}} \cdot |TSH| \right)^2 + \left(\sqrt{\vartheta_{TLH} + \chi_{TLH}} \cdot |TLH| \right)^2 \right]^{1/2} \quad (7)$$

$$\begin{aligned} \vartheta_{TSH} &= \left(\frac{2\sigma_D}{D} \right)^2 + \left(\frac{\sigma_H}{H} \right)^2 + \left(\frac{\sigma_{PS}}{\rho_S} \right)^2 + \left(\frac{\sigma_C}{C} \right)^2 \\ \vartheta_{TLH} &= \left(\frac{\sigma_{BM_L}}{BM_L} \right)^2 + \left(\frac{\sigma_C}{C} \right)^2 \\ \chi_{TSH} &= \left(\frac{(\sigma_D)^2}{D^2} \right)^2 + \left(\frac{2\sigma_D\sigma_H}{DH} \right)^2 + \left(\frac{2\sigma_D\sigma_{PS}}{D\rho_S} \right)^2 + \left(\frac{2\sigma_D\sigma_C}{DC} \right)^2 + \left(\frac{\sigma_H\sigma_{PS}}{H\rho_S} \right)^2 + \left(\frac{\sigma_H\sigma_C}{HC} \right)^2 + \left(\frac{\sigma_{PS}\sigma_C}{\rho_S C} \right)^2 \\ \chi_{TLH} &= \left(\frac{\sigma_{BM_L}\sigma_C}{BM_L C} \right)^2 \end{aligned} \quad (8)$$

In addition to the explained variables of Equation 5 and Equation 6, σ_{TCH} and σ_C are the precisions associated to tons cane per hectare (TCH) and cane spatial density (C), respectively, and χ_{TSH} and χ_{TLH} are the second order terms for tons stalk per hectare (TSH) and tons leaf per hectare (TLH). Similarly to Equation 5, the same error propagation assumptions apply and the results of Table 7 are expressed in percentages of the precisions relative to the average TCH estimates for the corresponding state. The precisions associated with the TSH and TLH are grouped separately (labeled group 1 and group 2, respectively) in order to show their magnitude relative to the TCH estimate at that state. The function fitting precisions again are approximated by the overall RMSE and in this case are based on the measured data and the fitted functions of cane spatial density, ρ_S , for group 1 and leaf biomass, BM_L , for group 2. Similarly to before, for these modeled parameters, only the fitting precisions are taken into account and their corresponding idealization precisions are disregarded. Also, the ESU-related precisions are again based on the averaged s.d. and the function fitting precisions, per stage. For group 1 +2, the idealization precision (averaged s.d. without ρ_S and BM_L function fitting precisions) is included as well. This is used for comparison with the variogram at the end of this section.

Field	Early state	Mid state	Late state
F1	13%	9%	9%
F2	9%	8%	7%
F3	19%	7%	6%
F4	31%	7%	7%
Average	18%	8%	7%

Table 6. Relative precisions with respect to TCH as a result from perturbation of the measurements in the logistic fits. The results per ESUs are averaged per field.

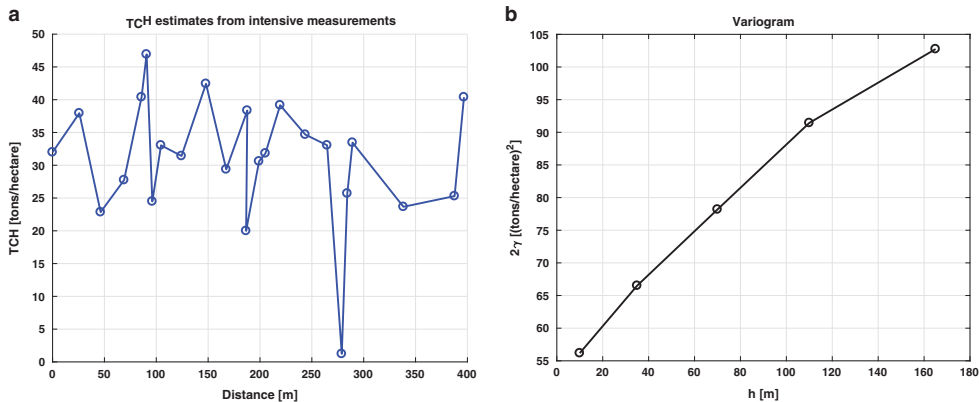


Figure 11. Biomass estimates from intensive measurements. (a) biomass estimates in TCH versus distance from intensive measurements at field F2; (b) corresponding variogram, where the variance (2γ) in (TCH)² is function of the lag distance h .

The overall interpretation is that the leaf precision component, group 2, contributes less to the total cane precision, group 1+2, than the stalk precision component, group 1, at mid and late stage, which is logical given the overall weight dominance of stalk biomass to the cane biomass. The early stage leaf precision component is indeed dominated by the function fitting precision, due to the earlier mentioned relatively large fitting error and large weight portion of the plant during this state. This group also shows that in the early state the precisions are always larger than the mid and late states, which is due to the fact that the relative contribution of leaf biomass to the cane biomass decreases over time. Also, similar to before and as expected, the instrument operation precisions are always lower than the idealization precisions and the main contributor to the latter is the local variability in plant parameters. In case of group 2, the effect of local variability is apparently almost entirely dominating the idealization precision. The fact that the total precisions at location of group 2 is equal to the propagated sum of the idealization precisions and function fitting precisions shows that the contribution of leaf biomass to the idealization precision is small compared to the cane spatial density. This confirms the finding in the previous section that the contribution of cane spatial density to the total precisions is dominant.

A more detailed assessment of the results reveals that the instrument operation precisions of group 1 shows lower values for mid state than for the other stages, which is considered to be in the expected margin of error. In other words, it could be assumed these precisions should be viewed as constant over the stages. The total precision within ESU shows that these are comparable though slightly smaller (again assumed to be within the expected margin of error) than the total precisions at location, which means that the variability within an ESU is similar to the variability at the location. For all second order terms of the previous equations it holds that they only marginally contribute, up to 2% point, to the corresponding precisions, indicating that all higher order terms can be safely ignored.

The total precisions at location of Table 7 (based on error propagation) are generally larger than the precisions of Table 6 (based on profile perturbations) and there is a clear difference in temporal development of the precisions. For example, when TCH would be directly estimated

Group 1: stalk component (TSH)	Early state	Mid state	Late state
Idealization precision (Instrument operation precision)	31% (19%)	38% (17%)	36% (21%)
Function fitting precision	12%	12%	11%
Total precision at location	30%	32%	29%
Total precision within ESU	27%	27%	28%
Group 2: leaf component (TLH)	Early state	Mid state	Late state
Idealization precision (Instrument operation precision)	22% (4%)	6% (1%)	2% (1%)
Function fitting precision	26%	6%	5%
Total precision at location	34%	8%	5%
Total precision within ESU	28%	6%	5%
Group 1+2: total cane (TCH)	Early state	Mid state	Late state
Idealization precision (Instrument operation precision)	38% (19%)	38% (14%)	36% (12%)
Function fitting precision	29%	14%	12%
Total precision at location	45%	33%	29%
Idealization precision within ESU	26%	24%	25%
Total precision within ESU	39%	27%	28%

Table 7. Relative precisions with respect to tons cane per hectare (TCH) as a result from error propagation from repeated measurements. The results are given for the stalk and leaf components individually and for the total cane.

from the measurements through Equation 3 and Equation 1, the precision is up to 45% of the TCH value (Table 7). This is significantly less precise than when the logistic fitting method of cane biomass and cane spatial density is used, where the precision is on average 18% in early state to 7% in late state of the TCH value (Table 6). This significantly lower precision is a result from the assumption that the boundary conditions of the general cane biomass growth curve can be used to constrain the field-dependent cane biomass growth curves in combination with the assumption both these curves and the ESU-dependent cane spatial density curves can be modeled by logistic functions. The advantage of this approach is that at any point in time (e.g. at the time of remote sensing images) in the growth period the TCH can be acquired, whereas with the direct measurements, interpolation and extrapolation needs to be carried out which induces another uncertainty. Hence, the modeled TCH curves per ESU are recommended for further use, e.g. for compare with remote sensing patterns in time and space.

The final set of measurements for describing the TCH variability was taken on a single date, 19 December 2014 (66 days after start of growth), in field F2. These were taken along a profile, starting from point F2-IM1 in Fig. 4 until 400 meters north from that point, every 5, 10 and 20 meters along that line, resulting in 24 local measurement locations. At every location, leaf length, stalk height, stalk thickness, cane spatial density, plant spatial density, LAI and soil moisture were measured. At each location, biomass values were computed from the biomass re-fitted logfit functions per field of Fig. 9 multiplied by the measured cane spatial densities. The results are shown as a profile along the measurement line and as a variogram, see Fig. 11. The variance (2γ as function of lag distance h) in each bin was estimated from 36 to 61 sample pairs. The associated s.d. as percentage of the averaged estimated biomass (approximately 30 tons/ha) is 24% within the spatial dimension of an ESU ($h = 20$ meters) and ranges to nearly 33% at the maximum calculated distance ($h = 160$ meters). The former value is lower than the early state idealization precision of group 1+2 but similar to the idealization precision within ESU of group 1+2 in Table 7.

Usage Notes

In (Data Citation 1), the readme files explains the location of the files and folders. All raw measurements records can be found in one Excel sheet, which is subsequently used in MATLAB for further processing and analysis.

References

1. Pammenter, N. W. & Allison, J. C. S. Effects of treatments potentially influencing the supply of assimilate on its partitioning in sugarcane. *J. Exp. Bot.* **53**, 123–129 (2002).

2. Vieira, M. A. *et al.* Object Based Image Analysis and Data Mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. *Remote Sens. Environ.* **123**, 553–562 (2012).

3. Kölln, O. T. *et al.* Fertigated Sugarcane Yield and Carbon Isotope Discrimination ($\Delta^{13}C$) Related to Nitrogen Nutrition. *Sugar Tech* **18**, 391–400 (2016).

4. Mariano, E. *et al.* Biomass and nutrient content by sugarcane as affected by fertilizer nitrogen sources. *Crop Sci* **56**, 1234–1244 (2016).

5. Franco, H. C. J. *et al.* Nitrogen in sugarcane derived from fertilizer under Brazilian field conditions. *Field Crops Res* **121**, 29–41 (2011).
6. Nawi, N. M., Chen, G. & Jensen, T. In-field measurement and sampling technologies for monitoring quality in the sugarcane industry: a review. *Precis. Agric.* **15**, 684–703 (2014).
7. Rudorff, B. & Batista, G. Yield estimation of sugarcane based on agrometeorological-spectral models. *Remote Sens. Environ.* **192**, 183–192 (1990).
8. Gutiérrez-Miceli, F. A. *et al.* Effects of partial defoliation on sucrose accumulation, enzyme activity and agronomic parameters in sugar cane (*Saccharum* spp.). *J. Agron. Crop Sci.* **190**, 256–261 (2004).
9. Sandhu, H. S. *et al.* Relationships among leaf area index, visual growth rating, and sugarcane yield. *J. Am. Soc. Sugar Cane Technol* **32**, 1–14 (2012).
10. Johnson, R., Viator, R., Veremis, J., Richard, E. & Zimba, P. Discrimination of sugarcane varieties with pigment profiles and high resolution, hyperspectral leaf reflectance data. *J. Assoc. Sugar Cane Technol* **28**, 63–75 (2008).
11. Oliveira, R. A. D. *et al.* Área foliar em três cultivares de cana-de-açúcar e sua correlação com a produção de biomassa. *Pesq. Agropec. Trop* **37**, 71–76 (2007).
12. Gomathi, R. *et al.* Physiological Studies on Ratoonability of Sugarcane Varieties under Tropical Indian Condition. *Am. J. Plant Sci* **4**, 274–281 (2013).
13. Simões, M. D. S., Rocha, J. V. & Lamparelli, R. A. C. Growth indices and productivity in sugarcane. *Sci. Agric.* **62**, 23–30 (2005).
14. Gilbert, R. A. *et al.* Sugarcane Response to Mill Mud, Fertilizer, and Soybean Nutrient Sources on a Sandy Soil. *Agron. J.* **100**, 845 (2008).
15. Wiedenfeld, R. Water stress during different sugarcane growth periods on yield and response to N fertilization. *Agr. Water Manage.* **43**, 173–182 (2000).
16. Thorburn, P. J. *et al.* Prioritizing Crop Management to Increase Nitrogen Use Efficiency in Australian Sugarcane Crops. *Front. Plant Sci.* **8**, 1–16 (2017).
17. Lin, H., Chen, J., Pei, Z., Zhang, S. & Hu, X. Monitoring Sugarcane Growth Using ENVISAT ASAR Data. *IEEE Trans. Geosci. Remote Sens* **47**, 2572–2580 (2009).
18. Simões, M. D. S., Rocha, J. V. & Lamparelli, R. A. C. Orbital spectral variables, growth analysis and sugarcane yield. *Sci. Agric.* **66**, 451–461 (2009).
19. Baghdadi, N., Cresson, R., Todoroff, P. & Moinet, S. Multitemporal observations of sugarcane by TerraSAR-X images. *Sensors* **10**, 8899–8919 (2010).
20. Baghdadi, N., Boyer, N., Todoroff, P., El Hajj, M. & Bégué, A. Potential of SAR sensors TerraSAR-X, ASAR/ENVISAT and PALSAR/ALOS for monitoring sugarcane crops on Reunion Island. *Remote Sens. Environ.* **113**, 1724–1738 (2009).
21. Simões, M. D. S., Rocha, J. V. & Lamparelli, R. A. C. Spectral variables, growth analysis and yield of sugarcane. *Sci. Agric.* **62**, 199–207 (2005).
22. Picoli, M. C. A., Lamparelli, R. A. C., Sano, E. E. & Rocha, J. V. Imagens multipolarizadas do sensor Palsar/Alas na discriminação das fases fenológicas da cana-de-açúcar. *Pesq. Agropec. Bras* **47**, 1307–1316 (2012).
23. Simões, M. D. S., Rocha, J. V. & Lamparelli, R. A. C. Análise do comportamento espectral de uma área comercial de cana-de-açúcar em duas safras por meio de radiometria de campo. *In SBSR 2367–2374* (Belo Horizonte: Brazil, 2003).
24. Abdel-Rahman, E. M. & Ahmed, F. B. The application of remote sensing techniques to sugarcane (*Saccharum* spp. hybrid) production: a review of the literature. *Int. J. Remote Sens.* **29**, 3753–3767 (2008).
25. LAI-2000 Plant Canopy Analyzer-Operating Manual <https://licor.app.boxenterprise.net/s/q6hrj6s79psn7o8z2b2s> (LI-COR Inc., 1992).
26. Baret, F. *et al.* VALERI: a network of sites and a methodology for the validation of medium spatial resolution land satellite products <http://w3.avignon.inra.fr/valeri/documents/VALERI-RSESubmitted.pdf> (NRA (2003).
27. *Sugar Cane* http://www.naandanjain.com/uploads/catalogfiles/sugar-cane-2/Suger_cane_booklet_100613F.pdf (NaanDanJain Irrigation Ltd., 2013).

Data Citations

1. Molijn, R. A. *Delft University of Technology* <https://doi.org/10.4121/uuid:37112e18-f794-4d66-a8cd-7f1e92af09fc> (2015).

Acknowledgements

The authors would like to acknowledge the European Space Agency (ESA) for providing the Radarsat-2 data under the framework of project CIP.16849, the Google Earth Engine team for providing access to the Sentinel-1 and Landsat-8 data, the Japan Aerospace Exploration Agency (JAXA) for providing the ALOS-2 data under the Research Announcement-6, number 149250, and the Brazilian Space Agency (INPE) for providing the Canasat maps of the São Paulo region. In addition, we like to thank the colleagues at FEAGRI, Unicamp, for their assistance with taking the ground measurements, in particular Diego della Justina, Carlos Wachholz de Souza, Walter Rossi Cervi, Rubens Lamparelli and Ali Mousivand. Also, we like to thank Ali Mousivand for his expert view on taking the LAI measurements and the correction to actual LAI in particular. Finally, we acknowledge and are grateful for the access to and safety in the field provided by Jefferson Rodrigo Batista de Mello, Eduardo Caetano Ceará, Pedro Lian Barbieri and Alex Thiele Paulino. The work was carried out in BE-Basic project FES0905 and partly carried out within the framework of the joint BE-Basic FAPESP project 2013/50942-2.

Author Contributions

R.A.M. designed and carried out the ground measurement campaign, developed the methodologies, applied statistical techniques, analyzed the data, interpreted and validated the data and results, and wrote the manuscript. L.I. assisted in designing and carrying out the ground measurement campaign, framing the statistical analysis, interpreting the data and results, reformulated parts of the manuscript and reviewed the manuscript. J.V.R. established the relation and an agreement with the owner of the sugarcane fields (together with Rubens Lamparelli), assisted in validating and approving the results and reviewed the manuscript. R.F.H. initiated the project and funding, assisted in framing the statistical analysis and reviewed the manuscript.

Additional Information

Competing interests: The authors declare no competing interests.

How to cite this article: Molijn, R. A. *et al.* Ground reference data for sugarcane biomass estimation in São Paulo state, Brazil. *Sci. Data* 5:180150 doi: 10.1038/sdata.2018.150 (2018).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files made available in this article.

© The Author(s) 2018

SCIENTIFIC DATA

OPEN

Author Correction: Ground reference data for sugarcane biomass estimation in São Paulo state, Brazil

Ramses A. Molijn, Lorenzo Iannini, Jansle Vieira Rocha & Ramon F. Hanssen

Correction to: *Scientific Data* <https://doi.org/10.1038/sdata.2018.150>, published online 07 August 2018

Following publication, it was noticed that the horizontal brackets labelling the two groups of precisions present in Equation 7 are incorrectly rendered in the PDF version of this Data Descriptor. The correct Equation 7 is as follows:

$$\sigma_{TCH} = \left[\left(\frac{\sqrt{\vartheta_{TSH} + \chi_{TSH}} \cdot |TSH|}{1} \right)^2 + \left(\frac{\sqrt{\vartheta_{TLH} + \chi_{TLH}} \cdot |TLH|}{2} \right)^2 \right]^{1/2}$$

In addition, in the Biomass subsection of the Methods section in both the HTML and PDF versions, the term “ESUs” is incorrectly rendered as “ESUâ€™s” and the term ESUBs is incorrectly rendered as “ESUBâ€™s”.

Finally, throughout the manuscript, references to sections and subsections include the prefixes “sec.” and “sub-sec.”, respectively. These prefixes and any hyphen between the reference words that follow the prefixes can be ignored.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

3

SPACE-BASED SAR AND OPTICAL REMOTE SENSING FOR SUGARCANE PRODUCTIVITY MONITORING

*O que me tranquiliza
é que tudo o que existe,
existe com uma precisão absoluta.
O que for do tamanho de uma cabeça de alfinete
não transborda nem uma fração de milímetro
além do tamanho de uma cabeça de alfinete.
Tudo o que existe é de uma grande exatidão.
Pena é que a maior parte do que existe
com essa exatidão
nos é tecnicamente invisível.
O bom é que a verdade chega a nós
como um sentido secreto das coisas.
Nós terminamos adivinhando, confusos,
a perfeição.*

Clarice Lispector

THIS chapter comprises one journal article and two conference proceedings. The article has been published in Remote Sensing, Earth Observations and Crop Models for Sustainable Agricultural Management, in 2019 [1]. The proceedings have been published in Proceedings of IGARSS (IEEE International Geoscience and Remote Sensing Symposium), in 2016 [2], and in Proceedings of IGARSS (IEEE International Geoscience and Remote Sensing Symposium), in 2018 [3]. The journal article demonstrates how C-band and L-band SAR and optical observations can be used for the monitoring of sugarcane productivity based on the ground measurements that were described in the previous

chapter. It includes a sensitivity analysis of the SAR and optical signals to sugarcane biomass development, precipitation events and sensor configurations. It also presents the conditions during which these signals can be used effectively for growth monitoring in terms of growth stages and precipitation conditions. Finally, the temporal consistencies between the remote sensing images per sensor and between sensors are demonstrated.

The first proceeding provides insights into the effectiveness of using satellite-based SAR and optical imagery over sugarcane fields for the detection of persistent patterns in time and space. The trends of growth as measured by the two sensors are compared and the challenges for sugarcane growth monitoring are demonstrated. The ground reference measurements from the previous chapter were also used here to limit the time span of observations for the analysis. The second proceeding demonstrates how spatial gaps in a sugarcane field affect C-band and L-band SAR signals. The variability of these effects with varying spatial averaging windows is addressed together with the impacts of precipitation conditions and vegetation heights on the signal's behavior. For this study, a ground reference dataset was used that was collected during another measurement campaign than described in the previous chapter.

REFERENCES

- [1] R. A. Molijn, L. Iannini, J. V. Rocha, and R. F. Hanssen, *Sugarcane productivity mapping through C-band and L-band SAR and optical satellite imagery*, *Remote Sensing* **11** (2019), 10.3390/rs11091109.
- [2] R. A. Molijn, L. Iannini, R. F. Hanssen, and J. Vieira Rocha, *Sugarcane growth monitoring through spatial cluster and temporal trend analysis of radar and optical remote sensing images*, in *International Geoscience and Remote Sensing Symposium (IGARSS)*, Vol. 2016-November (2016) pp. 7141–7144.
- [3] R. A. Molijn, L. Iannini, C. W. De Souza, D. D. Justina, J. V. Rocha, and R. F. Hanssen, *The effects of sugarcane productivity anomalies on L-band and C-band SAR signals*, in *International Geoscience and Remote Sensing Symposium (IGARSS)*, Vol. 2018-July (IEEE, 2018) pp. 5371–5374.



Article

Sugarcane Productivity Mapping through C-Band and L-Band SAR and Optical Satellite Imagery

Ramses A. Molijn ^{1,*} , Lorenzo Iannini ¹, Jansle Vieira Rocha ² and Ramon F. Hanssen ¹

¹ Geoscience and Remote Sensing, Delft University of Technology, 2628 CN Delft, The Netherlands; l.iannini@tudelft.nl (L.I.); r.f.hanssen@tudelft.nl (R.F.H.)

² Faculdade de Engenharia Agrícola (FEAGRI), Unicamp, Campinas 13083-875, Brazil; jansle@unicamp.br

* Correspondence: r.a.molijn@tudelft.nl

Received: 9 March 2019; Accepted: 1 May 2019; Published: 9 May 2019



Abstract: Space-based remote sensing imagery can provide a valuable and cost-effective set of observations for mapping crop-productivity differences. The effectiveness of such signals is dependent on several conditions that are related to crop and sensor characteristics. In this paper, we present the dynamic behavior of signals from five Synthetic Aperture Radar (SAR) sensors and optical sensors with growing sugarcane, focusing on saturation effects and the influence of precipitation events. In addition, we analyzed the level of agreement within and between these spaceborne datasets over space and time. As a result, we produced a list of conditions during which the acquisition of satellite imagery is most effective for sugarcane productivity monitoring. For this, we analyzed remote sensing data from two C-band SAR (Sentinel-1 and Radarsat-2), one L-band SAR (ALOS-2), and two optical sensors (Landsat-8 and WorldView-2), in conjunction with detailed ground-reference data acquired over several sugarcane fields in the state of São Paulo, Brazil. We conclude that satellite imagery from L-band SAR and optical sensors is preferred for monitoring sugarcane biomass growth in time and space. Additionally, C-band SAR imagery offers the potential for mapping spatial variations during specific time windows and may be further exploited for its precipitation sensitivity.

Keywords: sugarcane growth monitoring; SAR and optical remote sensing; precipitation effects; saturation effects

1. Introduction

Sugarcane is the number one globally cultivated crop in terms of production quantity, more than the second and third crops, maize and rice, combined. Brazil is the largest sugarcane producer, amounting to almost 40% of total global production [1], while São Paulo state hosts more than 60% of Brazil's sugarcane acreage [2,3]. Over the last 15 years, this acreage more than doubled [4]. The main products are sugar and bioethanol; the latter allows Brazil to reduce the national automotive's gasoline consumption by more than half [2]. This illustrates the importance of sugarcane production and research in São Paulo, which hosts the area of study.

Sugarcane is a semiperennial crop; after each growth cycle, typically lasting for 12 to 18 months for Brazilian plantations [5], new ratoons emerge from the same root system. Since the yield decreases over time, the plants and the root system are generally removed after five to seven years, and new shoots are planted. Indicatively, the growth cycle follows phenological phases as depicted in Figure 1. During the last phase, the sucrose content in the stem accumulates and senescence of the leaves occurs [5–7].

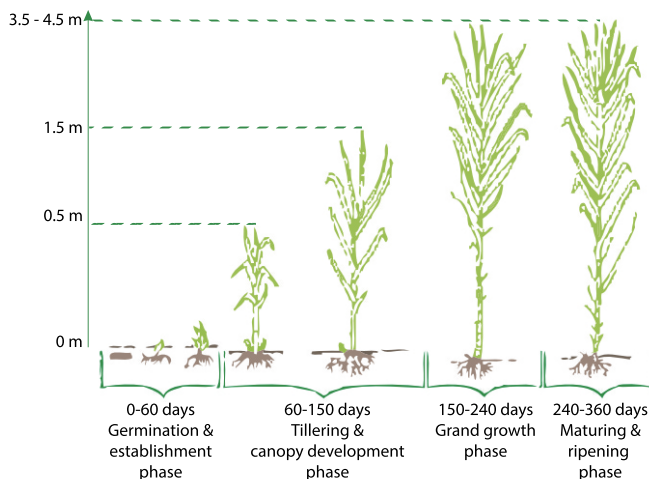


Figure 1. Phenological phases of sugarcane and their corresponding time frames, as well as indicative heights and plant geometries. Adapted and modified from NaanDanJain [8].

The sensitivity of remote observations to crop conditions was largely investigated for crop-monitoring uses [9–11] as well as to improve the understanding of the plant–microwave interaction [6,12]. Review works by McNairn and Brisco, Steele-Dunne et al. [13,14], and a study by Moran et al. [12] give a general introduction on the factors that affect Synthetic Aperture Radar (SAR) signals over the course of crop growth. The signal backscatter is mainly influenced by three mechanisms: the interaction of the microwaves directly with the vegetation, directly with the soil, and the interaction between vegetation and the underlying soil. Regarding the vegetation contribution, the backscatter coefficient is governed by its dielectric properties and its geometrical aspects. Its dielectric properties are mainly dependent on the vegetation water content and water droplets on the plant after precipitation events. The contribution from the soil is similarly governed by its geometry (roughness and slope) and by its moisture. The backscatter coefficient is further dependent on the configuration of the sensor, including viewing geometry, wavelength, and polarization. The soil backscatter contribution for C-band SAR is small from the leaf-development stage onward [12], while it is more significant for L-band SAR signals with standing vegetation [14]. In addition, the impact of the incidence angle on backscatter was found to be minimal compared to the effects from soil and vegetation conditions for C-band SAR [12], but more apparent for L-band SAR [6].

The monitoring capabilities of sugarcane growth with optical signals was described by several studies mainly focusing on biophysical-parameter extraction [3,9,15–17] and yield estimation [10,16,18,19]. Varying rates of success were reported, and the main limiting factors were attributed to the scarce availability of optical images and the complex relation between observed signals and the estimated parameter or feature. By integrating agrometeorological data with vegetation indices, better results were achieved, though still not operationally applicable due to challenges in the agrometeorological modelling [10]. Another solution was found in the use of optical data acquired by unmanned aerial vehicles at the time of interest [9]. This study also showed that spatial patterns in height estimations from optical signals were in agreement with observed heterogeneities in the ground-reference data. Furthermore, a short study on the value of C-band SAR (Radarsat-2 and Sentinel-1) and optical imagery (Landsat-8), acquired over sugarcane fields [20], found that SAR imagery contains fewer spatial features that are statistically significant and persistent in time than optical imagery.

Far fewer studies were published on the capabilities of SAR signals for the mapping and monitoring of sugarcane than were published for other crops like wheat, barley, and maize. Commonly, the works on sugarcane cover the use of SAR signals for retrieving sugarcane height and harvest events. Two of the foremost publications [6,21] show that X-band signals from TerraSAR-X and C-band signals from ASAR saturate at a lower sugarcane height than L-band signals from ALOS and NDVI signals from SPOT sensors. It was observed that X-band signals experience saturation between 0.5 and 1 m, and C-band signals after 1 m. L-band signal backscatter increased significantly until 1.5 m, moderately from 1.5 to 2 m, and marginally from 3 m onward. In addition, a decrease in NDVI and C-band radar backscatter close to harvest was demonstrated, which was linked to the severe drying of the sugarcane. This drop was found to be on the same order for harvested fields where no (assumed) drying occurred before harvest, leading to ambiguity on the detection of harvest events. For L-band signals, a clear relationship was observed between HH backscatter and NDVI over the course of sugarcane growth. Here, the authors relate their simultaneous declines to water stress of growing maturing sugarcane plants.

Other works, led by Unicamp in Brazil, focused on the behavior of ALOS signals with changing sugarcane conditions [11,22,23]. It was found that signals can be used to discriminate between the first phases of sugarcane growth until the grand growth phase commences, after which the signals saturate. In addition, it was concluded that the backscatter acquired in HH increased significantly when rows perpendicular to the look direction were viewed (in the order of 1.2 dB), similar to what was observed in Baghdadi et al. [6], but only for the first eight months (until the last phenological phase). This phenomenon is also named 'flashing field' in the literature. No significant effect of the row direction was observed for HV signals. Emphasis was placed on the need of including rain events when interpreting the signals because of the significant signal increase, although this was not quantified.

In light of these literature findings, we identified a scarcity in published research on the extraction of spatial patterns from remote sensing imagery acquired over crops, especially sugarcane, their spatial consistency over time, and the effects from precipitation and varying beam-acquisition modes. The temporal development of spatial patterns in relation to the ground-measured biophysical parameters of sugarcane is also underexposed. Accordingly, in this manuscript, we first demonstrate the effects of sugarcane conditions on optical and SAR signals, with specific focus on biomass growth and precipitation events. Subsequently, we extracted the spatial patterns that are most consistent in time and compared these patterns between the different sensors in order to discover their agreements and differences. Finally, we related these patterns to intrafield sugarcane-biomass variations. The term consistency is commonly used in this document as an abbreviated denotation for the temporal consistency of spatial features in remote sensing imagery, which is further mathematically defined in Section 3.2.

Since the mentioned literature shares the finding that cross-polarization signals are generally more effective in mapping and monitoring sugarcane than copolarization signals, emphasis is put on the analysis of HV signals in this paper; HH signals are addressed, but not with the same level of detail.

2. Study Area and Data

2.1. Fields of Interest

The four sugarcane fields of interest are owned by one of the largest sugarcane producers and energy companies in Brazil. They are located 20 km from Piracicaba in São Paulo state. All fields are within 15 km from each other (see Figure 2), connected by interfield roads of two to ten meters wide. The agriculture in this region is dominated by sugarcane plantations and is relatively flat. The fields vary in size, ratoon cycle, and time of growth and harvest (see Table 1), but have the same soil type.

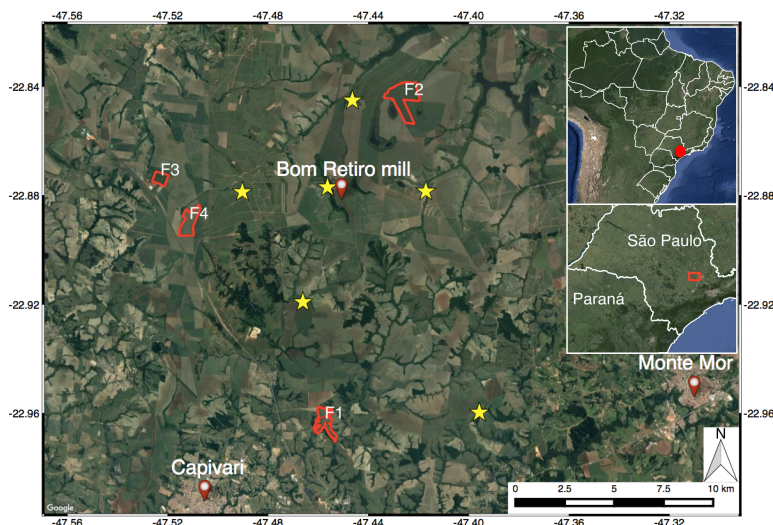


Figure 2. Location of the four fields of interest, indicated by red polygons, and weather stations, indicated by yellow stars.

Table 1. Characteristics of the four sugarcane fields; growth and harvest dates are not applicable to the entire field for F1 and F2.

Field Name	Ratoon	Area (ha)	Start of Growth	Harvest	Center Coordinate (Latitude, Longitude)
F1	1st cycle	58	30/10/2014	07/10/2015	-22.9607° , -47.4578°
F2	1st cycle	115	14/10/2014	07/12/2015	-22.8558° , -47.4228°
F3	2nd cycle	25	15/08/2014	26/07/2015	-22.8738° , -47.5230°
F4	9th cycle	59	01/08/2014	21/07/2015	-22.8895° , -47.5108°

2.2. Ground-Reference Data

From October 2014 to October 2015, an extensive set of ground-reference measurements were taken, amounting to a total of more than 3500 individual measurements on LAI, biometrics (including stalk height and thickness, leaf length, and cane spatial density), and biomass parameters (including cane water content and cane mass density). In Molijn et al. [24] we present in detail the methodologies for acquiring the measurements, the measured parameters, and the applied modelling and uncertainty analyses. The dataset itself was published for public use [25]. To summarize, two types of measurement locations were selected, spread over four fields (see Table 1 and Figure 3). The first type of locations consists of 15 Elementary Sampling Units (ESUs) that were selected based on the intrafield variability seen in Landsat-8 (LS8) NDVI images of previous years. The second type consists of six ESUs for biomass (ESUBs). Each location covers approximately an area of 20 by 20 m and was, on average, revisited 15 times during the growth period (see also Section 2.3). At the ESUBs, LAI and biometrics, consisting of sugarcane thickness, height, and cane spatial density, were measured, as well as dry and wet biomass. At the ESUs, only biometric measurements were taken and used as input parameters in a biomass-estimation equation, which was calibrated and validated by the measured biomass samples. This methodology allows for estimating, in a nondestructive manner, the sugarcane biomass in each ESU. Since only biometrics and LAI were measured at these locations, ground conditions were not affected and could be used for the analysis of the remote

sensing signals. The goodness of fit, expressed as the coefficient of determination, between estimated and measured biomass was 0.90, which gave us sufficient confidence in the validity of the method for estimating biomass at the required ESUs for this study. Biomass samples were also used for retrieving cane mass density and leaf mass per cane.

It should be noted that, for Fields F2, F3, and F4, the start of growth dates apply for the entirety of each field. For Field F2, sugarcane harvesting lasted for several weeks due to delays. Consequently, we analyzed this field until its first harvest event. For Field F1, certain parts of the field were uniformly grown over the course of the remote sensing observations. The parts that hosted ESUs and ESUBs were extracted and individually incorporated in the analyses based on their start of growth dates.

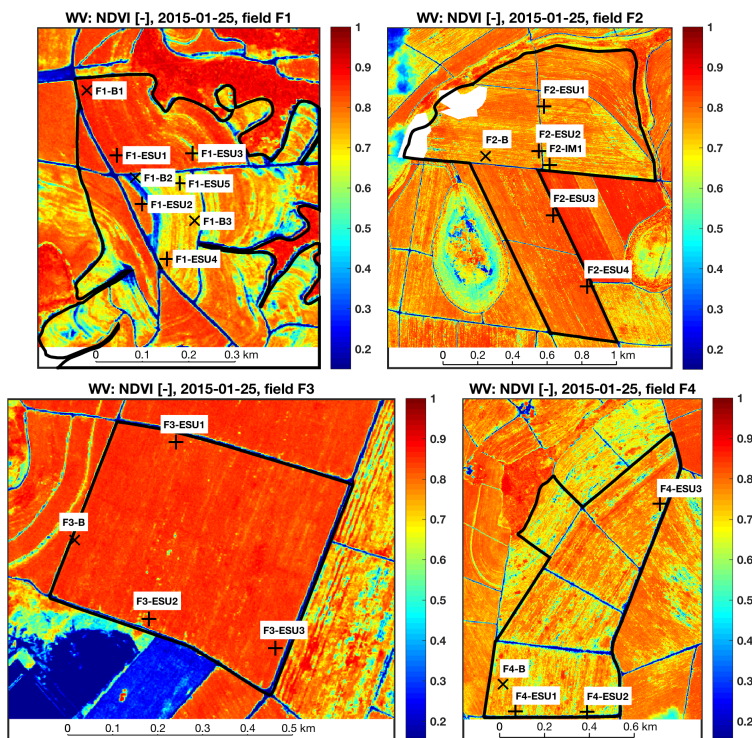


Figure 3. Delineation of the four studied sugarcane fields. Background map shows NDVI from WorldView-2 on 25 January 2015. Plus signs (+) and crosses (x) show, respectively, the location of the elementary sampling units (ESUs) and elementary sampling units for biomass (ESUBs). White patches in Field F2 are masked clouds. North direction is up.

In this manuscript, sugarcane biomass is defined as cane biomass, which is the the biomass of a single stalk (i.e., stem) together with the biomass of its leaves. The product of cane biomass with cane spatial density gives the biomass per unit area, commonly expressed as tons cane per hectare (TCH) [26–29]. The TCH estimates of all four fields were used for backscatter analysis (Section 3.1). For subsequent analyses, only the TCH estimates of Field F2 were used due to three reasons. First, the time series of the satellite imagery covers the entire growth period of this field as opposed to Fields F3 and F4. Second, it is

the only field with four ESUs (see Figure 3), and in this field the largest intrafield TCH differences were measured (see the left part of Figure 4). Here, the figure's uncertainty bars represent one standard deviation on either side of the mean profile based on the computed uncertainties associated with the measurement error and local variability; see Molijn et al. [24] for more detailed explanations. Third, on 19 December 2014 (66 days after start of growth, cane height up to one meter) in Field F2, 24 intensive measurements were taken for an indication of intrafield biomass variability. In total, the locations were spread over 400 m with spaces of 5, 10, or 20 m north of point F2-IM1 (see Figure 3). The resulting biomass estimates along this profile (see the right part of Figure 4) showed gaps and peaks up to 50% higher than the field average TCH (approximately 30 tons per hectare) and no evident spatial trend. These variations are a direct result of variabilities in cane spatial density.

The development of a selection of the measured parameters during the ground campaign (see Figure 5) shows a deflection point for leaf length and stalk thickness during the canopy-development phase and a continuous increase of stalk height until harvest. The ratio of leaf biomass over cane biomass shows that leaf biomass is dominant during, approximately, the first 50 days (i.e., establishment phase), after which the stalk biomass dominates and the leaf biomass contribute rapidly declines. The main reason is the decline in leaf wet content and increase in stalk mass density combined with stalk growth [24].

Finally, daily cumulative precipitation measurements were collected by seven weather stations. At least one weather station was within 10 km distance of each field. In terms of detected rain, the stations were in 96% agreement with each other, based on which we made the assumption that all fields were subject to the same condition of being precipitated or not precipitated.

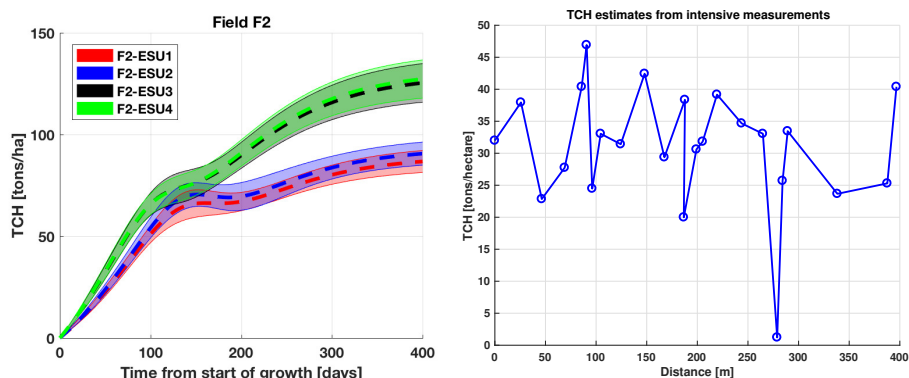


Figure 4. Estimated tons cane per hectare (TCH) over growth time for each elementary sampling unit (ESU) of Field F2, along with the associated uncertainties (left). F2-ESU1 and F2-ESU2 nearly overlap; similarly for F2-ESU3 and F2-ESU4. Profile of biomass estimates 400 m along from the intensive measurement at Field F2 (right).

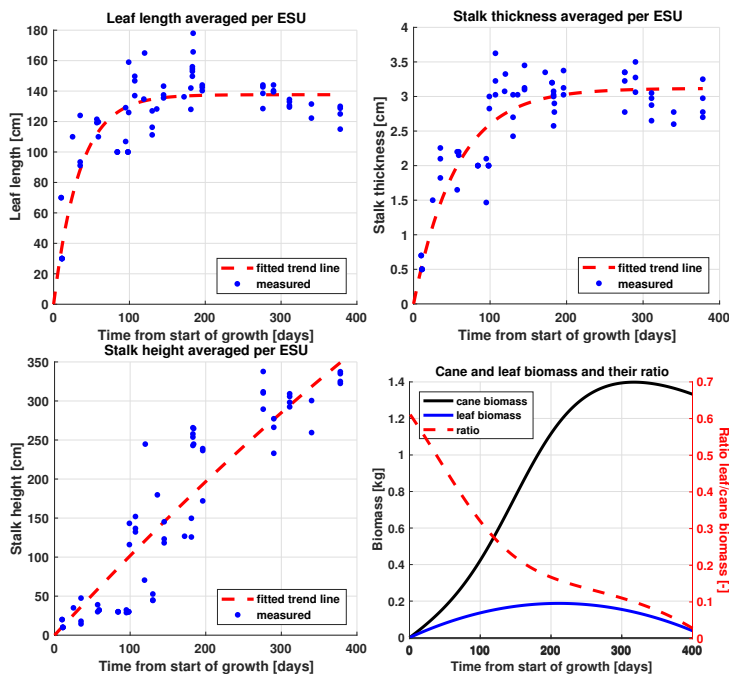


Figure 5. Temporal developments of measured leaf length, stalk thickness, and stalk height, averaged per ESU and accompanied by fitted trend lines. The cane and leaf biomass (in kg) per cane and their ratio over time are based on the trend lines of individual stalk- and leaf-biomass measurements.

2.3. Remote Sensing Data

2.3.1. Acquisitions and Preprocessing

Sensor characteristics vary in terms of wavelengths, polarizations, resolutions, and beam angles (see Table 2). Radarsat-2 (RS2) and ALOS-2 data were multilooked to a 30 by 30 m resolution. Sentinel-1 (S1) images were resampled to the same 30 by 30 m resolution grid. All radar images were radiometrically calibrated, geometrically corrected, terrain-corrected, and thermal-noise-corrected. SAR images were labelled as precipitation (or rain)-affected based on the time of acquisition and daily cumulative precipitation measurements from the nearest weather station. No precipitation threshold was applied.

The acquisition schemes of ground-reference and remote sensing data (Figure 6) show that the entire growth cycles are covered by ground and space measurements, except for SAR data during the first three months of Fields F3 and F4. Of the 29 LS8 acquisitions available, 11 had clear-sky observations for all fields. For all sensors, the number of acquisitions were counted between 1 July 2014 and 31 December 2015, i.e., just before the first start of growth and after the last harvest event of the fields. WorldView imagery consists of two WorldView-2 (WV2) images and one WorldView-3 (WV3) image. The WV2 images were acquired on 25 January and 11 August 2015. The WV3 image was acquired after the harvest events on 21 March 2016, and was used for additional visual inspection of the fields.

Table 2. Characteristics of available remote sensing data. ¹ RS2 = Radarsat-2, S1 = Sentinel-1, LS8 = Landsat-8, WV = WorldView; ² FQ: Fine Quad, W: Wide, S: Standard, EW: Extra Wide mode, SM: Stripmap, ASC: ascending pass, DSC: descending pass; ³ incidence angles; ⁴ UTC and approximate times, local time -2 or -3 hours; ⁵ regridded by ESA to medium resolution GRD product of 93 m × 87 m resolution; ⁶ five images with HV + HH, and three images with only HH; ⁷ approximate SLC product resolutions at field in meters; [†] rain label based on the daily cumulative previous to the day of acquisition; * rain label based on the daily cumulative on the day of acquisition. Reported resolutions and angles are mainly based on sensor and product descriptions [30–36].

Sensor ¹	Instrument	Mode ²	Polarizations/ Bands Used	Nominal Resolution (Slant Range × Azimuth) ⁷	Angle Fields ³ (°)	Acquisition Frequency/Time ⁴	Images
RS2	C-band (5.5 cm)	FQ16W (ASC)	HH + HV +	5.2 × 7.6	35.7–36.3	24 days/21:45 *	14
		FQ21W (DSC)	VH + VV		40.9–41.4	24 days/08:32 [†]	8
		S5 (ASC & DSC)			40.9–41.4	24 days /21:45 * & 08:32 [†]	2
		S6 (ASC)	HH + HV	13.5 × 7.7	41.5–41.9	24 days/21:49 *	5
		S7 (ASC)			48.4–48.9	24 days/20:55 *	9
S1	C-band (5.5 cm)	EW (DSC)	HH + HV	11.5 × 43 ⁵	36.0–36.6	12 days/08:38 [†]	33
ALOS-2	L-band (24 cm)	SM (ASC & DSC)	HH + HV or HH	6.0 × 4.3	31.3–42.8	14 days /03:49 [†] & 14:43 [†]	8 ⁶
LS8	Optical	Reflectances	RGB + NIR	30 × 30	NA	16 days/10:10	29
WV	Optical	Reflectances	RGB + NIR1	<2 × <2	NA	NA	3

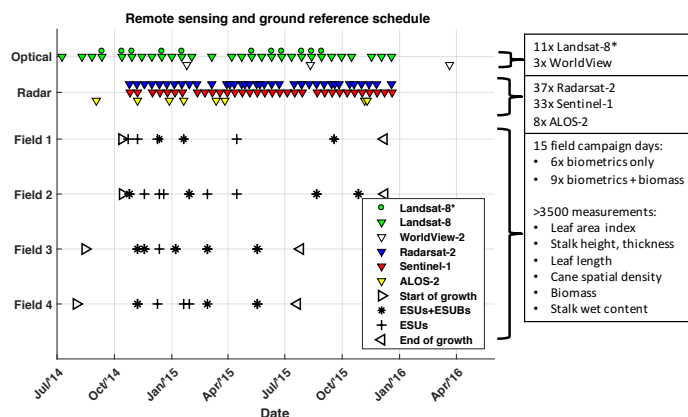


Figure 6. Temporal distribution of remote sensing images and ground-reference measurement dates. (Right box) summary of acquisitions, including number of image acquisitions per sensor, counted over growth duration. Landsat-8* symbols, LS8 acquisitions without cloud obstruction for all fields.

The remote sensing images of Field F2 taken by the different sensors, acquired approximately at the same date (see Figure 7), illustrate the spatial variabilities of the signals per sensor and their differences between sensors. Apart from the difference in backscatter magnitude between the C-band and L-band, the images were typically noisy and show no clear pattern agreement between the sensors. The most visually dominant field features included intrafield roads (low NDVI) versus the sugarcane plants (high NDVI). Outside the field, the most important feature was the water-drainage channel along the upper field border, as visible in the ALOS-2 image. This strip naturally contains more water, especially along the upper-right border due to elevation differences (see Section 2.3.3).

As described in Section 1, one of the factors causing variability in SAR backscatter intensities over vegetation is a changing incidence angle. This applies to the RS2 images as well due to the various acquisition modes. Even though the effect of incidence angle was reported to be minimal compared to the effects from varying soil and vegetation structure conditions, we applied an incidence-angle correction on RS2 images in order to more accurately compare the signals from RS2 and S1. Overall, considering the growth characteristics of sugarcane, we assumed high backscatter attenuation through the canopy based on the described literature and the high volumetric density of sugarcane. Consequently, we applied the following incidence-angle correction, based on the cosine dependency as proposed by Attema and Ullaby [37]:

$$\sigma_{RS2}^0 = \sigma_{RS2}^0 \frac{\cos \theta_{S1}}{\cos \theta_{RS2}}$$

where σ_{RS2}^0 is the beam-adjusted RS2 backscatter coefficient, σ_{RS2}^0 is the original RS2 backscatter coefficient, θ_{RS2} is the RS2 incidence angle over the fields (which differs per mode), and θ_{S1} is the S1 incidence angle over the fields (taken as 36.3°). Notice that this angle correction is equivalent to using the Gamma nought (γ_0)-normalized backscatter. As an illustrative example of the effects of this equation, we applied the correction on one S1 EW image (swath width 400 km) covering more than 40,000 sugarcane fields that were delineated by the Canasat project [4,38]. Before correction, the median signal backscatter ranged between -14.3 dB at a 29° and -16.1 dB at 47° incidence angle (see the left part of Figure 8). After correction, the spread of the median signal backscatter ranged from -14.6 dB at a 29° to -15.3 dB at 47° incidence angle; the median at 36.3° before and after correction remained -14.9 dB (see the right part of Figure 8).

Since the RS2 signal radiometric stability was smaller than 1 dB [30], we accepted this method for reducing the spread.

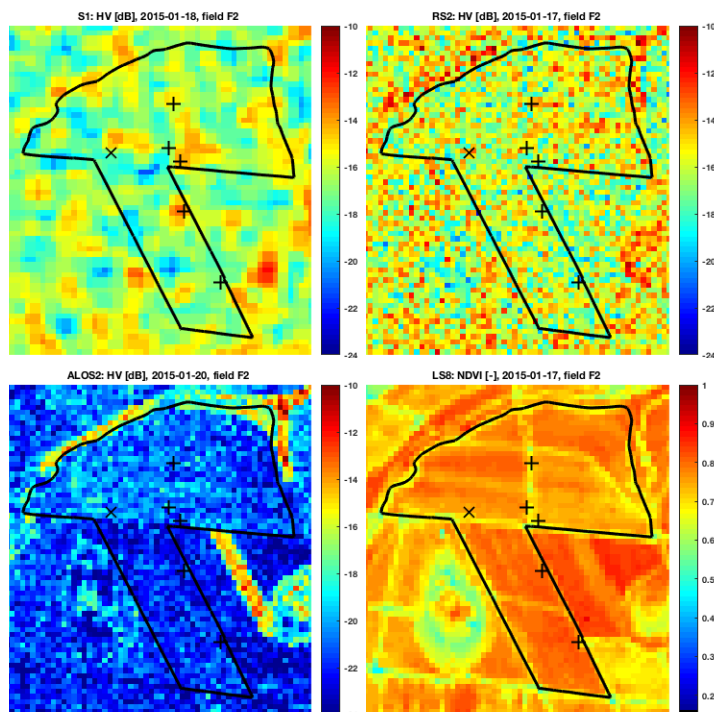


Figure 7. Sample images from available sensors, spatially averaged to 30 by 30 m pixel spacing, at approximately the same date for Field F2, around 100 days after start of growth, with a sugarcane height of 1 m. Plus signs (+) and crosses (×), location of ESUs and ESUs for biomass (ESUBs), respectively. North direction is up and field spans approximately 1.4 km northward by 1.7 km eastward.

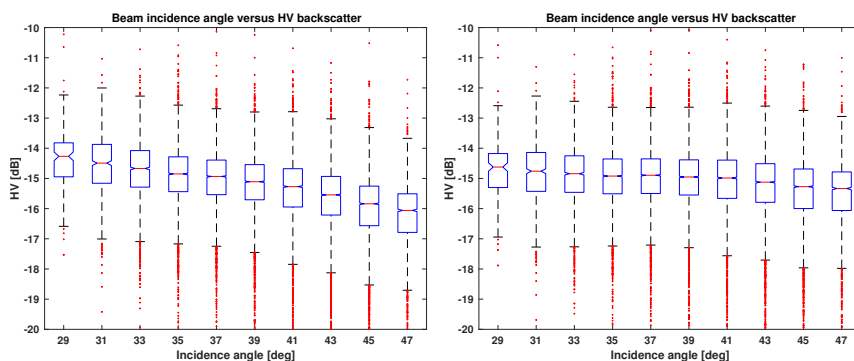


Figure 8. S1 EW HV backscatter versus incidence angles, before incidence-angle correction (left) and after incidence-angle correction (right).

2.3.2. Signal-Noise Budgets and Accuracies

For computation of the temporal consistency of spatial patterns from remote sensing imagery (as is further explained in Section 3.2), sensor-specific signal-noise budgets are quantified. For SAR sensors, we considered radiometric stability (ζ_{RAD}), maximum Noise Equivalent Sigma Zero (ζ_{NESZ}) and speckle noise (ζ_{SPECKLE}) as contributions to the total noise budgets. The first noise, ζ_{RAD} , acted as an multiplicative factor, whereas the second, ζ_{NESZ} , had an additive nature. Both were taken from nominal sensor and product specifications [30,32,33,39] (see Table 3). ζ_{SPECKLE} is dependent on the number of looks and, hence, on the spatial averaging window. For intra- and intersensor imagery-consistency analysis, we resampled the signals of all sensors over a spatial window of 90 by 90 m (the approximate S1-EW GRD product resolution). Based on the original SLC ground resolutions, the resulting numbers of looks are 11 for S1-EW, 128 for RS2-FQ, 55 for RS2-S, and 186 for ALOS-2. For ζ_{SPECKLE} , it was assumed that the scattering medium in the averaging window was homogeneous. Total noise ζ_{TOTAL} is quantified for three reference backscatter values that are indicative for the backscatter range found at the sugarcane fields (see Table 3).

As an explanatory example, the noise contributions of S1-EW were specified as follows. ζ_{RAD} (expressed as 3σ) is a constant equivalent to 0.5 dB, ζ_{NESZ} is equivalent to -22 dB and needs to be treated as relative to the signal [40], contributing 0.1, 0.3, and 1.4 dB to the three reference signals of -10 , -15 , and -20 dB, respectively. ζ_{SPECKLE} is dependent on the number of looks (eleven in this case), but not on the reference signal, and is equivalent to 1.4 dB. As a result, ζ_{TOTAL} is approximately 1.7 dB for a backscatter coefficient of -15 dB, whereby ζ_{SPECKLE} is the major contributor to the noise for high-valued signals and ζ_{NESZ} becomes increasingly important for the low-valued signals.

Table 3. Signal-noise budget associated with data from different sensors and modes for 90 by 90 m spatially resampled SAR signals, with total noise computed for three backscatter intensity scenarios. Radiometric stability and NESZ values were extracted from sensor and product descriptions [30,32,33,39].

Parameter	S1	RS2-FQ	RS2-S	ALOS-2
$\zeta_{\text{RAD}}(3\sigma)$	0.5 dB	1 dB	1 dB	1.2 dB
ζ_{NESZ}	-22 dB	-35 dB	-29 dB	-28 dB
ζ_{SPECKLE}	1.4 dB	0.4 dB	0.6 dB	0.3 dB
$\zeta_{\text{TOTAL}}@-10$ dB	1.5 dB	0.5 dB	0.7 dB	0.5 dB
$\zeta_{\text{TOTAL}}@-15$ dB	1.7 dB	0.5 dB	0.7 dB	0.5 dB
$\zeta_{\text{TOTAL}}@-20$ dB	2.7 dB	0.5 dB	0.8 dB	0.5 dB

With regard to the noise associated with NDVI from LS8, Akdim et al. [41] reported NDVI root mean square error (RMSE) values of 0.047 and 0.033 through comparing simulated NDVI values of LS8 with SPOT-4 images and LS8 with RapidEye images, respectively. To verify and improve these values, we assessed the temporal NDVI variation over stable vegetation, for which nearby native forest areas were selected, and computed the RMSE based on consecutive images with a maximum time difference of 32 days (i.e., maximum two revisits apart). We assumed that the differences of surface reflectances caused by the vegetation itself within this time frame were insignificant with respect to sensor noise. We also acknowledge that (illumination and sensor) geometry and topographic conditions affect NDVI variations [42,43] and contribute to this noise. Computation resulted in an additional RMSE of 0.027 and, through averaging the reported RMSEs (i.e., by taking the square root of the mean of the square of the RMSEs), we concluded on a total NDVI noise budget of 0.037.

Geolocation accuracies, defined as the RMSE estimated from point target monitoring, were smaller than 6 m for the relevant beam modes of S1, RS2, and ALOS-2 [44–47]. The geolocation products of the optical sensors are commonly specified using circular error at the 90th percentile (CE90), which, for LS8 is less than 20 m [48,49], and for the WorldView sensors this is less than 5 m [36,50,51].

2.3.3. Elevation

The widely used Shuttle Radar Topography Mission (SRTM) data product provided by NASA was used for the elevation information of the sugarcane fields. Acquired in 2000, the DEM is offered as a 30 m resolution product. Since it was derived based on C-band and X-band interferometric SAR (InSAR) techniques, several studies reported that errors in SRTM increase with increasing vegetation cover, especially for forests, due to the limited penetration of the X-band and C-band signals to the ground [52,53]. The methodologies these and similar works propose for correcting overestimations are based on vegetation masks and coarse-resolution-estimated vegetation heights. Hence, for this study we assumed these elevations could be used for interpreting relative in-field ground-elevation differences at our time and area of interest. The variations in elevation of Field F2 (Figure 9) show that adjacent to the boundaries of the upper part of the field, especially at the lowest-elevated places with the highest slopes, native vegetation strips are situated, serving as drainage systems for excess water discharge and water preservation.

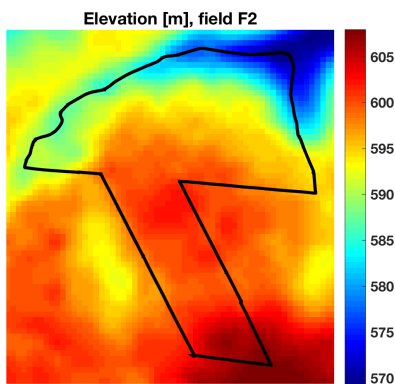


Figure 9. Elevation of Field F2. North direction is up and field spans approximately 1.4 km northward by 1.7 km eastward.

3. Methods

3.1. Backscatter Analysis

In order to investigate the effect of increasing biomass on remote sensing signals, the average of the modeled ESU biomass values was taken, expressed in TCH, per sugarcane field at each remote sensing acquisition. It is assumed that these averages are representative for the biomass conditions of the entirety of the corresponding field. Subsequently, the average of the remote sensing signals was taken per sugarcane field and plotted against TCH values. For each TCH bin, the average and standard deviation were taken for illustrating development and spread over time. Together with the TCH axis, the time of growth and stalk-height axes is visualized as well, though only for approximate and indicative purposes since these are not strictly linearly related to the TCH values. We defined the saturation point of the remote sensing signal as the first point in time when it does not increase with increasing biomass. For visualizing the effects of rain on SAR backscatter, box plots were produced based on SAR signals after saturation points.

3.2. Intrasensor Imagery-Consistency Analysis

For each sensor, the temporal coherence of the spatial features in remote sensing imagery was examined. For this, we introduce a spatiotemporal signal-consistency measure, based on which time windows with maximum consistencies are deduced. Remote sensing images were first spatially averaged over a 90 by 90 m window (approximately the S1 GRD product resolution, see Section 2.3). Subsequently, in order to extract pattern dynamics, we applied *z-score* normalization ($Z_{i,p}$) for each pixel (p) per image (i), by subtracting the field average of the signals (μ_i) from the pixel value ($X_{i,p}$) and normalizing by the field standard deviation of the signals (σ_i):

$$Z_{i,p} = \frac{X_{i,p} - \mu_i}{\sigma_i} \quad (1)$$

For the temporal consistency of the images, we compared all N_i available images from the start of growth until harvest with each other, i.e., for $\frac{N_i(N_i-1)}{2}$ possible combinations. Consistency metric C is based on the ratio between log-likelihoods:

$$C = 1 - \frac{\log L_0}{\log L_1} \quad (2)$$

The first element, $\log L_0$, represents the log-likelihood of the null hypothesis that the two images belong to the same spatial distribution, and hence share the same spatial pattern, and is expressed by:

$$\log L_0 = -\frac{1}{N_p} \sum_{p=1}^{N_p} \frac{(Z_{i,p} - Z_{j,p})^2}{\sigma_{Z_{i,p}}^2 + \sigma_{Z_{j,p}}^2} \quad (3)$$

The second element, $\log L_1$, represents the log-likelihood of the alternative hypothesis that the two images do not share the same spatial pattern, but instead a random permutation of image values:

$$\log L_1 = E \left[-\frac{1}{N_p} \sum_{p=1}^{N_p} \frac{(Z_{i,p} - \mathfrak{S}_{N_r}(Z_{j,p}))^2}{\sigma_{Z_{i,p}}^2 + \sigma_{Z_{j,p}}^2} \right] \quad (4)$$

Here, N_p is the number of *z-score* values in the field and N_r is the number of iterations for which the random permutations, $\mathfrak{S}_{N_r}$, of the *z-score* values of image j , Z_j , is conducted. This N_r was heuristically set to 100. In addition, $\sigma_{Z_{i,p}}^2$ and $\sigma_{Z_{j,p}}^2$ are the sensor and acquisition mode-specific variances per pixel p in *z-score*-equivalent values, equal to pixel-specific $\frac{\epsilon_{\text{TOTAL},p}^2}{\sigma_i^2}$ and $\frac{\epsilon_{\text{TOTAL},p}^2}{\sigma_j^2}$, respectively. The corresponding total noise budgets were presented in Table 3. For LS8, noise is based on the RMSE value as reported in Section 2.3, without taking into account noise reduction from spatial averaging. The resulting consistency metric, with $-1 \leq C \leq 1$ defining the range, is equal to the upper limit (one) for two equal images, is equal to zero for two perfectly random images, and is equal to the lower limit for two inversely correlated images. The pixel variances give importance to the samples by weighing the differences between pixel values. Since pixels with high backscatter are generally associated with low variances (see Table 3), more weight is given to the outcome of the absolute differences between the pixels. Hence, comparisons between images with distinct and similar patterns result in a higher consistency value as compared to images with similar but less distinct patterns. All SAR images are labelled for rain events; for the optical images, the potential effects of rain conditions on consistency were not analyzed and hence the images are not labelled.

3.3. Intersensor Imagery-Consistency Analysis

From previously presented consistency matrices per sensor, we computed an average *z-score* image from the *z-score* images present in the time windows with highest consistencies. The similarities of these average *z-score* images between the SAR and optical sensors were measured by the (Pearson) correlation.

3.4. Intrafield Variability Analysis

In addition to the relationship of field-averaged remote sensing signals with sugarcane biomass (Section 3.1), we also analyze the relation between spatial patterns in the averaged *z-score* images and the intrafield sugarcane biomass estimates, quantified as TCH. This is examined by two approaches: based on the TCH estimates at the ESU locations (varying in space and time), and based on the TCH estimates at the intensive-measurement (IM) locations (varying only in space).

3.4.1. ESU locations

Since the ESUs cover approximately one LS8 resolution cell, the average *z-score* images are based on remote sensing images that were resampled to a 30 by 30 m grid. For S1, given its coarse resolution, this results in oversampling. For WV2, the original resolution *z-score* values were averaged that were located within the ESU dimensions. The selections of the remote sensing images are based on the time windows with maximum consistency, as were defined in Section 3.2. For each sensor, the *z-score* value was extracted that was nearest to the corresponding ESU locations. In addition, per sensor and per ESU, the range of TCH values was taken from the start until the end of the time window based on the modeled TCH graphs (see Figure 4). For WV2, both images were analyzed.

3.4.2. IM locations

Biomass estimations along the spatial profile in Field F2 (see the right part of Figure 4) were correlated to two sets of *z-score* values:

- Based on averaged *z-scores* within the window of maximum temporal consistencies.
- Based on the single image per sensor that was closest to the intensive-measurement acquisition time. For the WV2 image (acquired 103 days after start of growth, on 25 January 2015), the original-resolution *z-score* value was taken closest to the recorded GPS location.

4. Results

4.1. Backscatter Analysis

For C-band HV (and, similarly, for C-band HH, which is not visualized here) the saturation point of the signal occurred at approximately 25 tons/ha (90 cm stalk height) for rain-affected signals and approximately 35 tons/ha (120 cm stalk height) for signals not affected by rain (see Figure 10). These saturation points coincided with the canopy-development phase, after which the increase in leaf length stagnated and the grand growth phase of the stalk commenced, with increasing stalk biomass dominance over leaf biomass (see Figures 1 and 5). Afterward, both signals that were affected by rain and not affected by rain tended to converge. In addition, signals affected by rain showed backscatter values for bare soil and marginal canopy that were close to those for fully mature sugarcane. This is contrary to signals not affected by rain, which showed clear differences.

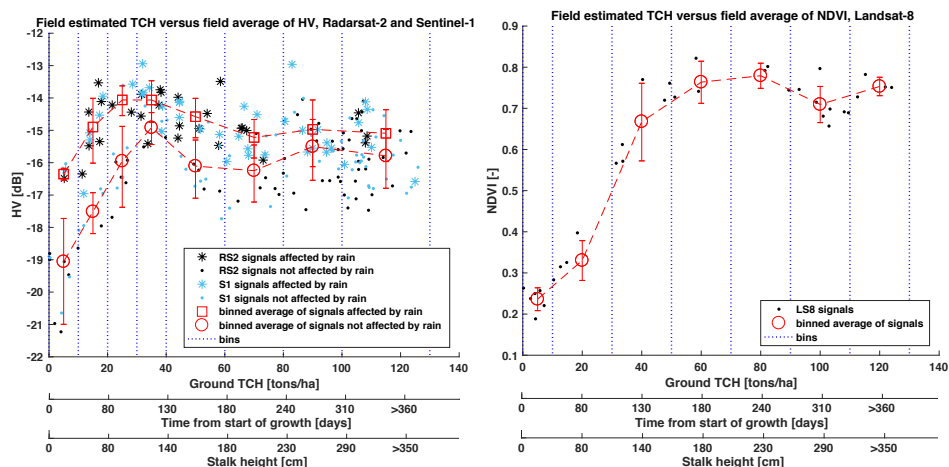


Figure 10. RS2 and S1 combined HV (left) and LS8 NDVI (right) signals versus binned values of estimated TCH, showing the effect of rain events on the signal. RS2 values were corrected for beam angles to the S1 angle. Additionally, the approximate and indicative axes of time from start of growth and stalk height are added. Bars illustrate standard deviations of remote sensing signals within each TCH bin.

For both HV and HH, the median of the signals affected by rain after the saturation point significantly differed (with 95% confidence) from the median of the signals not affected by rain after the saturation point (see Figure 11). In addition, the figure demonstrates that the magnitude of this difference is comparable for both polarizations.

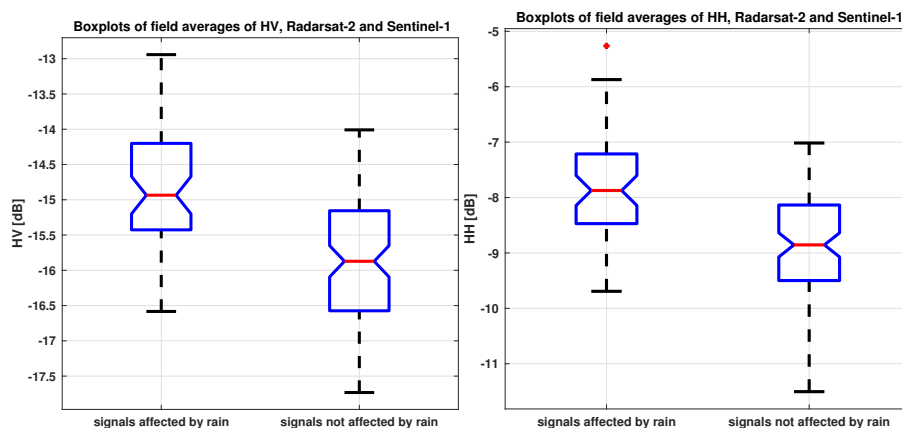


Figure 11. Box plots of RS2 and S1 HV (left) and HH (right) signals combined over sugarcane fields after their saturation points, showing the effect of rain events on the signal. Central (red) line indicates the median, and bottom and top box edges represent the 25th and 75th percentiles, respectively. Whiskers spanning to the most extreme data points were not considered outliers, and cover approximately 99% of the values. Notches represent the 95% confidence intervals of the median.

For NDVI, the saturation point occurred at 80 tons/ha (240 cm stalk height), which coincided with the transition from the grand growth phase to the maturing and ripening phase.

For ALOS-2 signals in HV and HH, signals did not show evident saturation points with increasing biomass (see Figure 12). Due to the limited number of available images, signals affected by rain and not affected by rain were combined for the averaging per bin. Nevertheless, similarly to C-band SAR, the L-band signals, especially in HH, showed sensitivity to surface wetness.

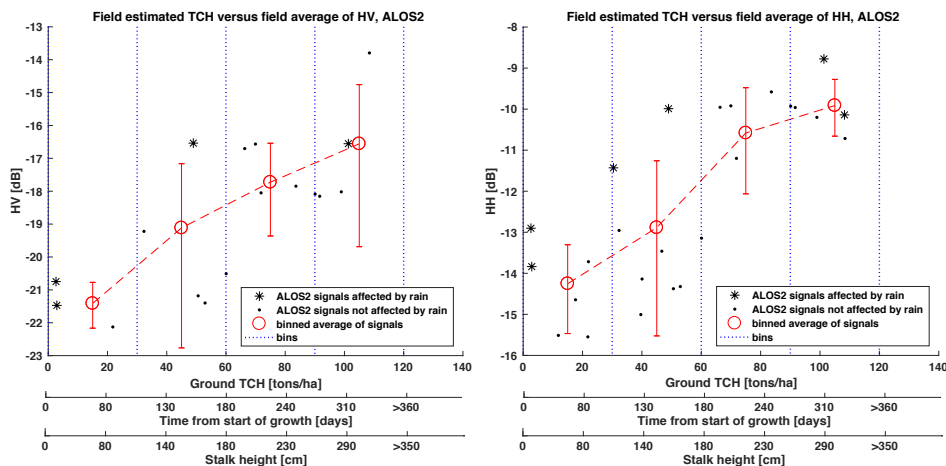


Figure 12. ALOS-2 HV (left) and HH (right) signals versus TCH, with signal averages taken from combined signals affected by rain and not affected by rain. Approximate and indicative axes of time from start of growth and stalk height were added.

4.2. Intrasensor Imagery-Consistency Analysis

From the consistency matrices between images over time (see Figure 13), different dynamics can be observed. First of all, the ranges of consistencies differ per sensor, whereby the consistency is highest for LS8, followed by ALOS-2, RS2 and finally S1. Also the time frame during which highest consistencies are observed differ.

For LS8 NDVI, images acquired during early crop growth showed relatively low consistency with later images. The image pair taken 63 and 95 days after growth was more similar due to the relatively small difference between the dates (two revisits' difference). The time window with the most similar images occurred from the late grand growth stage (approximately 210 days after growth) until harvest. The (not visualized) consistency of the WV2 NDVI images taken 103 and 301 days after growth (on 25 January and 11 August 2015, respectively) was 0.4, which is equal to the LS8 consistency between images taken on 95 and 303 days after growth, and lower than those between later LS8 NDVI images.

For ALOS-2, the highest consistency was found between two consecutively acquired images during the grand growth phase. The maximum value for ALOS-2 images was comparable to the highest consistencies found for LS8.

For the C-band SAR sensors, especially S1, most consistency values ranged between zero and 0.4, whereby the majority of the images showed a high degree of randomness in spatial patterns over time. For RS2, a window with higher values was found between approximately 200 and 330 days after growth. Before showing the details on this window and providing the corresponding explanation, we show a

more clear view on the variation in consistencies over time by taking the average over time windows for both sensors (see Figure 14). Each value in these matrices represents the average of the consistencies in Figure 13 that fall in the (triangular) window from the image indicated on the horizontal axis until the image indicated on the vertical axis. Through this representation, it could be found that images until approximately 200 days after start of growth were generally not similar with each other and hence showed a high degree of temporal randomness in spatial patterns. An exception is a minor time window that coincides with the period around the saturation point. For both C-band SAR sensors, the highest consistencies occurred for the images acquired during the second half of the grand growth phase until the maturing phase (i.e., between approximately 200 and 320 days after start of growth), after which the ripening phase commenced and images started to decorrelate. This time window commenced after the signal-saturation point (see Figure 10) when the backscatter decreased and coincided with the decrease in leaf biomass, though before severe leaf senescence occurred.

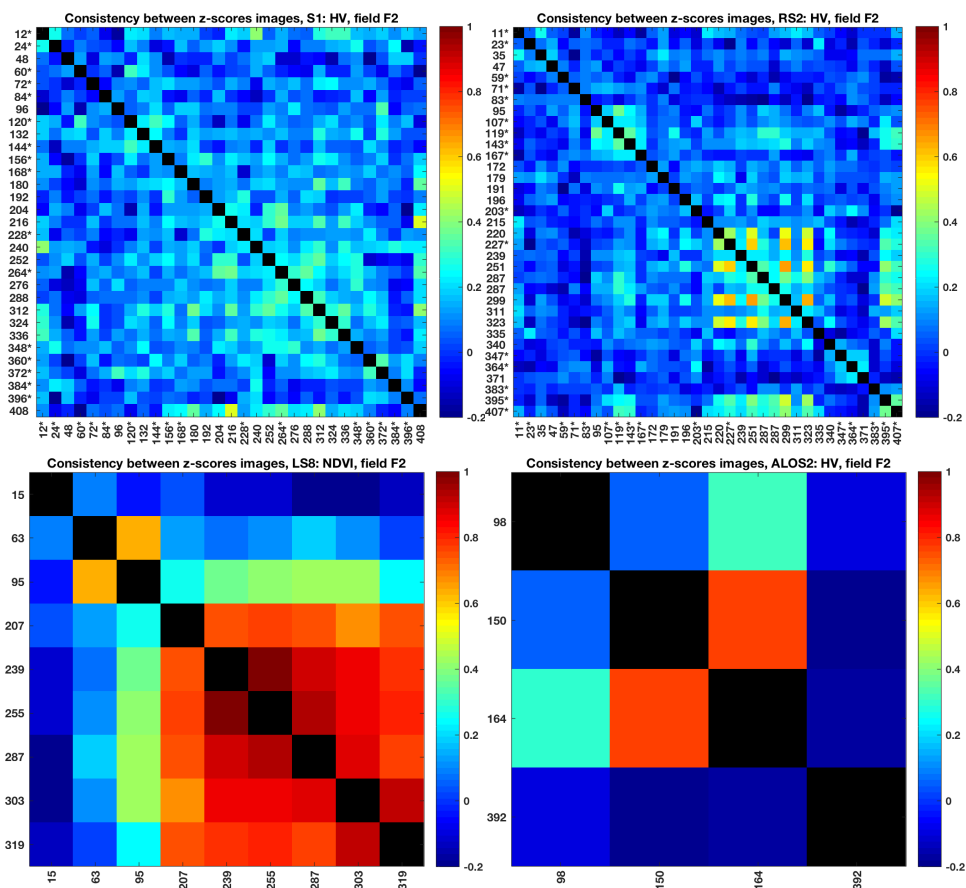


Figure 13. Consistencies of all image combinations for different sensors. Both axes show days from start of growth and asterisks indicate which acquisitions were rain-affected. Diagonal values were omitted and colored black. The harvest of this field was 419 days after start of growth.

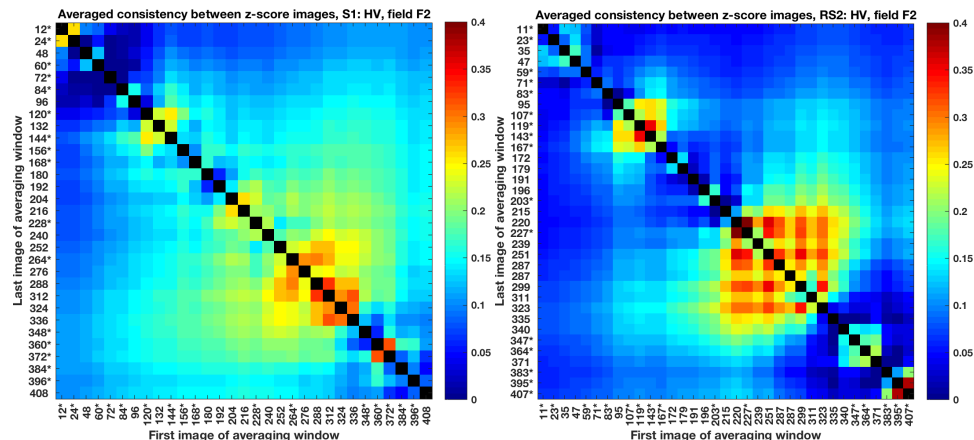


Figure 14. Consistencies for S1 HV (left) and RS2 HV (right) averaged over time windows based on the elements in Figure 13 with the first image on the horizontal axis and the last image on the vertical axis. Both axes present days from start of growth and asterisks indicate which acquisitions were rain-affected. Diagonal values were omitted and colored black.

The subset with highest consistencies for RS2 from Figure 13 accompanied by beam-mode information shows that the highest similarities were found between images taken in the FQ16W and S6 modes from ascending passes (see the left matrix of Figure 15). This is confirmed by the averaged consistencies for all beam-mode combinations within the window from 220 to 323 days after start of growth (see the right matrix of Figure 15). Contrary to the relatively high consistencies between these two modes, their consistencies with S7-ascending images are close to zero. This can be explained by the relative high incidence angle for S7, approximately 49° , with respect to the angles for FQ16W and S6, approximately 36° and 42° . The hour of acquisition for these modes was approximately similar (all around 21:00 UTC). The two images that were taken on the same day (287 days after growth) showed that changing beam modes (FQ21-descending and S7-ascending) results in low image similarity. In addition to their difference in incidence angle and look direction, low consistency may also be caused by the FQ21-ascending mode acquiring in the morning (when potential dew causes more surface wetness) versus the S7-ascending mode acquiring in the late evening (when the heat of the day causes lower moisture content).

Specifically regarding the effect of rain on similarities, the box plots of Figure 16 show that the median of the consistencies of the no-rain-to-no-rain (NR2NR) image pairs was higher than the median of the combined no-rain to rain (NR2R) and rain-to-no-rain (R2NR) image pairs, and the rain-to-rain (R2R) image pairs. This can be ascribed to the saturation effects and smaller spread of percentiles (see Figures 10 and 11) for the rain-affected images.

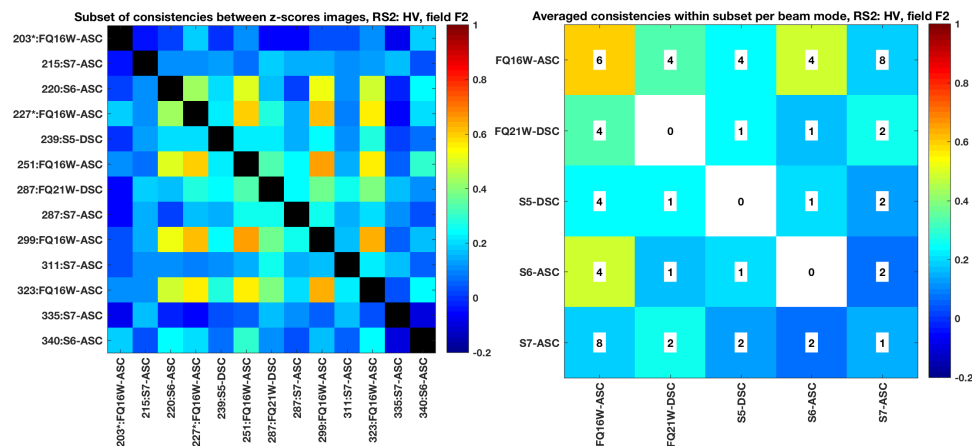


Figure 15. Subset of the RS2 HV consistency matrix of Figure 13 accompanied by beam-mode information (left). Both axes present days from start of growth and asterisks indicate which acquisitions were rain-affected. Diagonal values were omitted and colored black. In addition, averages of the consistencies between beams with the number of unique image pairs are displayed in each cell (right). White cells have no value as no image pair exists for those specific modes.

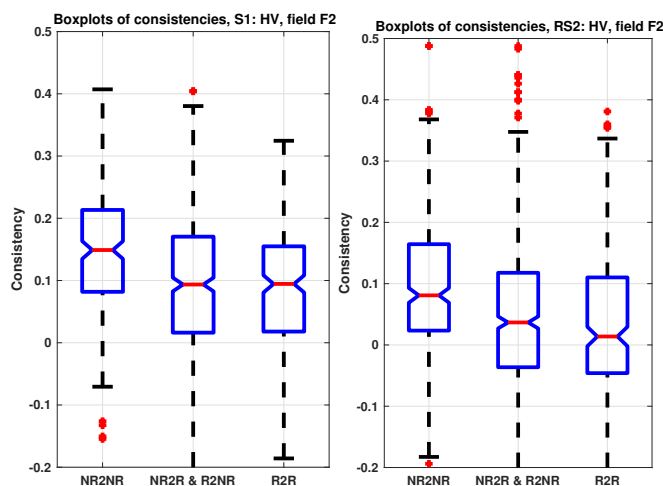


Figure 16. Box plots of all consistencies for S1 HV (left) and RS2 HV (right) per rain-condition pair, representing no rain to no rain (NR2NR), no rain to rain and rain to no rain combined (NR2R and R2NR), and rain to rain (R2R). Central (red) line indicates the median, and bottom and top box edges represent the 25th and 75th percentiles, respectively. Whiskers span to the most extreme data points not considered outliers and cover approximately 99% of the values. Notches represent the 95% confidence intervals of the median.

Hence, it may be deduced that the most similar images acquired by the C-band sensors could be found during the earlier-mentioned time window for images that were not affected by rain. In addition,

specifically for RS2 images that were acquired with changing beam modes, images with a high incidence angle and changing orbital passes (i.e., ascending and descending) should be avoided. Based on these findings, the following sensor-specific criteria may be summarized for selecting images that showed maximum temporal consistency:

- S1 HV: time window from 204 days until and including 336 days after start of growth, and only with nonrain-affected images, resulting in nine images.
- RS2 HV: time window from 220 days until and including 323 days after start of growth, only with nonrain-affected images and only for images acquired in the FQ16W and S6 modes, both ascending, resulting in four images.
- ALOS-2 HV: time window from 150 days until and including 164 days after start of growth, only with nonrain-affected images, resulting in two images.
- LS8 NDVI: time window from 207 days until and including 319 days after start of growth, regardless of rain conditions, resulting in six images.
- WV2 NDVI: only the image taken 301 days after start of growth, because Field F2 was partly masked by clouds in the image taken 103 days after start of growth (see Figure 3).

These *z-score* images were averaged for each sensor and visualized in Figure 17.

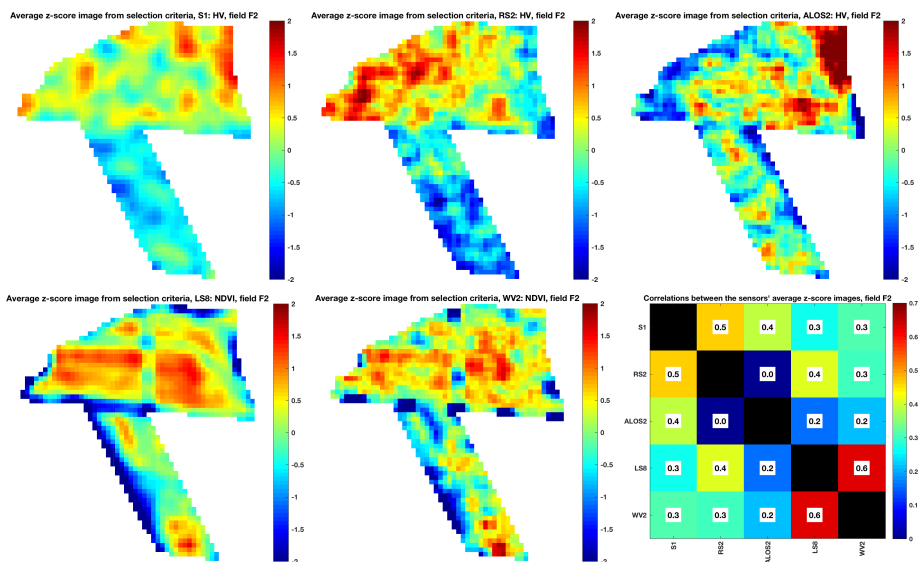


Figure 17. Field F2 average *z-scores* from images specified by the selection criteria for the different sensors (field plots), and corresponding correlations of these images between the different sensors (bottom right). Correlation coefficients are displayed in each cell, and diagonal values were omitted and colored black. The north direction is up, and the field spans approximately 1.4 km northward by 1.7 km eastward.

4.3. Intersensor Imagery-Consistency Analysis

The highest correlations between the averaged *z-score* images (see the lower right matrix in Figure 17) occurred between the optical sensors, followed by correlations between the C-band sensors. The images of both LS8 and WV2 showed agreement on the intrafield road features (see also the NDVI image of Field F2 in Figures 3 and 7) and regional differences between the northern and southern part of the field.

For the C-band SAR sensors, the images showed locally contrary behavior in the northern part of the field, i.e., the eastern part featured high values for S1 and low values for RS2, and conversely for the western part. More corresponding behavior was found in the southern part of the field. On a coarser level, the northern part of the field showed higher values than the southern part for both sensors.

For ALOS-2, the patterns with the highest *z-score* values occurred in the northeastern part of the field, coinciding with the lowest field elevations. As was described in Section 1, soil backscatter contribution, which is particularly dependent on soil-moisture content, is under standing vegetation conditions higher for L-band SAR than for C-band SAR. Consequently, higher values in the field's northeastern corner may be associated to high soil and plant moisture. The patterns in the southern part are not evidently consistent with image patterns of other sensors, and ground observations cannot give conclusive explanations for this.

4.4. Intrafield Variability Analysis

4.4.1. ESU Locations

Table 4 shows that the differences in *z-score* values per sensor are not indicative of ESU biomass differences, except for the WV2 image taken 103 days after growth. Due to this sensor's fine resolution and hence higher geolocation accuracy, in combination with the absence of speckle noise, the signals more accurately present the state of vegetation at such small locations. The difference between the August acquisitions is that these two images were taken before and after the saturation point.

Table 4. TCH ranges between the first and last image in window of images of estimated TCH and associated *z-score* values per sensor for all ESUs of Field F2. ¹ WV2 image taken 103 days after start of growth; ² WV2 taken 301 days after start of growth.

Name	ESU1	ESU2	ESU3	ESU4
TCH range S1	67 to 83	70 to 87	91 to 121	92 to 122
<i>z-score</i> S1	0.1	0.3	−0.5	−0.2
TCH range RS2	69 to 82	72 to 86	96 to 119	98 to 121
<i>z-score</i> RS2	1.4	−0.5	0.1	−0.9
TCH range ALOS-2	66 to 66	70 to 70	76 to 79	76 to 80
<i>z-score</i> ALOS-2	0.5	−0.4	−0.1	−0.6
TCH range LS8	68 to 82	70 to 86	92 to 119	93 to 120
<i>z-score</i> LS8	−0.2	0.1	0.3	−0.9
TCH range WV2 ¹	80	83	116	118
<i>z-score</i> WV2 ¹	1.1	1.1	1.2	1.2
TCH range WV2 ²	80	83	116	118
<i>z-score</i> WV2 ²	1.1	1.1	1.0	1.1

4.4.2. IM Locations

Based on the correlations (see Table 5), it may be deduced that none of the time window-averaged *z-score* images was indicative for locating biomass differences along such a measurement profile. This can be expected due to the low correlations between images close to the IM date, as shown in Figures 13 and 14, as well as the high local variability within a 30 by 30 m resolution cell and geolocation accuracy. Regarding the *z-score* images that were closest to the measurements, the first WV2 image (acquired 103 days after start of growth) showed the highest potential for indicating these differences in the field.

Table 5. Correlation coefficients between estimated TCH values at the intensive measurement locations and z-scores of images per sensor.

Sensor	Window Averaged	Closest Date
S1	0.1	0
RS2	0	−0.1
ALOS-2	0.2	0.3
LS8	−0.1	0
WV	−0.2	0.6

5. Discussion

The results presented the effects of sugarcane biomass growth, precipitation, and sensor characteristics on remote sensing signals, in particular from C-band SAR, L-band SAR, and optical sensors. These effects were first examined through the relationship between field-averaged remote sensing signals and sugarcane biomass. Second, remote sensing imagery was analyzed for the statistical coherence of spatial features in time. The identification of time windows per sensor allowed for the effective extraction of the most consistent patterns. It was found that rain events have their effect on this as well, specifically for the C-band SAR sensors. Subsequently, the patterns were extracted and compared between the various sensors. Finally, the relationship between these patterns and measured intrafield sugarcane biomass was investigated. Hereafter, these elements will be compared and discussed from an integrated perspective.

5.1. Backscatter Analysis

The first variant of sugarcane productivity monitoring was presented through the relationship between field-averaged remote sensing signals and sugarcane biomass. The saturation points of the signals differ per wavelength and, for SAR, are dependent on the wetness conditions of the surface. For C-band signals, saturation occurs earlier when affected by rain conditions than when unaffected by rain conditions. After saturation, backscatter values decrease, which can be ascribed to a combination of two effects. First, the increasing dominance of stalk biomass over leaf biomass (see Figures 1 and 5) causes a reduction in the depolarizing effect of volumetric scattering and an increase in the vertical copolarizing effect due to higher microwave interaction with the vertical stalks, and thus an attenuation of the cross-polarized signals. Second, there is a decreasing contribution of microwave interaction with water volume in the plant due to a decline in wet stalk content and leaf senescence (see Molijn et al. [24]). Both interactions were also (partly) reported by Baghdadi et al. [6,21] and further supported by the biophysical developments observed in Simoes et al., and Vieira et al. [5,17,19]. Even though the signals after the saturation points decreased and the two signal profiles at high biomass values tended to converge, the profile affected by rain clearly differed from the profile unaffected by rain. Furthermore, the perceived converge of the two profiles with increasing biomass values suggests that the contribution from microwave interaction with the volumetric features dominates over the contribution from surface wetness. Hence, the response of C-band SAR signals to surface wetness is a function of vegetation biomass. Additionally, the results showed comparable C-band backscatter values for bare soil when affected by rain with respect to the values for mature cane irrespective of rain conditions. Such confusion between signals acquired under wet-soil conditions and under mature sugarcane conditions was also observed for X-band SAR, for example by Baghdadi et al. [21]. Furthermore, Molijn et al. [54] showed that C-band backscatter fluctuations in time, caused by precipitation events, could be used for improving the classification of vegetation conditions. Inversely, when vegetation conditions are known, these signals may be used for detecting rain events and, in combination with the relationship between temporal backscatter development and biomass water content, may enhance sugarcane drought monitoring.

NDVI signals saturate at a later stage (approximately at 80 tons/ha, 240 cm stalk height), and also afterward show a signal decrease. This coincides with the decrease in leaf biomass (see Figure 5) and ongoing senescence. ALOS-2 cross-polarized and copolarized signals did not reveal clear saturation points and showed longer sensitivity to sugarcane growth, which can be ascribed to the longer wavelength [6]. Additionally, the copolarized signals (HH) suggest there was also an effect of rain on backscatter intensities.

5.2. Intrasensor Imagery-Consistency Analysis

The time windows during which image features were most consistent differ between the various types of sensors. In addition, image consistencies were found to be generally lower for SAR sensors than for the optical sensors, with lowest values for S1 imagery and a relatively high value for ALOS-2. For the C-band SAR sensors, temporal averaging of the consistency metric allowed for the identification of these time windows. These were, as may be expected, comparable between S1 and RS2 imagery. A minor window that was identified coincided with the period around the saturation point. The general and principal window spans from the late grand growth phase until the maturing phase, after which the ripening phase commenced and images started to decorrelate. As was discussed in the previous section, during this time frame the C-band SAR signals showed, on average, lower backscatter values than at saturation point. For both sensors, images that were not affected by rain showed higher consistencies than when affected by rain. Consequently, it may be deduced that rain induces fluctuations in signal backscatter causing increased disorder in spatial patterns. For RS2 specifically, image pairs that were acquired under different beam modes should be avoided. Furthermore, one should be aware of the time of acquisition, as it may increase the chance of surface wetness (e.g., caused by morning dew).

For ALOS-2, it was found that the spatial patterns from two images taken consecutively within one revisit, not affected by rain and during the grand growth phase, showed the highest temporal agreement, comparable to values that were found for the optical sensors. Regarding LS8, the most effective time window spanned from the late grand growth phase until harvest. Finally, consistency between the two WV2 images was similar to what was observed for the LS8 images acquired during the same growth phase.

From a combined view on the time windows for the C-band and optical sensors, it may be noticed that the highest consistencies generally occurred during the final growth stages. Based on the descriptive and mathematical definition of the term consistency given in Section 3.2, specifically on the effect of pattern distinctiveness, the higher values were likely caused by increased spatial-biomass differences over the course of sugarcane growth, which is supported by the measurements in Figure 4 and by the observed distinct patterns in the averaged *z-score* images (see Figure 17).

5.3. Intersensor Imagery-Consistency Analysis

The comparisons between the most consistent patterns extracted for each sensor showed that highest agreements are found between the optical sensors. Local intrafield roads and the regional patterns in the field are especially common. For C-band SAR, agreements were generally observed for regional patterns. Spatial agreements and differences between the optical and SAR images should especially be further investigated. Special attention should be paid to the suitability of comparing the averaged *z-score* patterns when time windows differ between the sensors.

5.4. Intrafield Variability Analysis

Despite the limitations of the ground measurements, the results indicated that high-resolution optical-satellite imagery has the most potential for mapping intrafield sugarcane productivity differences.

From an integrated perspective on previous findings, it is recommended to focus on the relationship between the physical properties of the signals and the extracted patterns. Since patterns from optical

imagery may be linked to canopy conditions, and patterns from SAR imagery may be linked to volume and soil-moisture conditions, their integration may indicate fertility and moisture issues. The different sensitivities may also be used for the detection of field anomalies, including gaps and plant substitution (such as weed infestations). This should be supported by more detailed ground measurements, specifically on soil moisture, plant moisture, and sugarcane biomass. It is recommended to focus on taking these measurements during the overlap between the specified time windows, from the late grand growth phase until the early maturing phase.

6. Conclusions

In conclusion, for sugarcane productivity mapping, satellite imagery offers several opportunities. For monitoring field-averaged sugarcane-biomass growth, signals from both SAR and optical sensors can be used effectively during certain growth stages. C-band SAR imagery generally offers smallest time frames for growth monitoring, especially when the signals are affected by precipitation. Adversely, for L-band SAR imagery, saturation effects were not observed; hence, it offers growth-monitoring potential during the entire growth period. Nevertheless, L-band signals also appeared to have been affected by precipitation events. It is recommended that the precipitation sensitivity of both cross-polarized and copolarized signals from C-band and L-band sensors should be further explored for the detection of these events and therefore for drought monitoring as well. Optical imagery offers effective time frames that range between those from C-band and L-band imagery.

For monitoring intrafield sugarcane-biomass differences, the presented analyses gave an indication of the selection of images to investigate. Local and regional patterns can be obtained through temporally averaging spatial features within these windows. Especially for the SAR imagery, this reduces the noisy appearances of features and facilitates analysis for physical interpretation. Outside of these time windows, high variations of spatial-feature locations in time complicate this assessment. Comparing the resulting patterns between sensors shows that the highest agreements were found between the C-band SAR sensors and between the optical sensors. Furthermore, the intersensor comparison revealed that, in general, pattern agreements are found on a regional scale and differences on a local scale. As such, integration between sensors can be used for obtaining additional indications of regional productivity differences that may be related to water and fertility concerns. The mapping of local differences may serve as indications for local variations in field quality that may be related to plant failure and substitution (such as weed infestations). The collected ground measurements were too limited for proper validation of these specific potentials. Nevertheless, the distinctive interaction mechanisms between remote sensing signals and observed features may be exploited. For example, the sensitivity of SAR signals to water content and volumetric plant properties could be used for the detection of plant substitution where optical signals could be less successful due to the limited sensitivity to plant geometry. Several limitations associated to satellite imagery should be respected during these analyses, in particular, differences in time windows between the sensors, signal noise (especially for SAR imagery), and spatial resolution.

Finally, through relating patterns to the collected ground measurements, it was found that high-resolution optical images are most effective for providing indications on intrafield biomass variations. When similar ground campaigns are henceforth undertaken, it is recommended to acquire spatially dense plant and moisture measurements, including plant failure and plant substitution, on the scale of the SAR sensors' spatial resolutions from the late grand growth phase until the early maturing phase.

Author Contributions: R.A.M. conceived of and designed the study, designed and carried out the ground-measurement campaign, designed the methodologies, analyzed and interpreted the data and results, and wrote the manuscript. L.I. assisted in designing and carrying out the ground-measurement campaign, framing the statistical analyses, interpreting the data and results, reformulated parts of the manuscript, and reviewed the manuscript. J.V.R. established a relation and an agreement with the owner of the sugarcane fields (together with Rubens Lamparelli), assisted in validating and approving the results, and reviewed the manuscript. R.F.H. initiated the overall project and funding, endorsed the design, and approved the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to acknowledge the European Space Agency (ESA) for providing the Radarsat-2 data under the framework of project C1P.16849, the Google Earth Engine team for providing access to the Sentinel-1 and Landsat-8 data, the Japan Aerospace Exploration Agency (JAXA) for providing the ALOS-2 data under Research Announcement-6, number 149250, and the Brazilian Space Agency (INPE) for providing the Canasat maps of the São Paulo region. In addition, we would like to thank the colleagues at FEAGRI, Unicamp, for their assistance with taking the ground measurements, in particular, Diego della Justina, Carlos Wachholz de Souza, Walter Rossi Cervi, Rubens Lamparelli, and Ali Mousivand. Finally, we acknowledge and are grateful for the access to and safety in the sugarcane fields provided by Jefferson Rodrigo Batista de Mello, Eduardo Caetano Ceará, Pedro Lian Barbieri, and Alex Thiele Paulino. The work was carried out in BE-Basic project FES0905, and partly carried out within the framework of the joint BE-Basic FAPESP project 2013/50942-2.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Food and Agriculture Organization of the United Nations. *Food and Agriculture Data (FAOSTAT)*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2017.
- Carpio, L.G.T.; Simone de Souza, F. Optimal allocation of sugarcane bagasse for producing bioelectricity and second generation ethanol in Brazil: Scenarios of cost reductions. *Renew. Energy* **2017**, *111*, 771–780. [CrossRef]
- Friedrich Theodor Rudorff, B.; Alves Aguiar, D.; Adami, M.; Pereira Galvao Salgado, M. Frost damage detection in sugarcane crop using MODIS images and SRTM data. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 5709–5712. [CrossRef]
- INPE. *Canasat*; Instituto Nacional de Pesquisas Espaciais: São Paulo, Brazil, 2017. Available online: <http://www.dsr.inpe.br/laf/canasat/en/crop.html> (accessed on 1 December 2017).
- Vieira, M.A.; Formaggio, A.R.; Rennó, C.D.; Atzberger, C.; Aguiar, D.A.; Mello, M.P. Object Based Image Analysis and Data Mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. *Remote Sens. Environ.* **2012**, *123*, 553–562. [CrossRef]
- Baghdadi, N.; Boyer, N.; Todoroff, P.; El Hajj, M.; Bégué, A. Potential of SAR sensors TerraSAR-X, ASAR/ENVISAT and PALSAR/ALOS for monitoring sugarcane crops on Reunion Island. *Remote Sens. Environ.* **2009**, *113*, 1724–1738. [CrossRef]
- Wiedenfeld, R. Water stress during different sugarcane growth periods on yield and response to N fertilization. *Agric. Water Manag.* **2000**, *43*, 173–182. [CrossRef]
- NaanDanJain Irrigation Ltd. *Sugar Cane*; Technical Report; NaanDanJain Irrigation Ltd.: Post Naan, Israel, 2013.
- De Souza, C.H.W.; Lamparelli, R.A.C.; Rocha, J.V.; Magalhães, P.S.G. Height estimation of sugarcane using an unmanned aerial system (UAS) based on structure from motion (SfM) point clouds. *Int. J. Remote Sens.* **2017**, *38*, 2218–2230. [CrossRef]
- Rudorff, B.; Batista, G. Yield estimation of sugarcane based on agrometeorological-spectral models. *Remote Sens. Environ.* **1990**, *192*, 183–192. [CrossRef]
- Picoli, M.C.A.; Lamparelli, R.A.C.; Sano, E.E.; Rocha, J.V. The use of ALOS/PALSAR data for estimating sugarcane productivity. *Eng. Agrícola* **2014**, *34*, 1245–1255. [CrossRef]
- Moran, M.S.; Alonso, L.; Moreno, J.F.; Cendrero Mateo, M.P.; de la Cruz, D.F.; Montoro, A. A RADARSAT-2 Quad-Polarized Time Series for Monitoring Crop and Soil Conditions in Barrax, Spain. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1057–1070. [CrossRef]
- McNairn, H.; Brisco, B. The application of C-band polarimetric SAR for agriculture: A review. *Can. J. Remote Sens.* **2004**, *30*, 525–542. [CrossRef]

14. Steele-Dunne, S.C.; McNairn, H.; Monsivais-Huertero, A.; Judge, J.; Liu, P.W.; Papathanassiou, K. Radar Remote Sensing of Agricultural Canopies: A Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 2249–2273. [\[CrossRef\]](#)
15. Johnson, R.; Viator, R.; Veremis, J.; Richard, E.; Zimba, P. Discrimination of sugarcane varieties with pigment profiles and high resolution, hyperspectral leaf reflectance data. *J. Assoc. Sugar Cane Technol.* **2008**, *28*, 63–75.
16. Simões, M.D.S.; Rocha, J.V.; Lamparelli, R.A.C. Orbital spectral variables, growth analysis and sugarcane yield. *Sci. Agric.* **2009**, *66*, 451–461. [\[CrossRef\]](#)
17. Simões, M.D.S.; Rocha, J.V.; Lamparelli, R.A.C. Growth indices and productivity in sugarcane. *Sci. Agric.* **2005**, *62*, 23–30. [\[CrossRef\]](#)
18. Duveiller, G.; López-Lozano, R.; Baruth, B. Enhanced Processing of 1-km Spatial Resolution fAPAR Time Series for Sugarcane Yield Forecasting and Monitoring. *Remote Sens.* **2013**, *5*, 1091–1116. [\[CrossRef\]](#)
19. Simões, M.D.S.; Rocha, J.V.; Lamparelli, R.A.C. Spectral variables, growth analysis and yield of sugarcane. *Sci. Agric.* **2005**, *62*, 199–207. [\[CrossRef\]](#)
20. Molijn, R.A.; Iannini, L.; Hanssen, R.F.; Rocha, J.V. Sugarcane growth monitoring through spatial cluster and temporal trend analysis of radar and optical remote sensing images. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016, pp. 7141–7144. [\[CrossRef\]](#)
21. Baghdadi, N.; Cresson, R.; Todoroff, P.; Moinet, S. Multitemporal observations of sugarcane by TerraSAR-X images. *Sensors* **2010**, *10*, 8899–8919. [\[CrossRef\]](#)
22. Picoli, M.C.A.; Lamparelli, R.A.C.; Sano, E.E.; Rocha, J.V. Imagens multipolarizadas do sensor Palsar/Alos na discriminação das fases fenológicas da cana-de-açúcar. *Pesqui. Agropecuária Bras.* **2012**, *47*, 1307–1316. [\[CrossRef\]](#)
23. Picoli, M.C.A.; Lamparelli, R.A.C.; Sano, E.E.; De Mellod, J.R.B.; Rocha, J.V. Effect of sugarcane-planting row directions on ALOS/PALSAR satellite images. *GISci. Remote Sens.* **2013**, *50*, 349–357. [\[CrossRef\]](#)
24. Molijn, R.A.; Iannini, L.; Rocha, J.V.; Hanssen, R.F. Detailed ground reference data for sugarcane biomass estimation in São Paulo state, Brazil. *Sci. Data* **2018**, *5*, 1–18. [\[CrossRef\]](#)
25. Molijn, R. *Sugarcane Ground Reference Data over Four Fields in São Paulo State, Brazil, Taken in 2014 and 2015*; Delft University of Technology: Delft, The Netherlands, 2015. [\[CrossRef\]](#)
26. Kölln, O.T.; de Castro Gava, G.J.; Cantarella, H.; Franco, H.C.J.; Uribe, R.A.M.; da Rocha Pannuti, L.E.; Trivelin, P.C.O. Fertigated Sugarcane Yield and Carbon Isotope Discrimination ($\Delta^{13}C$) Related to Nitrogen Nutrition. *Sugar Tech* **2016**, *18*, 391–400. [\[CrossRef\]](#)
27. Mariano, E.; Leite, J.M.; Vieira-Megda, M.X.; Ciampitti, I.A.; Vitti, A.C.; Faroni, C.E.; Franco, H.C.J.; Trivelin, P.C.O. Biomass and nutrient content by sugarcane as affected by fertilizer nitrogen sources. *Crop Sci.* **2016**, *56*, 1234–1244. [\[CrossRef\]](#)
28. Franco, H.C.J.; Otto, R.; Faroni, C.E.; Vitti, A.C.; Almeida de Oliveira, E.C.; Trivelin, P.C.O. Nitrogen in sugarcane derived from fertilizer under Brazilian field conditions. *Field Crops Res.* **2011**, *121*, 29–41. [\[CrossRef\]](#)
29. Nawi, N.M.; Chen, G.; Jensen, T. In-field measurement and sampling technologies for monitoring quality in the sugarcane industry: A review. *Precis. Agric.* **2014**, *15*, 684–703. [\[CrossRef\]](#)
30. MDA. *Radarsat-2 Product Description*; Technical Report; MacDonald, Dettwiler and Associates: Richmond, BC, Canada, 2014.
31. Collecte Localisation Satellites (CLS). *Sentinel-1 Product Definition*; Technical Report; European Space Agency (ESA): Paris, France, 2016.
32. ESA. *ESA's Radar Observatory Mission for GMES Operational Services*; Technical Report sp-1322/1; European Space Agency: Noordwijk, The Netherlands, 2012.
33. JAXA. *PALSAR-2 Level 1.1/2.1/1.5/3.1 CEOS SAR Product Format Description*; Technical Report; Japan Aerospace Exploration Agency: Tsukuba, Japan, 2016.
34. USGS. *Landsat 8 (L8) Data Users Handbook*; Technical Report; United States Geological Survey: Sioux Falls, SD, USA, 2016.
35. DigitalGlobe. *World View-2 Design and Specifications—Data Sheet*; Technical Report; DigitalGlobe: Westminster, CO, USA, 2012.

36. DigitalGlobe. *World View-3 Design and Specifications—Data Sheet*; Technical Report; DigitalGlobe: Westminster, CO, USA, 2013.
37. Attema, E.P.W.; Ulaby, F.T. Vegetation modeled as a water cloud. *Radio Sci.* **1978**, *13*, 357–364. [\[CrossRef\]](#)
38. Rudorff, B.F.T.; de Aguiar, D.A.; da Silva, W.F.; Sugawara, L.M.; Adami, M.; Moreira, M.A. Studies on the Rapid Expansion of Sugarcane for Ethanol Production in São Paulo State (Brazil) Using Landsat Data. *Remote Sens.* **2010**, *2*, 1057–1076. [\[CrossRef\]](#)
39. JAXA. *ALOS-2/PALSAR-2 Calibration and Validation Results*; Technical Report; Japan Aerospace Exploration Agency: Tsukuba, Japan, 2016.
40. El Hajj, M.; Baghdadi, N.; Zribi, M.; Angelliaume, S. Analysis of Sentinel-1 radiometric stability and quality for land surface applications. *Remote Sens.* **2016**, *8*, 406. [\[CrossRef\]](#)
41. Akdim, N.; Alfieri, S.M.; Habib, A.; Choukri, A.; Cheruiyot, E.; Labbassi, K.; Menenti, M. Monitoring of irrigation schemes by remote sensing: Phenology versus retrieval of biophysical variables. *Remote Sens.* **2014**, *6*, 5815–5851. [\[CrossRef\]](#)
42. Bhandari, S.; Phinn, S.; Gill, T. Assessing viewing and illumination geometry effects on the MODIS vegetation index (MOD13Q1) time series: Implications for monitoring phenology and disturbances in forest communities in Queensland, Australia. *Int. J. Remote Sens.* **2011**, *32*, 7513–7538. [\[CrossRef\]](#)
43. Ponzone, F.J.; da Silva, C.B.; dos Santos, S.B.; Montanher, O.C.; dos Santos, T.B. Local illumination influence on vegetation indices and plant area index (PAI) relationships. *Remote Sens.* **2014**, *6*, 6266–6282. [\[CrossRef\]](#)
44. Williams, D.; Wang, Y.; Chabot, M.; Dantec, P.L.; Wu, Y.; James, K.; Thompson, A.; Vigneron, C. *Image Quality and Calibration Update*; Technical Report; MDA Corporation: Richmond, BC, Canada, 2015.
45. Hillman, A.; Rolland, P.; Périard, R.; Luscombe, A.; Chabot, M.; Chen, C.; Martens, N. RADARSAT-2 continuing system operations and performance. In Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium, Honolulu, HI, USA, 25–30 July 2010; pp. 3228–3231.
46. Schubert, A.; Miranda, N.; Geudtner, D.; Small, D. Sentinel-1A/B Combined Product Geolocation Accuracy. *Remote Sens.* **2017**, *9*, 607. [\[CrossRef\]](#)
47. Shimada, M. Advanced Land Observing Satellite-2 (ALOS-2) and PALSAR-2—Performance and the calibration. In Proceedings of the 3rd Advanced Course on Radar Polarimetry, Frascati, Italy, 19–23 January 2015; JAXA: Frascati, Italy, 2015.
48. Storey, J.; Choate, M.; Lee, K. Landsat 8 operational land imager on-orbit geometric calibration and performance. *Remote Sens.* **2014**, *6*, 11127–11152. [\[CrossRef\]](#)
49. Markham, B.; Storey, J.; Morfitt, R. *Landsat-8 Sensor Performance and Data Quality Update*; Technical Report; NASA: Pasadena, CA, USA, 2013.
50. Bresnahan, P.C. *Geolocation Accuracy Evaluations of WorldView-1 and WorldView-2*; Civil Commercial Imagery Evaluation Workshop; National Geospatial-Intelligence Agency: Boulder, CO, USA, 2011.
51. Bresnahan, P.; Brown, E.; Henryvazquez, L. *WorldView-3 Absolute Geolocation Accuracy Evaluation*; Joint Agency Commercial Imagery Evaluation Workshop; National Geospatial-Intelligence Agency: Tampa, FL, USA, 2015.
52. O’Loughlin, F.E.; Paiva, R.C.D.; Durand, M.; Alsdorf, D.E.; Bates, P.D. A multi-sensor approach towards a global vegetation corrected SRTM DEM product. *Remote Sens. Environ.* **2016**, *182*, 49–59. [\[CrossRef\]](#)
53. Zhogolev, A.; Savin, I. The influence correction of boreal forest vegetation on SRTM data. *Geocarto Int.* **2016**, *6049*, 1–34. [\[CrossRef\]](#)
54. Molijn, R.A.; Iannini, L.; Dekker, P.L.; Magalhães, P.S.; Hanssen, R.F. Vegetation characterization through the use of precipitation-affected SAR signals. *Remote Sens.* **2018**, *10*, 1647. [\[CrossRef\]](#)



SUGARCANE GROWTH MONITORING THROUGH SPATIAL CLUSTER AND TEMPORAL TREND ANALYSIS OF RADAR AND OPTICAL REMOTE SENSING IMAGES

R.A. Molijn¹, L. Iannini¹, R.F. Hanssen¹, J. Vieira Rocha²

¹Delft University of Technology, The Netherlands

²UniCamp, Brazil

ABSTRACT

During the 2014-2015 sugarcane growth season in São Paulo, Brazil, a considerable dataset was acquired consisting of space-based remote sensing images from radar and optical sensors, together with intensive ground measurements. In this work, images from the Sentinel-1, Radarsat-2 and Landsat-8 satellites are used to test the effectiveness of satellite-based indicators in sugarcane growth monitoring. A two-fold hypothesis testing is applied, in order to find statistically significant emerging hot spots and cold spots, both in space and time. Especially the comparison of results from the radar and optical sensors gives an insight into the difference in capability of these sensors to detect spatial and temporal patterns and trends.

Index Terms— Sugarcane, hypothesis testing, radar, optical, growth monitoring

1. INTRODUCTION

The purpose of the study is to observe the growth of the sugarcane through the interpretation of data from various space-based remote sensors. Sugarcane plays an important role in the world's food and energy market place, being the number one crop in the world in terms of production quantity (FAOSTAT, 2014). Brazil is the world's number one producer of sugarcane, which provides fuel for around 40% of the national gasoline market.

Several works were dedicated to growth analysis of sugarcane over time, mainly focusing on yield prediction using vegetation indices retrieved from optical imagery [1], [2], and the effects of sugarcane growth on SAR signals with an emphasis their saturation in time [3], [4]. These last works conclude that the intensities in the HV polarization from C-band (ASAR) signals shows clearer trends with increasing sugarcane heights than intensities in co-polarizations, saturating at one meter height (within three months after start of growth).

For the sugarcane growth season 2014/2015, four fields have been intensively monitored through collecting ground reference data [5] and remote sensing data. Through local

statistics analysis the similarities between observed clusters in remote sensing images and their temporal trends are statistically tested for significance and presented. The remote sensing data include Radarsat-2 Fine Quad, Sentinel-1 Extra Wide and Landsat-8 images.

The cluster analysis can yield an insight into the formation of hot and cold spots over time, which are in this stage of the study related to the optical retrieved NDVI and radar retrieved HV. Following the local statistics analysis of the remote sensing images, the evolution of the optical and radar signals with growing sugarcane is assessed. From the field reference data, the Leaf Area Index (LAI) is used to extract the vegetative part of the sugarcane growth season. Due to the height of the sugarcane (up to four meters when mature) and the short wavelength of C-band radar, saturation effects are expected.

2. STUDY AREA AND DATA DESCRIPTION

The map delineating the overview of the sugarcane fields study area is illustrated in Figure 1 including the fields' positions and the Radarsat-2 image footprints. Table 1 shows the characteristics of the fields.

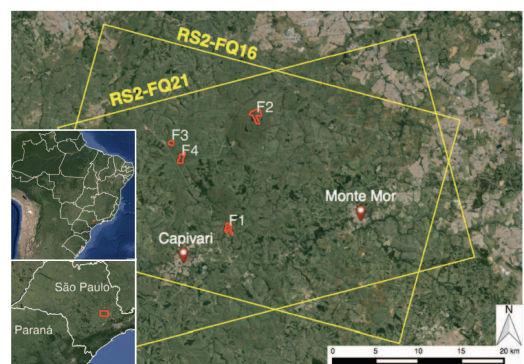


Figure 1: Overview of the study area with the Radarsat-2 (RS2) footprints used for the sugarcane analysis. The red polygons indicate the visited fields (denoted by F and the field number).

Field name	Ratoon	Area [ha]	Start of growth	Harvest
F1	1st cycle	58	30/10/2014	07/10/2015
F2	1st cycle	115	14/10/2014	07/12/2015
F3	2nd cycle	25	15/08/2014	26/07/2015
F4	9th cycle	59	01/08/2014	21/07/2015

Table 1: Sugarcane fields' characteristics with growth and harvest dates, which are for F1 and F2 not applicable to the entire field.

Figure 2 shows the remote sensing acquisitions and the field campaigns, whereby a distinction is made between elementary sampling units (ESUs) of which only biometrics are taken and elementary sampling units biomass (ESUB) of which biometrics, soil moisture and plant moistures are taken. At all instances in time LAI was taken.

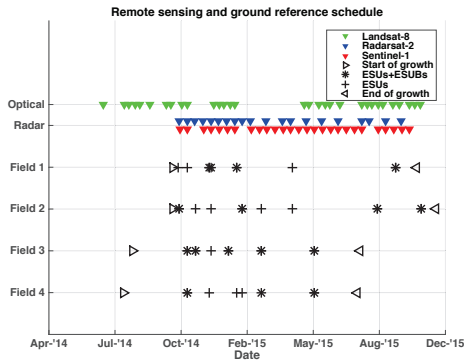


Figure 2: Remote sensing and field campaigns acquisitions scheme for the sugarcane research. 'F' denotes field, 'ESU' the Elementary Sampling Unit, 'B' Biomass.

Sensor	Mode	Bands	Resolution	Images
Radarsat-2	FQ16	Quad-pol	5.2m x 7.6m	14
Radarsat-2	FQ21	Quad-pol	5.2m x 7.6m	8
Radarsat-2	FQ21	Quad-pol	5.2m x 7.6m	8
Radarsat-2	S5	HH+HV	7.7m x 13.5m	2
Radarsat-2	S6	HH+HV	7.7m x 13.5m	4
Radarsat-2	S7	HH+HV	7.7m x 13.5m	5
Sentinel-1	Extra-wide (EW)	HH+HV	93m X 87m*	27
Landsat-8	Surface Reflectances	RGB+NIR+SWIR1/2	30m x 30m	12

Table 2: Characteristics of available remote sensing data. *Re-gridded by ESA to medium resolution product, original spatial resolution 20m x 40m.

The SAR observations, Table 2, have been averaged and gridded to the Landsat-8 cells. Figure 3 shows the temporal profiles of NDVI and HV of two of the four fields (as indicative for the other two fields as well), whereby the data gaps in the NDVI are mostly present in the rain season.

3. METHODOLOGY

Before relating the features collected in the field to the remote sensing indicators, it is important to perform a quality check on the remote sensing images in order to

determine whether the features in space and time are statistically significant or not.

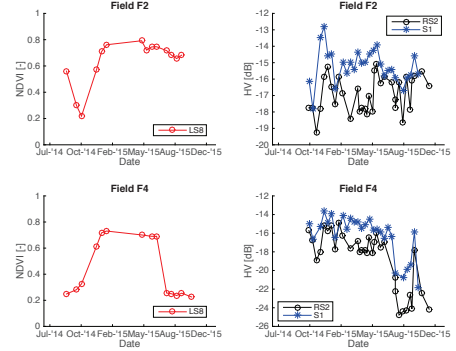


Figure 3: Time series of NDVI from Landsat-8 (LS8) and HV from Radarsat-2 (RS2) and Sentinel-1 (S1) averaged over the sugarcane fields.

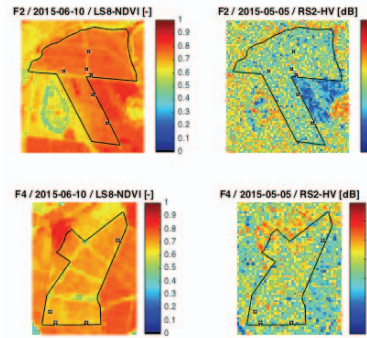


Figure 4: Example images of NDVI and HV taken of the sugarcane fields.

One of the techniques to test this significance is the Getis-Ord (G_i^*) statistic, or a weighted z-score, which is also used in several other studies that analyze remote sensing images of vegetation [6]–[8]:

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - (\sum_{j=1}^n w_{i,j})^2}{n-1}}} \quad (1)$$

Here x_j is the attribute value for feature j , $w_{i,j}$ is the weight in space between features i and j , which in this case is based in inverse distance weighting with power parameter equal to 2, and n is the total number of features. In addition,

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (2)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \quad (3)$$

The resulting statistic is a weighted z-score, which can be tested against a user-defined significance level to accept or reject the two-tailed null hypothesis of each feature being a random pattern or representing a statistically significant cluster. Significantly high and significantly low values indicate so-called hot spots and cold spots. When applied over time, emerging hot spots can be discovered. For this, the Mann-Kendall test can be applied, which has been used in various recent works that focus on vegetation trend analysis (e.g. [9], [10]) and is described in mathematical detail in [11], [12]. This nonparametric technique tests whether the null-hypothesis, H_0 : *no monotonic trend*, should be rejected and the alternative hypothesis, H_a : *monotonic trend* is present, should be accepted. When the observations are chronologically sorted (i.e. resulting in a three dimensional stack of spatial images in time), the observation vector at one point in space looks like:

$$\underline{x} = [x_1, x_2, x_3, \dots, x_{n-1}, x_n] \quad (4)$$

where n is the number of statistically significant observations. The following sign indicator function can be computed for $\frac{n(n-1)}{2}$ possible combinations:

$$sgn(x_j - x_k) = \begin{cases} 1 & \text{if } (x_j - x_k) > 0 \\ 0 & \text{if } (x_j - x_k) = 0 \\ -1 & \text{if } (x_j - x_k) < 0 \end{cases} \quad (5)$$

where $j > k$.

Subsequently, the sum of the differences, S_d , can be computed:

$$S_d = \sum_{k=1}^{n-1} \sum_{j=k+1}^n sgn(x_j - x_k) \quad (6)$$

which is an integer and comes down to the number of positive difference minus the negative differences and is positive if observations tend to be larger later in time and is negative if observations tend to be smaller later in time. In order to test whether this sum of differences is indeed significant, if $n < 10$ the value of S_d can be directly used to calculate the Kendall tau statistic, τ_s , against the critical values for Kendall's rank correlation coefficient:

$$\tau_s = \frac{2 * S_d}{n * (n - 1)} \quad (7)$$

If $n \geq 10$ the variance is [12]:

$$\sigma_d = \sqrt{\frac{1}{18} \left[n(n-1)(2n+5) - \sum_{p=1}^g t_p(t_p-1)(2t_p+5) \right]} \quad (8)$$

where g is the number of tied groups and t_p the number of ties in the p^{th} group of ties. Ties, or more correctly denoted as tied data values, is a subset of the time series consisting of the values with more than one occurrence. For example, a time series with the measurements: {0.1, 0.2, 0.1, 0.3, 0.4, 0.3, 0.5, 0.3, 0.7, 0.7} has three tied groups ($g = 3$), namely $t_1 = 2$ for tied value 0.1, $t_2 = 3$ for tied value 0.3 and $t_3 = 2$ for tied value 0.7. Since remote sensing values consist

usually not of identical numbers, the number of tied groups is expected to be zero or close to zero. In any case, with the variance computed, the standard test statistic Z_S can be applied:

$$Z_S = \begin{cases} \frac{S_d - 1}{\sigma_d} & \text{if } S_d > 0 \\ 0 & \text{if } S_d = 0 \\ \frac{S_d + 1}{\sigma_d} & \text{if } S_d < 0 \end{cases} \quad (9)$$

A positive Z_S means the data tends to increase, a negative Z_S means the data tends to decrease. Subsequently this statistic can be used in the hypothesis testing, whereby the null hypothesis (no trend) is rejected if $Z_S > Z_{\alpha/2}$ and α is the Type I error.

Based on the number of available observations, the τ_s or Z_S should be used to test the null hypothesis. For both the Gi* statistic (in space) and the Mann-Kendall statistic (in time) a significance level of 0.05 was applied.

The sum of differences is carried out on the time span between the start of growth until maximum LAI (200 days after growth), which is taken from the ground observations (which is comparable to the NDVI time series in Figure 3) of each specific field.

4. RESULTS AND DISCUSSION

The just described methodologies are applied on the SAR data of fields F1 and F2 and on the optical data of all fields. The reason for the SAR selection is the temporal coverage of the sensors during the growth of the sugarcane until 200 days, as can be viewed in Figure 3.

Table 3 shows the resulting percentage of pixels that were labeled as spatial cluster and trend (i.e. emerging hot and cold spots), as well as the percentage of pixels labeled as trend-only (i.e. not considering the cluster analysis).

Field name	Indicator	Statistically significant pixels: Cluster + trend	Statistically significant pixels: Trend only
F1	S1: HV	8%	5%
F2	S1: HV	7%	1%
F1	RS2: HV	8%	10%
F2	RS2: HV	3%	3%
F1	LS8: NDVI	0%	0%
F2	LS8: NDVI	8%	94%
F3	LS8: NDVI	15%	99%
F4	LS8: NDVI	40%	99%

Table 3: Results from the techniques per sensor and per field.

Considering only the table, Radarsat-2 shows comparable results with the Sentinel-1 results, with low percentages of significant pixels in space and time. The reason that the trend-only results can be lower than the cluster and trend results is due to the lower number of statistically significant observations, and Equation (7) would be used instead of (8).

When solely applying the trend analysis on the optical data, almost all pixels of F2, F3 and F4 show a statistically significant increasing trend (though with only four images in time), which is indeed expected considering the clear increase in NDVI as shown in Figure 3. The reason F1 does not show trends in NDVI is because only three images were useful due to cloud coverage, which is too low for the temporal statistical analysis. When also incorporating the cluster information, it shows that emerging hot and cold spots occur for 8% to 40% of the field for F2, F3 and F4.

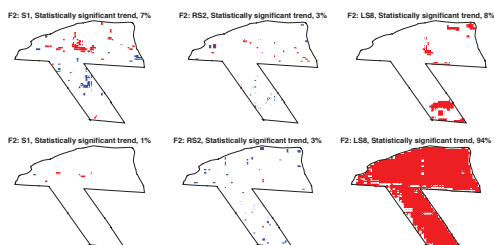


Figure 5: Emerging hot spots (red) and cold spots (blue) for S1, RS2 and LS8 in the upper graphs and positive trend-only (red) and negative-trend-only (blue) results in the lower graphs, all applied to field F2 (north is up).

Figure 5 shows the spatial results from the emerging hot spot and trend analyses. From visual inspection, it seems that the clusters from S1 and RS2 are about at the same geographical places, whereby the higher percentage of S1 could be due to its significantly lower resolution. In addition, the south parts experience in general more negative trends. The difference in saturation and interaction of optical and radar waves in vegetation are likely causes of the major differences in results, though in-depth comparison of the NDVI and HV graphs at this stage of research is too uncertain. The trend-only results of NDVI show clearly the field-wide significant increase, which is expected considering the graph in Figure 3. The positions with no trend coincide with intra-field roads (about 2 meters wide).

5. CONCLUSIONS

The results show that there are clear clusters and trends observable for the optical data, which is not the case for the SAR data. The most likely reason for this is the rapid saturation occurring in C-band SAR over tall and fast growing crops like sugarcane, in addition to the noisy character of the SAR data.

There is an important point of attention with respect to the temporal availability, which holds especially to the optical time series. These results are hence dependent on the distribution of observations in time and could have turned out differently with other distributions. One could assume this effect would not make much difference when

considering the steady increase in NDVI in Figure 3. This would be different for SAR time series with its faster saturation and noisier character, while also recognizing on the other hand that SAR in practice is temporally more equally distributed than optical. A point of consideration is that the method assigns the binary signs over time, which does not allow for gradations in the magnitude of trends.

Foreseen next steps involve the validation of the trends with respect to the already collected ground reference data.

6. ACKNOWLEDGEMENTS

The authors would like to acknowledge the European Space Agency (ESA) for providing the SAR data under the framework of project C1P.16849. The work has been carried out within the framework of the joint BE-Basic FAPESP project: 2013/50942-2.

7. REFERENCES

- [1] E. M. Abdel-Rahman and F. B. Ahmed, "The application of remote sensing techniques to sugarcane (*Saccharum spp. hybrid*) production: a review of the literature," *Int. J. Remote Sens.*, vol. 29, no. 13, pp. 3753–3767, Jul. 2008.
- [2] B. Rudorff and G. Batista, "Yield estimation of sugarcane based on agrometeorological-spectral models," *Remote Sens. Environ.*, vol. 192, no. August, pp. 183–192, 1990.
- [3] N. Baghdadi, C. Cresson, P. Todoroff, and S. Moinet, "Multitemporal observations of sugarcane by TerraSAR-X images," *Sensors (Basel)*, vol. 10, no. 10, pp. 8899–919, Jan. 2010.
- [4] N. Baghdadi, N. Boyer, P. Todoroff, M. El Hajj, and A. Bêgué, "Potential of SAR sensors TerraSAR-X, ASAR/ENVISAT and PALSAR/ALOS for monitoring sugarcane crops on Reunion Island," *Remote Sens. Environ.*, vol. 113, no. 8, pp. 1724–1738, Aug. 2009.
- [5] R. A. (Ramses); Molijn, "Sugarcane ground reference data over four fields in São Paulo state, Brazil, taken in 2014 and 2015." Delft University of Technology, 2015.
- [6] J. Barrell and J. Grant, "Detecting hot and cold spots in a seagrass landscape using local indicators of spatial association," *Landsc. Ecol.*, vol. 28, no. 10, pp. 2005–2018, 2013.
- [7] M. L. Lin, C. M. Chu, and B. W. Tsai, "Drought risk assessment in western inner-mongolia," *Int. J. Environ. Res.*, vol. 5, no. 1, pp. 139–148, 2011.
- [8] M. Wulder and B. Boots, "Local spatial autocorrelation characteristics of Landsat TM imagery of a managed forest area," *Can. J. Remote Sens.*, vol. 27, no. 1, pp. 69–77, 2001.
- [9] J. Tüshaus, O. Dubovyk, A. Khamzina, and G. Menz, "Comparison of Medium Spatial Resolution ENVISAT-MERIS and Terra-MODIS Time Series for Vegetation Decline Analysis: A Case Study in Central Asia," *Remote Sens.*, vol. 6, no. 6, pp. 5238–5256, 2014.
- [10] S. Wang, X. Chen, X. Li, G. Zhang, and T. Yang, "Trends of NDVI, precipitation and their relationship in different forest ecological zone of China during 1982 to 2006," vol. 9260, p. 92602H, 2014.
- [11] M. G. Kendall, "A New Measure of Rank Correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938.
- [12] R. M. Hirsch and J. R. Slack, "A Nonparametric Trend Test for Seasonal Data With Serial Dependence," *Water Resour. Res.*, vol. 20, no. 6, pp. 727–732, 1984.

THE EFFECTS OF SUGARCANE PRODUCTIVITY ANOMALIES ON L-BAND AND C-BAND SAR SIGNALS

Ramses A. Molijn¹, Lorenzo Iannini¹, Carlos Wachholz de Souza², Diego Della Justina², Jansle Vieira Rocha², Ramon F. Hanssen¹

1. Geoscience and Remote Sensing, Delft University of Technology, Delft, The Netherlands
2. School of Agricultural Engineering, State University of Campinas (Unicamp), Campinas, Brazil

ABSTRACT

SAR as an active remote sensing technique is capable of providing insights into the physical features of agricultural vegetation. However, the noisy nature of SAR signals makes the direct conversion to effective productivity metrics challenging. This study sheds light on the effect of gaps present in a sugarcane field on L-band and C-band SAR signals and demonstrates the variability of this effect with changing spatial averaging windows, changing precipitation conditions and changing vegetation height.

Index Terms— SAR signals, sugarcane gaps, precipitation, crop monitoring, spatial analysis

1. INTRODUCTION

The use of SAR data for mapping and monitoring vegetation dynamics has been widely researched during the last decades, but after the launch of Sentinel-1 an uptake in research activity can be observed. Especially for agricultural purposes, several review works [1, 2] provide a state-of-the-art of the techniques that utilize SAR signals to monitor crop development, to estimate crop productivity, to classify crop types and to model the interaction between the crop's physical features and the microwaves. The first review ([1]), and other works like [3], describe how crop-related dynamics retrieved by SAR data differ for different wavelengths (particularly C-band and L-band). Specifically to the sugarcane crop, due to their wavelength difference, saturation for L-band (e.g. from ALOS-2) occurs after two to three meter and for C-band (e.g. from Sentinel-1 and Radarsat-2) after 1 meter [3]. Since SAR data are, next to the geometrical features of the plant, also sensitive to the dielectric content of the illuminated medium (e.g. soil and vegetation) [4, 1], the saturation height decreases when precipitation occurs.

Sugarcane is a semi-perennial crop; after each growth cycle of 12 to 18 months, new stalks grow from the existing root system. These so-called ratoons usually last five to seven years due to declining sucrose productivity as time progresses. When new sugarcane is planted, shoots are inserted into the ground along pre-specified rows. During this

process, machine faults occur and over time sprouts default due to soil, disease, climatic conditions and harvest operations [5]. The resulting gaps in the field directly affect the yield and hence the profitability of the sugarcane field. When mature, the height of the plant is three to four meter and the leaves remain on the stalks but are subject to senescence.

2. STUDY AREA AND DATA

The study site is located in the west of São Paulo state, Brazil, near Euclides da Cunha, and surrounded mostly by sugarcane fields and pasture fields. It is 400 meter elevated above sea level, moderately sloped (<12%) and the climate is subtropical with an average temperature of 22°C and annual precipitation of 1200 mm [6]. The field of interest spans approximately 100 ha and contains sugarcane, variety RB-86-7515, with 1.5 meter row spacing. The sugarcane was planted in September 2014 and harvested in October 2015.

The gaps (Fig. 1) in the sugarcane field were estimated from UAV imagery that was acquired on November 3, 2014. At that time, the sugarcane was approximately 50 cm to one meter high. The precise methodology of the gap estimation can be found in [6]. The resulting (vectorized) gaps were compared to 54 field collected samples using tape measure. The relationship between the observed and estimated gap lengths was found to be “highly satisfactory” with a R^2 of 0.97 [6]. Based on the physical causes of gap manifestation and [5] we assume the gaps are persistent in time and hence are stationary over the considered growth period. In addition to the gap vector data, also the field boundaries were available, allowing for minimizing the inclusion of signals affected by intra-field and inter-field roads.

The SAR data consist of images acquired by three spaceborne radar sensors, all taken in cross-polarization mode, see Table 1 and Fig. 2. All radar images were radiometric calibrated, geometrically corrected, terrain corrected and georeferenced. All ALOS-2 images were taken during ascending passes with incidence angles at the field of 30°, except for the image taken on January 21, 2015, with an angle of 39°. The images of Radarsat-2 and Sentinel-1 were all acquired during

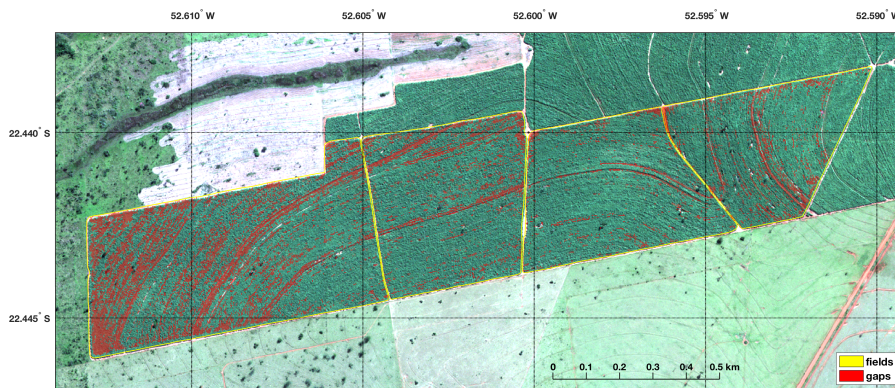


Fig. 1. Sugarcane field of interest, enclosed by yellow field boundaries and illustrating the location of the estimated sugarcane gaps. The background shows a true color, 2 meter spatial resolution WorldView-2 images taken in August 2015.

Sensor	Mode ¹	Resolution ²	Acquisition time ³	Images
ALOS-2	SM	6.0 x 4.3 m	~04:00	7
Radarsat-2	UF	1.6 x 2.8 m	08:57	2
Sentinel-1	IW	5 x 20 m	08:54	11

Table 1. Characteristics of the available SAR data. ¹SM: Stripmap, UF: Ultra-Fine, IW: Interferometric Wide; ² nominal resolution (slant range x azimuth); ³UTC, local time -2 or -3 hours.

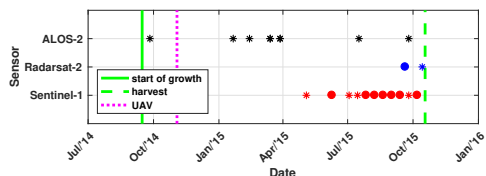


Fig. 2. Remote sensing acquisition scheme and growth dates. The asterisks denote precipitation-affected acquisitions.

descending passes and were taken at an incidence angle of 35.0° and 42.6°, respectively.

For the precipitation information, daily cumulatives of two weather stations were used, named Euclides da Cunha and Fazenda Ponte Branca, located 11 km and 7 km from the field, respectively. For flagging the SAR images as affected by precipitation or not affected by precipitation, the time of SAR image acquisition and the amount of precipitation one day before and on the day of acquisition were taken into account.

3. METHODOLOGY

First of all, the vectorized gaps were rasterized to a 1 meter by 1 meter spatial grid. The width of the gaps (i.e. equal to the row spacing of 1.5 m) is preserved by assigning gap ratios to grid cells adjacent to the those that were intersected by the gap vector. The SAR signals were nearest-neighbor interpolated (over-sampled) to the same grid. Then, for each SAR image and the gap map, spatial moving averaging with varying kernel window sizes was applied. We used window 3 meter by 3 meter (W3) based on the Radarsat-2 resolution, window 7 meter by 7 meter (W7) based on the ALOS-2 resolution and window 21 meter by 21 meter (W21) based on the Sentinel-1 resolution. For showing the effect of a larger window than the resolution for Sentinel-1, we also added window 37 meter by 37 meter (W37) and 59 by 59 meter (W59).

With the spatially averaged SAR images and gap map, we produce bivariate histograms by grouping the SAR signals and gap fractions into 2D bins. In addition, we group the SAR signals into bins of gap fractions for showing the average and standard deviations of the SAR signals per binned gap fraction. Since these statistics were computed on the signals in powers and converted back to dB, the standard deviations are not symmetric with respect to their corresponding average. The bin ranges of gap fraction for the histograms and statistics were truncated to exclude bins with low number of samples.

For each case, a linear regression was fitted and the regression coefficient, i.e. slope as backscatter change over unity gap fraction, denoted as β is extracted. The comparison of these coefficients is used show the effect of precipitation on the sensitivity of the SAR signals from gaps and to show the differences in sensitivity between sensors.

(ALOS-2) Date [yyyy-mm-dd]	DAS	Height [m]	W7	W21	W37
2014-09-26*	11	0-0.5	1.7	1.2	1.7
2015-01-21*	128	1-1.5	0.6	1.2	2.2
2015-02-13*	151	1-1.5	0.3	0.6	1.2
2015-03-14*	180	1.5-2.5	-1.6	-2.2	-3.4
2015-03-28*	194	1.5-2.5	-2.0	-2.5	-3.8
2015-07-17*	305	3-4	-1.1	-1.3	-1.9
2015-09-25*	375	3-4	-1.4	-2.1	-2.4

(RS2) Date [yyyy-mm-dd]	DAS	Height [m]	W3	W7	W37
2015-09-20	370	3-4	-0.5	-1.0	-2.0
2015-10-14*	394	3-4	0.0	-0.1	-0.9

(S1) Date [yyyy-mm-dd]	DAS	Height [m]	W21	W37	W59
2015-05-04*	231	2.5-3	-0.6	-0.4	0.1
2015-06-09	267	3-4	-0.3	-0.5	-0.4
2015-07-03*	291	3-4	-0.2	0.1	0.4
2015-07-15*	303	3-4	0.2	0.5	0.5
2015-07-27	315	3-4	-0.5	-0.7	-0.5
2015-08-08	327	3-4	-1.1	-1.5	-2.3
2015-08-20*	339	3-4	0.0	-0.1	-0.2
2015-09-01	351	3-4	-1.1	-1.4	-1.7
2015-09-13	363	3-4	-0.6	-0.8	-0.8
2015-09-25*	375	3-4	-0.3	-0.3	0.0
2015-10-07	387	3-4	-1.2	-0.9	-1.2

Table 2. ALOS-2 (top), Radarsat-2 (middle) and Sentinel-1 (bottom) regression coefficients, accompanied by the days after start of growth (DAS) and sugarcane height. The asterisks indicate the acquisitions that were affected by precipitation.

4. RESULTS AND DISCUSSION

For each sensor, several typical histograms are presented to demonstrate the difference in SAR signal sensitivity to gap fraction with varying vegetation condition, with varying windows and with varying precipitation condition. For the latter, only consecutive acquisitions within one month are included to minimize the effect of changing vegetation conditions. The entirety of the results are summarized through their regression coefficients per window and per sensor (Table 2).

Regarding changing vegetation condition, positive ALOS-2 regression coefficients are found for young cane and negative coefficients for maturing cane (top and bottom ALOS-2 plots of Fig. 3, respectively). This transition point for precipitation-affected ALOS-2 images occurs around 2 meter sugarcane height (Table 2), which can be explained by the dominant sensitivity of L-band signals to soil moisture over volume scattering until the transition point and vice versa afterwards. In other words, before 2 meter plant height, the

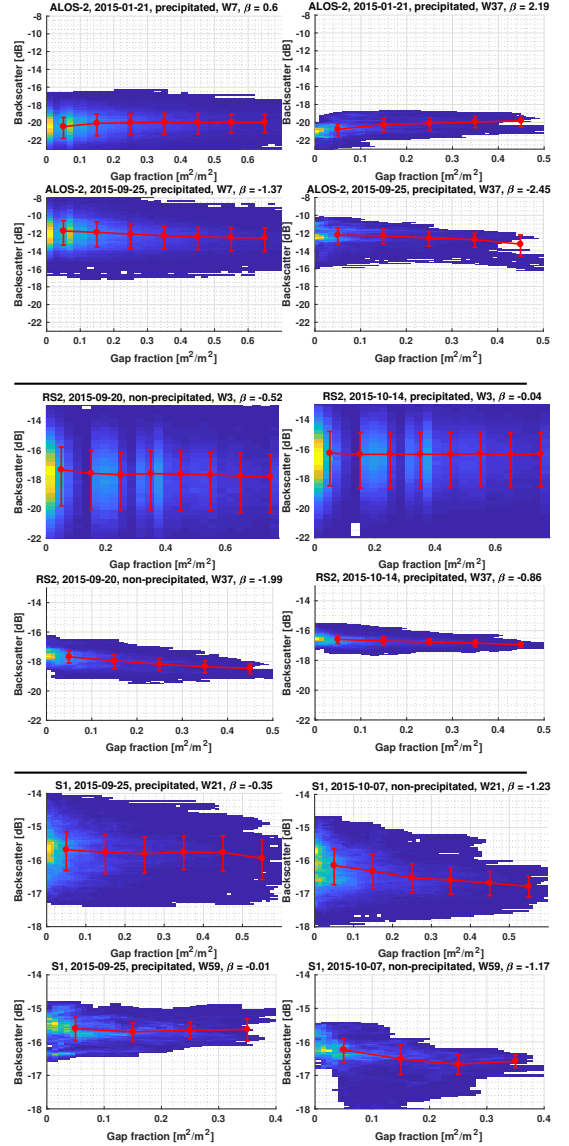


Fig. 3. ALOS-2 (top), Radarsat-2 (RS2, middle) and Sentinel-1 (S1, bottom) bivariate histograms (colored), and average and standard deviations per binned gap fraction (red markers) for different dates, windows and precipitation conditions. The regression coefficients are denoted as β .

soil moisture (which is in all cases relatively high due to the precipitation conditions) results in higher backscatter at the gaps than the soil moisture gives at the covered parts. When the volume scattering-related signal component becomes dominant over the soil moisture-related component, the backscatter is lower at the gaps than at the covered parts.

As for the effect of varying averaging windows, the Radarsat-2 and ALOS-2 results show that larger windows mainly result in smaller spread of backscatter and higher absolute regression coefficients. This can be mainly attributed to a reduction in SAR speckle noise and a smaller effect of geolocation misalignment between the SAR pixels and the gap map. In addition, the coefficients for the precipitation-affected ALOS-2 images at a sugarcane height of 3–4 meter are generally more negative than those for the precipitation-affected C-band SAR images. The likely explanation is the smaller wave penetrative depth for C-band SAR at these sugarcane heights, and hence the saturation difference.

Regarding the precipitation conditions, the results from Radarsat-2 and Sentinel-1 show that, in general, the regression coefficients of precipitation-affected images are significantly closer to zero than observed with non-precipitation-affected images. Especially for Radarsat-2, the increase in difference between coefficients over window size is apparent. This difference can be explained by the smaller difference in backscatter between the soil and the vegetation contributions when both are wet as compared to the dry condition.

Some exceptions to the mentioned observations are found for Sentinel-1 with respect to the varying windows. For example, Fig. 3 shows that the regression coefficients do not become more negative with larger averaging windows. This can be explained by the higher speckle noise caused by the sensor's lower resolution in combination with the limited size of the gap areas in the field.

5. CONCLUSIONS

The effect of gaps on SAR backscatter is mainly dependent on (1) the size of the spatial averaging window, (2) precipitation condition and (3) sugarcane height. Regarding the first, larger windows increase the absolute values of the regression coefficient and can be attributed to a reduction in the speckle noise and a reduction of the effects of geolocation misalignment between the SAR signals and the gaps. The second refers to the smaller absolute values of the coefficients when the vegetation and soil was moist due to precipitation, which can be linked to the sensitivity of the microwaves to dielectric properties. In particular, difference in backscatter between the soil and the vegetation contributions is smaller when both are wet as compared to the dry condition. The third, as was demonstrated by the ALOS-2 results, refers to positive coefficients for a sugarcane height up to two meter and negative coefficients afterwards. Finally, the effect of gaps is greater for ALOS-2 signals than for the C-band signals, which can

be ascribed to the penetrative depth of the two wavelengths. Hence, as a recommendation for future work, we suggest to study the potential of using the difference of the sign of the coefficients with varying vegetation height for actual sugarcane productivity anomaly detection using ALOS-2 signals.

6. ACKNOWLEDGEMENTS

The authors would like to acknowledge JAXA for the ALOS-2 data under the Research Announcement- 6, number 149250, ESA for the Radarsat-2 data under the framework of project CIP.16849 and the Google Earth Engine team for access to the Sentinel-1 data. In addition, we like to thank the colleagues at FEAGRI, Unicamp, for taking the ground measurements. The work was partly carried out within the framework of the joint BE-Basic FAPESP project: 2013/50942-2 and was supported by a grant from BE-Basic Foundation on behalf of the Dutch Ministry of Economic affairs.

7. REFERENCES

- [1] Susan C. Steele-Dunne, Heather McNairn, Alejandro Monsivais-Huertero, Jasmeet Judge, Pang Wei Liu, and Kostas Papathanassiou, "Radar Remote Sensing of Agricultural Canopies: A Review," 2017.
- [2] H. McNairn and B. Brisco, "The application of C-band polarimetric SAR for agriculture: a review," *Canadian Journal of Remote Sensing*, vol. 30, no. 3, pp. 525–542, jun 2004.
- [3] Nicolas Baghdadi, Nathalie Boyer, Pierre Todoroff, Mahmoud El Hajj, and Agnès Bégué, "Potential of SAR sensors TerraSAR-X, ASAR/ENVISAT and PAL-SAR/ALOS for monitoring sugarcane crops on Reunion Island," *Remote Sensing of Environment*, vol. 113, no. 8, pp. 1724–1738, aug 2009.
- [4] M. Susan Moran, Luis Alonso, Jose F. Moreno, Maria Pilar Cendrero Mateo, D. Fernando de la Cruz, and Amelia Montoro, "A RADARSAT-2 Quad-Polarized Time Series for Monitoring Crop and Soil Conditions in Barrax, Spain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1057–1070, apr 2012.
- [5] Sizuo Matsuoka and Rubismar Stolf, "Chapter 5 - Sugarcane tillering and ratooning: Key factors for a profitable cropping," *Sugarcane: Production, Cultivation and Uses*, pp. 137–157, 2012.
- [6] Carlos Henrique Wachholz de Souza, Rubens Augusto Camargo Lamparelli, Jansle Vieira Rocha, and Paulo Sergio Graziano Magalhães, "Mapping skips in sugarcane fields using object-based analysis of unmanned aerial vehicle (UAV) images," *Computers and Electronics in Agriculture*, vol. 143, no. October, pp. 49–56, 2017.

4

SPACE-BASED SAR AND OPTICAL REMOTE SENSING DATA FOR LAND COVER MONITORING

Formulierung technischer Patente war ein wahrer Segen für mich. Sie zwang zu vielseitigem Denken, bot auch wichtige Anregungen für das physikalische Denken. Endlich ist ein praktischer Beruf für Menschen meiner Art überhaupt ein Segen. Denn die akademische Laufbahn versetzt einen jungen Menschen in eine Art Zwangslage, wissenschaftliche Schriften in impressiver Menge zu produzieren — eine Verführung zur Oberflächlichkeit, der nur starke Charaktere zu widerstehen vermögen.

Albert Einstein

THIS chapter is under review at IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, submitted in 2019 [1]. The article demonstrates the performance improvement in land cover monitoring that can be achieved when integrating optical and SAR signals in a model that is based on a dynamic Bayesian network, in this case an adaptation of the Hidden Markov Model (HMM). The improvements are demonstrated as a function of image availability and as a function of time. In addition, the challenges are presented that arise when precipitation events and gaps in optical imagery occur. The ground measurement dataset was published in 2015 [2].

REFERENCES

- [1] R. A. Molijn, L. Iannini, P. L. Dekker, J. V. Rocha, and R. F. Hanssen, *Integration of space-borne SAR and optical data into a Hidden Markov Model-based approach for land cover monitoring (under review)*, (2019).
- [2] R. A. Molijn, *Land cover ground reference data in São Paulo state, Brazil, taken in 2015*, (2015), dataset, doi.org/10.4121/uuid:7b900822-4efe-42f1-9b6e-a099eda4ba02.

Integration of space-borne SAR and optical data into a Hidden Markov Model-based approach for land cover monitoring

Ramses A. Molijn, Lorenzo Iannini, Paco López Dekker, Jansle Vieira Rocha, Ramon F. Hanssen

Abstract—The timely and automated mapping of land cover from satellite is continuously subject to advances, encouraged by the increasingly dense time-series offered by radar and optical imagers. A key role in such mapping is played by the effective statistical handling of the temporal contextual information, especially in dynamic land scenarios with dominant agriculture use. Land cover tracking in such areas raises in fact the issue of heterogeneous crop timings and of their diverse cultivation patterns. In this paper, we propose the use of the Hidden Markov Model with regular sampling of the system state as an effective response to these challenges. Leveraging on the recursive nature of the model that links the up-to-date observations with the long-archived ones, the model allows to jointly account for the in-season crop growth patterns and for the inter-seasonal changes in land cover. The designed technique assimilates SAR cross-polarized backscatter and optical-derived NDVI, efficiently handling their irregular temporal availability and their gaps. We show that classification based on optical-only data generally outperforms SAR-only data, and that their integration further improves the performances. For this work, in particular, we used imagery from Sentinel-1, Radarsat-2 and Landsat and ground surveyed data over a study site in São Paulo state, Brazil.

Index Terms—SAR remote sensing, optical remote sensing, remote sensing data integration, land cover monitoring, Hidden Markov Model, sugarcane, crops, grasslands

I. INTRODUCTION

Accurate mapping of land cover and land cover changes allow for the making of environmental impact assessments, for example on biodiversity, greenhouse gas emissions, water quality and soil erosion. These are especially important in areas that experience a rapid expansion of certain land cover types over large areas. One typical example includes the mapping of land cover dynamics in São Paulo state, Brazil, where the acreage of sugarcane (an important bio-energy crop) has doubled over the last fifteen years [1], [2]. This expansion occurred mainly over grassland areas that are used as pasture fields for the grazing of cattle. The displacement of grassland to other regions (as an example of indirect land cover change) can put the preservation of natural vegetation under stress [3], [4].

Over the last two decades, advances in space-based remote sensors caused improvements in the quality of global land

cover (GLC) products, especially with respect to the spatial resolution (from tens of kilometers to tens of meters) and frequency of updates (from one single stationary map based on data over multiple years to almost near-real time maps). Currently available products are based on optical remote sensing data, featuring overall accuracies that generally range between 70-80% [5]. However, the classification of dynamic vegetation classes (in particular crops) typically performs significantly worse than the classification of relatively stationary land cover classes (e.g. natural vegetation and built-up) [6], [5]. These works report differences between GLC products up to 60%, mainly caused by confusions between crops and natural vegetation that show similar growth behavior. It was deduced that these disagreements are mainly caused by differences in types of used satellite sensors, classification methodologies, class taxonomies and ground reference data and georeferencing errors. In addition, the sensitivity of optical satellite sensors to cloud coverage in combination with the adopted methodology to cope with such data gaps can have an increasing polarizing effect on the performance differences, especially in (sub)tropical climates. In fact, the mentioned studies advocate that these criticalities make the majority of the land cover products ineffective for land cover change detection and cropland monitoring. In addition, these products are further limited by their classification taxonomy, whereby generally all crop types are grouped into one crop class. This makes them inadequate for analyzing the (in)direct land cover change that is caused by the expansion of a specific crop of interest, such as sugarcane.

The use of SAR for land cover classification, and in particular crop classification, is only successful if time series are taken into account [7]. However, the complexity of the signal caused by the complex interaction of the microwaves with the ground features make the implementation into conventional classification techniques not straightforward. Especially saturation effects, signal sensitivity to surface wetness (e.g. caused by precipitation and dew) and random speckle make the interpretation of the observations challenging. Although C-band SAR is currently widely available through the operation of Sentinel-1, in comparison to works based on optical-only data, still few works are found that describe methodologies and improvements to monitor of cropland based on SAR data. Most of the available and relevant works [8], [9], [10] use machine learning algorithms (commonly random forest) to conduct monitoring based on Sentinel-1 SAR imagery. High overall classification accuracies of more than 90% were found

Ramses A. Molijn, Lorenzo Iannini, Paco López Dekker and Ramon F. Hanssen are with the Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, The Netherlands, e-mail: r.a.molijn@tudelft.nl; l.iannini@tudelft.nl; f.lopezdekker@tudelft.nl; r.f.hanssen@tudelft.nl.

Jansle Vieira Rocha is with the Faculdade de Engenharia Agrícola (FEA-GRI), Unicamp, Campinas, e-mail: jansle@g.unicamp.br.

for single-crop mapping (rice), but substantially lower results (on average 55%) for the classification of multi-crop and natural vegetation. The work by [11] proposes a classification technique that uses the patterns of phenological stages for enhanced classification. For this, a probability based model ingesting a dense stack of multi-temporal Sentinel-1 imagery was implemented. The most probable sequence of crop stages determines the classification of the land cover. It was found that this technique on average yields a relatively high classification accuracy (more than 80%) and specifically outperforms the random forest approach for specific short-cycle crops.

When integrating optical and SAR data for land cover monitoring should cope with the frequent absence of observations (generally for optical) and a high degree of signal fluctuations (generally for SAR). Such challenges demand more complex model frameworks, capable of accommodating temporal contextual information. Among these, Hidden Markov Models (HMM) stand as a particularly convenient solution, due to their modeling flexibility and the availability of cost-effective solvers, such as the Viterbi algorithm. Hidden Markov Models became well known for their use in speech recognition [12], because of its power to relate the sequence of recognized words, instead of treating each recognized word individually, captured in a sound mathematical framework. As such, a logical sequence of words can be enforced, accepting that the probability of a recognized word based on a single observation is subordinate to the probability of that word in the context of previous recognized words. This overcomes the (partial) missing of recognized words and the corruption of the signal by noise. Similarly, the HMM can be used for land monitoring applications, where the land cover types and the phenological stages take the place of words and utterances. The authors of [13] implement for instance the HMM on optical time series for the temporal mapping of phenological states of crops (sugarcane, soybean and corn) and their discrimination from grassland and forest.

The objective of this paper is to find the differences in classification performance when integrating optical-only and SAR-only observations with respect to single-sensor observations, based on the application of the HMM for the classification of vegetation classes that commonly occur in São Paulo state. This paper should be viewed as an example of how the integration of optical and radar data can benefit classification accuracies.

II. STUDY AREA AND DATA

A. Area and ground reference data

The area of interest is located adjacent to the metropole Campinas in São Paulo state, Brazil, see Figure 1. The area is commonly referred to Brazil's heart of sugarcane due to the dominance of sugarcane plantations and the existence of sugarcane research centers and large sugar mills. The latter produce, in addition to sugar, bio-ethanol as a replacement fuel for cars.

Next to sugarcane, the area hosts agricultural crops, including mostly corn and (soy)bean. These are collectively grouped as annual crops, referring to the type of crops that

naturally complete their complete life cycle within one year. In addition, the area hosts vast ranges of grassland, which are commonly used as pasture fields for grazing cattle. The forest includes both native forests, mainly situated close to the water drainage streams, and production forests (mainly eucalyptus). The region has a humid subtropical climate with a hot and humid summer and a dry winter. This seasonality can cause vegetation to show signs of drought (i.e. lower biomass and greenness) during the winter months, especially for grasslands.

During several ground reference campaigns, conducted in 2015, more than 1000 sites were visited in this area and another close by area north of Campinas (around Mogi Guaçu). The ground reference dataset of the latter site is published and available for public use [14]. From the interpretation of Landsat and MODIS time series together with Google Earth imagery, approximately 5000 segments were digitalized. These were labeled as build-up, water, forest, grassland, sugarcane and annual crop for the years 2015 and 2016.

In addition to the location of the fields, the start and end of each crop cycle was time stamped. These dates (see Figure 2) show that the cycles are neither aligned synchronously nor uniformly spread over time. The sugarcane harvest period commences in the second half until the end of the year, before the wettest and hottest months of the summer. Typically, a sugarcane growth cycle lasts for 12 to 16 months [15]. After the harvest of sugarcane, the plants continue to grow from the same root system. After five to seven years of such 'ratoon' cycles, the roots systems are removed and replaced by new shoots. The annual crop fields host both single and double cropping systems. Single croppings usually include only summer crops, which last four to six months and are harvested around May. Afterwards, either the ground is abandoned and natural weed and grasses growth until the new cycle of summer crops, or winter crops are seeded. The latter crops are either irrigated or drought-resistant and harvested in spring. It was frequently observed in the field, that during the harvest of corn fields, the stalks were cut off just below the corn cob (roughly 0.5 meter height). The remaining (dead) stalks and natural grasses are sometimes left until the tillage of the field just before the next cropping cycle begins.

Finally, for labelling the SAR observations as affected by precipitation or not affected by precipitation, daily cumulative precipitation measurements were collected by seven weather stations in the area. This dataset covers the remote sensing observations from the beginning until June 2016.

B. Remote sensing data

The remote sensing data consist of SAR data from Radarsat-2 and Sentinel-1 and optical data from Landsat-8, see Table 1 and Figure 3. The Sentinel-1A and Sentinel-1B satellites acquire images over the study area from October 2014 onwards and from September 2016 onwards, respectively. To approximate the operational availability of Interferometric Wide SAR data provided by Sentinel-1 satellites, data from Radarsat-2 was incorporated into the classification procedure. Since the SAR satellites commonly acquire in the cross-polarization channel, only HV and VH are used. All SAR and optical

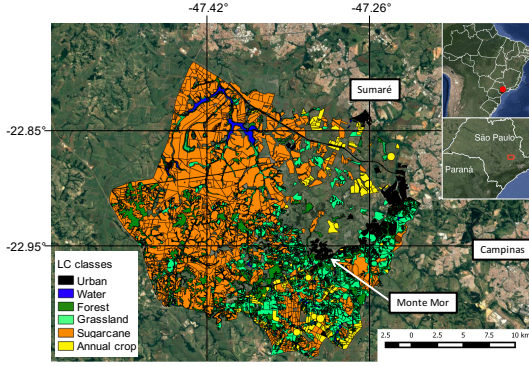


Figure 1. Study area, delineating the polygons labeled as land cover (LC) classes from ground reference data.

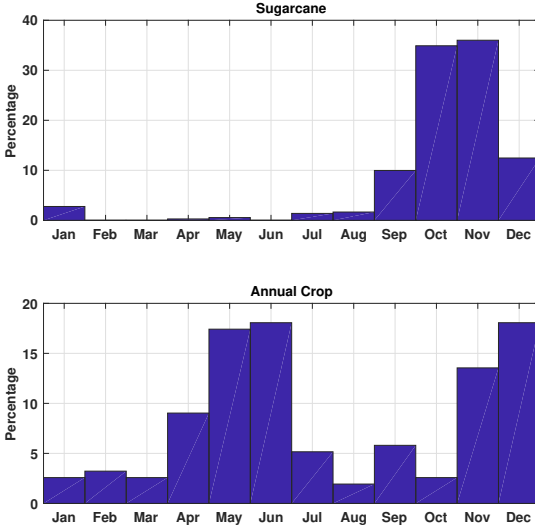


Figure 2. Temporal distribution of the start of growth for sugarcane and annual crop fields in the area of interest.

data were resampled to a common 30 by 30 meter spatial grid after multi-looking. All radar images were radiometrically calibrated, geometrically corrected, terrain corrected, thermal noise corrected and georeferenced. The optical remote sensing data consist of surface reflectances; based the near-infrared and red bands, NDVI was computed. The number of optical images in Table I and Figure 3 is based on a maximum image cloud cover percentage of 40%. The number of useful observations can differ per pixel due to remaining cloud coverage in each image.

III. METHODOLOGY

A. Hidden Markov Model framework and implementation

The following nomenclature, equations, descriptions and illustrations provide an excerpt of the essentials of the Hidden Markov Model (HMM) and are based on fundamental and applied works from [12], [13], [23], [24].

1) *Framework*: The essence of Hidden Markov Models (HMMs) lies in the modeling of causality relations in discrete data series and is based on two stochastic mechanisms (see Figure 4). The first (layer 1) regulates the underlying sequence of a pre-defined set of finite system states through transition probabilities. The second (layer 2) describes the observation probabilities belonging to each state, also referred to as emission probabilities. The actual (non-observable) sequence of states is not known by the user, attributing to the 'hidden' aspect of the model.

More formally, the model's mathematical elements and relations can be denoted as follows. The set of N individual states of the system is $\mathbf{S} = \{S_1, S_2, S_3, \dots, S_N\}$, the state at time t is recalled as q_t , the observation output at epoch t is the M -dimensional vector $\mathbf{y}_t = [y_1(t), y_2(t), y_3(t), \dots, y_M(t)]^T$. The output of the two stochastic mechanisms for a number of time steps, or epochs, T , is described by the sequences:

$$\begin{aligned} \mathbf{Y} &= \mathbf{y}_1 \mathbf{y}_2 \mathbf{y}_3 \dots \mathbf{y}_T \\ \mathbf{q} &= q_1 q_2 q_3 q_T \end{aligned} \quad (1)$$

The HMM is characterized by the model λ as a function of an element triplet, $\lambda = (A, B, \pi)$, where:

- $A = a_{i,j}$ represents the state transition probabilities - $a_{i,j}$ is the probability of the system to be in state S_j given the state of the previous epoch S_i , i.e. $a_{i,j} = P[q_t = S_j | q_{t-1} = S_i]$, where $1 \leq i, j \leq N$ and $\sum_{j=1}^N a_{i,j} = 1$;
- $B = b_j$ represents the observation output (emission) probabilities - $b_{j,t}$ is the probability at epoch t that observation vector $\mathbf{y}_t$ is observed in state S_j , i.e. $b_{j,t}(\mathbf{y}_t) = P[\mathbf{y}_t | q_t = S_j]$, where $1 \leq j \leq N$ and $1 \leq t \leq T$;
- $\pi = \pi_i$ represents the initial state probabilities - π_i is the probability of the system to be in state S_i at the first epoch, i.e. $\pi_i = P[q_1 = S_i]$, where $1 \leq i \leq N$.

The probability that one of the possible hidden sequence of states occurs is essentially a product of the three elements:

$$P(\mathbf{Y}, \mathbf{q} | \lambda) = \pi_{q_1} b_{q_1}(\mathbf{y}_1) \cdot a_{q_1 q_2} b_{q_2}(\mathbf{y}_2) \cdot a_{q_2 q_3} b_{q_3}(\mathbf{y}_3) \dots a_{q_{T-1} q_T} b_{q_T}(\mathbf{y}_T) \quad (2)$$

The sequence that is most probable can be retrieved by maximizing the likelihood of the whole vector space of sequences, $\mathbf{Q}$:

$$\hat{\mathbf{q}} = \arg \max_{\mathbf{q}} \{P(\mathbf{Y}, \mathbf{q} | \lambda)\} \quad (3)$$

For an efficient computation of this expression, the dynamic programming-based Viterbi algorithm was adopted, which is commonly used and well documented [12]. With every new observation, the model is able to recursively alter the full sequence of states when another path appears to be more probable.

Sensor ¹	Instrument	Mode ²	Polarizations/bands	Nominal resolution (slant range x azimuth) ⁷	Revisit / acquisition time ⁴	Images
RS2	C-band SAR	FQ16W (ASC)	HH+HV+VH+VV	5.2 m x 7.6 m	24 days / 21:45	14
		FQ21W (DSC)			24 days / 08:32	8
		S5 (ASC & DSC)			24 days / 21:45 & 08:32	2
		S6 (ASC)	HH+HV	13.5 m x 7.7 m	24 days / 21:49	5
		S7 (ASC)			24 days / 20:55	9
S1	C-band SAR	EW (DSC)	HH+HV	11.5 m x 43 m ⁵	12 days ⁴ / 08:38	48
		IW (DSC)	VV+VH	3.1 m x 22 m ⁶	12 days ⁴ / 08:38	10
LS8	Optical	Reflectances	RGB+NIR+SWIR1/2	30 m x 30 m	17 days / 10:10	39

Table I

CHARACTERISTICS OF AVAILABLE REMOTE SENSING DATA. ¹RS2 = RADARSAT-2, S1 = SENTINEL-1, LS8 = LANDSAT-8; ²FQ: FINE QUAD, W: WIDE, S: STANDARD, EW: EXTRA WIDE, IW: INTERFEROMETRIC WIDE, ASC: ASCENDING PASS, DSC: DESCENDING PASS; ³INCIDENCE ANGLES; ⁴REVISIT TIME IS BASED EACH SENTINEL-1 SATELLITE INDIVIDUALLY, ACQUISITION TIME IS APPROXIMATE AND IN UTC, LOCAL TIME -2/-3 HOURS;

⁵RE-GRIDDED BY ESA TO MEDIUM RESOLUTION GRD PRODUCT OF 93 M X 87 M RESOLUTION; ⁶RE-GRIDDED BY ESA TO HIGH RESOLUTION GRD PRODUCT OF 20 M X 22 M RESOLUTION; ⁷APPROXIMATE SLC PRODUCT RESOLUTIONS. REPORTED RESOLUTIONS ARE MAINLY BASED ON [16], [17], [18], [19], [20], [21], [22].

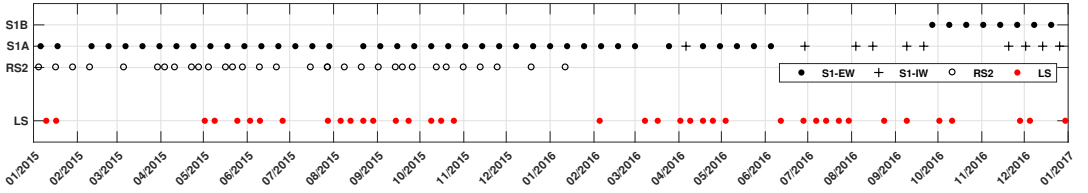


Figure 3. Remote sensing data availability, for Landsat (LS), Radarsat-2 (RS2) and Sentinel-1 satellites (S1A and S1B), separated by their interferometric wide (IW) and extra wide (EW) modes.

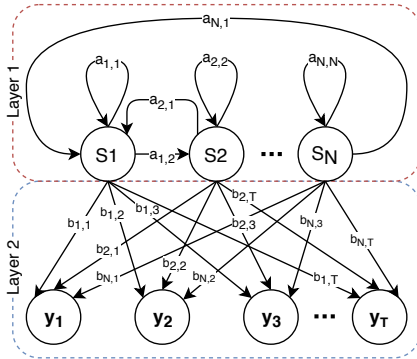


Figure 4. Basic scheme of the Hidden Markov model, including states (S_i), transition probabilities ($a_{i,j}$), emission probabilities ($b_{j,t}$) and observation variables (y_t). Adapted and modified from [13].

2) *Implementation*: The observation layer is tackled by adopting a Continuous Density HMM, that establishes continuous probability density functions for the state emission processes. The observable output of state i is modeled through the monomodal normal distribution:

$$b_{j,t} = \frac{\exp\left(-\frac{1}{2}(\mathbf{y}_t - \boldsymbol{\mu}_{j,t})^T \mathbf{C}_{j,t}^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_{j,t})\right)}{\sqrt{(2\pi)^M |\mathbf{C}_{j,t}|}} \quad (4)$$

with

$$\begin{aligned} \mathbf{y}_t &= [y^{NDVI}(t), y^{HV}(t)]^T \\ \boldsymbol{\mu}_{j,t} &= [\mu_j^{NDVI}(t), \mu_j^{HV}(t)]^T \\ \mathbf{C}_{j,t} &= \text{diag}(\nu_j^{NDVI}(t), \nu_j^{HV}(t)) \end{aligned} \quad (5)$$

where $\boldsymbol{\mu}_j(t)$ and $\mathbf{C}_j(t)$ are the mean vector and the covariance matrix, respectively. Integer M is the dimension of $\mathbf{y}_t$ and equal to two, as the selected data features representing the two sensors in this work are the NDVI index for the optical sensing and the cross-polarized backscatter, HV or VH, expressed in dB, for the radar sensing. The two variables are assumed to be uncorrelated, due a well-known difference in the sensors' sensitivity to the plant geometric and physical/dielectric configuration.

The set of N individual states of the system is the collection of the n_k states for each class with $k = 1 \dots K$. Two state-chain class models are designed, based on the temporal behavior of the classes:

- SD: state-dynamic ($n_k > 1$), reserved for the classes featuring temporally dynamic observation patterns and for which the start of the state sequence is typically asynchronous in the spatial domain
- SS: state-stationary ($n_k = 1$), reserved for the classes with regular observation patterns

The first configuration is particularly applicable to crops that experience distinct phenological developments, i.e. the from seeding and germination to ripening and harvest, which in this case are sugarcane and annual crops. The fragmentation of these crop classes into n_k states is based on the minimum

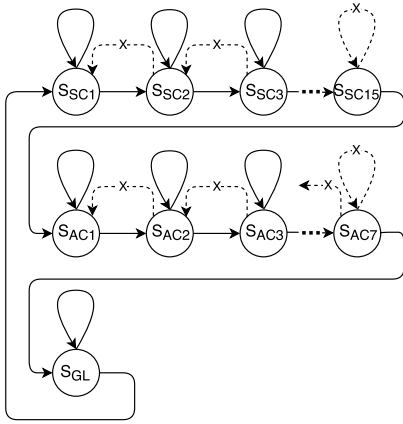


Figure 5. Scheme of an exemplary sequence, passing through all sugarcane (SC) states, followed by all annual crop (AC) states, followed by the grassland state (GL) and finally back to the first state of sugarcane. The sugarcane and annual crop classes follow the state-dynamic configuration, grassland follows the state-stationary configuration. The solid lines indicate the possible transitions, the dashed lines indicate the impossible transitions.

duration of the crop cycle divided by the revisit time of the Landsat satellites (16 days). With reference to the class descriptions given in Section II, for sugarcane $n_k = 15$ and for annual crops $n_k = 7$. The second configuration is used for the classes for which their physical development in time is simplified as a function of the day of year, hence modeling the seasonality patterns. In this study, these classes include forest and grassland.

Through the setting of the state transition probabilities, $a_{i,j}$, limitations can be enforced to the transition between states and the transition between classes. The between-states transition is of particular interest for the crop classes, for which it is biophysically impossible to go back in phenological stages as time progresses. This is also referred to as a left-right model, whereby the transitions can only go from one state to itself or to the subsequent state. In addition, the model needs to pass through all states for the observed to be labeled as that certain class before transitioning to another class.

The above mentioned features of the HMM are illustrated through an exemplary sequence in Figure 5. Here, a subset of possible transitions and impossible transitions is indicated, as well as the two types of state-chain configurations. The transition probabilities are initially based on expert knowledge and subsequently calibrated based on the sequences as observed by the HMM. Resultantly, and in accordance with the class descriptions given in Section II, the transition probability for repeating sugarcane cycles is higher than the transition probability from one sugarcane cycle to one annual crop and from one annual crop cycle to pasture. Hence, the sequence that is illustrated in Figure 5 is not likely, though not impossible.

Since the sequence of states allows for the repetition of states between epochs, flexibility is introduced in the form

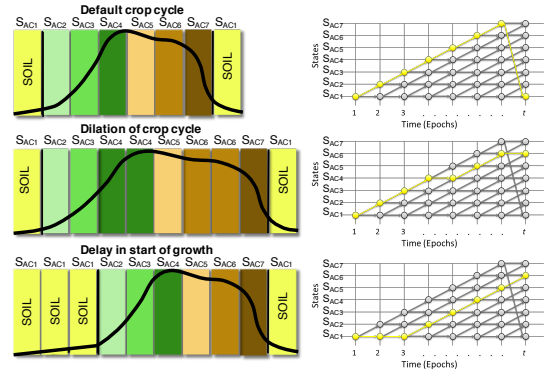


Figure 6. Flexibility in the model by accommodating the possibility of dilating the crop cycle lengths and delaying the start of growth. The crop cycle view on the left shows the order of annual crop states that are passed over the course of time. On top the indicative NDVI profiles are visualized. The trellis view on the right illustrates the Viterbi paths for these sequences.

of dilation of the crop cycle and delay in start of growth is flexible, see Figure 6. Here, the first state is related to the soil state and the last state is related to a harvest event. As was indicated in Figure 5, the state that can be associated to harvest should be passed only once for the sequence to be labeled as a finished cycle of one of the two crop classes. When a crop cycle lasts longer than the duration of the sequence of states, dilation is allowed by repetition of the same state until the last state. Since this is also allowed for the first state, a delay in start of growth is accommodated for as well.

The frequency of state estimation coincides with the observation interval Landsat acquisition scheme, i.e. 16 days. As a result, the SAR observations were resampled to the same observation scheme, based on the minimum backscatter within each interval. As such, we increase the probability that harvest events are registered by the selected SAR observations, since these are generally associated to low backscatter. In addition, when Radarsat-2 is available, these observations have priority over Sentinel-1 EW observations in order to simulate Sentinel-1 IW observations.

B. Training procedures

The estimation of the model λ consists in the estimation of its constitutive elements, i.e. the emission variables $\{\mu_{j,t}, C_{j,t}\}$ and the transition coefficients $a_{i,j}$. The training data are based on visual inspections both from on-ground observations as well as from medium and high resolution optical satellite imagery, as was explained in Section II-A. Basing on this information, the optical and radar data samples are arranged into sequences with a one-year duration. Two different criteria are adopted for SD and SS classes.

For SD classes (i.e. crop classes) the parameters of interest $\theta_k = \{(a_{k,soil}, a_{k,veg}), (\mu_{j,t}, C_{j,t})_{j \in \text{class } k}\}$ must be retrieved in presence of the hidden states q . Here, $a_{k,soil}$ and $a_{k,veg}$ are the state transition probabilities, where $a_{k,soil}$ refers to the first state of the class, whereas $a_{k,veg}$ refers to all the

other $n_k - 1$ class states. Model tying of the vegetation state probabilities has been adopted because of the limited available phenological reference data in the study area. For each class the marginal log likelihood function

$$\mathcal{L}(\theta_k) = \log P(\mathbf{Y} | \theta_k) = \sum_{\mathbf{q}} \log P(\mathbf{Y}, \mathbf{q} | \theta_k) \quad (6)$$

should be obtained by integrating the joint distribution over all possible hidden state combinations and maximized in order to extract the best candidate set of parameters. For this, we adopted a global search approach based on the well known Simulated Annealing method. The technique parameters, consisting in the initial temperature, reannealing temperature and in the maximum number of iterations, have been set heuristically by expert analysis.

In the case of SS classes, the continuous emission probabilities are readily computed through the sample mean and variance for each of the 16-days bins along the 1-year time-frame. Notice that the number of samples can be different for each bin and per sensor since the cloud gaps and the swath coverage are time- and location-dependent.

Finally, the between-class transitions regulate the resilience to class changes. In the case of SD classes, these coefficients are linked to the harvest state, $s = n_k$, when the transition is outgoing, or to the soil state, $s = 1$, when the transition is ingoing. For estimation of the transition probabilities, we use once again the yearly state and observation sequences, prepared for the training. Assume that at time t_0 , before the start of sequence $\mathbf{q}$, the true class is k_0 , and that the system initialization is exact, i.e. $\sum_{s_i \in \text{class } k_0} \pi_i = 1$. In the one-year time the HMM system is allowed to change to a new class $k \neq k_0$ if

$$LR(k_0, k) = \frac{P(\mathbf{Y} | \text{best } \mathbf{q} \subset \text{class } k)}{P(\mathbf{Y} | \text{best } \mathbf{q} \subset \text{class } k_0)} > \frac{1}{a_{k_0, k}}. \quad (7)$$

In case k is the true land cover class and Equation (7) is satisfied, we have a correct detection. Vice versa the system will register a missed detection if $LR(k_0, k) < \frac{1}{a_{k_0, k}}$. A similar analysis can be applied to the case in which no land cover change is occurring in reality. We can readily infer that Equation (7) would generate in such case a false alarm, or false detection of change. The rate of false alarms and of missed detection along a given time frame depends hence on the distribution of the likelihood ratios, and hence by the intrinsic class separability, in relation with their between-class transition coefficients. These latter can be consequently determined by setting a maximum accepted value for one of the two metrics or by searching for their joint optimization. For the problem in this work, a maximum false alarm rate of 5% on yearly sequences was selected as baseline criterion.

C. Performance assessment

The performance of the classification technique are presented based on two image availability schemes. The first is based on the manipulation of the amount of optical images, the second is based on the the real-world availability of observations as a function of time.

1) *As a function of optical image availability:* Here, the availability of optical images is simulated by introducing gaps. We selected three scenarios with 5, 10 and 15 optical images for the two years separately and averaged the results over these years. For SAR, due to its consistent availability, we use all available images and serves as a comparison for the optical results.

2) *As a function of time:* Here, the observation sequence of Figure 3 is followed. This means that in total, given the resampling to the 16-day observation interval, 39 optical images and 45 SAR images were used.

3) *Classification accuracies:* Of the ground reference dataset, 80% was used to train the classification algorithm and the complement 20% was used for validation. The performance of the classification technique is presented through confusion matrices and classification accuracies. Based on the confusion matrix, per class the following classification accuracy measures are deduced, and for the sake of clarity explained as follows:

- Producer's accuracy (PA): fraction of correctly classified samples with respect to all samples of the truth, which is directly related to the omission error, and can hence be interpreted as the accuracy from the classification map maker's perspective. For example, if 700 of the all actual 1000 sugarcane pixels are classified correctly as sugarcane, the PA is 70%.
- User's accuracy (UA): fraction of correctly classified samples with respect to all samples classified as this class, which is directly related to the commission error, and can hence be interpreted as the accuracy from the classification map user's perspective. For example, if 700 of the 2000 pixels that were classified as sugarcane are indeed sugarcane, the UA is 35%. Consequently, the remaining 1300 pixels that were classified as sugarcane are in fact other classes.

These accuracies are corrected for the differences in numbers of classified samples per class. Consequently, the overall accuracy is equal to the average of the producer's accuracies.

The general performance of the classification is expressed as the $\hat{\kappa}$ -accuracy, which compensates for the chance agreement of a random classifier:

$$\hat{\kappa} = \frac{\Pi_o - \Pi_e}{1 - \Pi_e} = \frac{\frac{1}{C} \sum_{i=1}^C r_{ii} - \frac{1}{C}}{1 - \frac{1}{C}} \quad (8)$$

Here, Π_o is overall accuracy and Π_e is the overall accuracy of correctly classifying a point purely by chance. These are computed based on the ratios in the confusion matrix, where r_{ii} is the ratio in row i and column i and the number of classes C . Hence, the $\hat{\kappa}$ -accuracy is generally always lower than the overall accuracy.

IV. RESULTS AND DISCUSSION

In this paper, two years of data were used, starting from the approximate start of the operational phase of the Sentinel-1 mission (beginning of 2015) until the end of the validity of the ground validation data (end 2016). However, since the performance generally increases with the number of observations,

these do not yet represent the maximum achievable accuracies of the technique, i.e. the full potential of the technique becomes apparent when taking into account several more years of observations. To put the potential of our adaptation of the HMM technique itself into perspective of other classifiers, based on almost 10 years of Landsat-8 data, the $\hat{\kappa}$ -accuracy reaches more than to 0.90 (overall accuracy 95%). Given the classification performances of the global land cover products mentioned in Section I, this is exceptionally high. With respect to the the earlier mentioned study [13], which also adopted the HMM for the classification of similar land cover types (overall accuracy of 90%), our results are similar.

Before presenting the classification performance accuracies, we illustrate some exemplary temporal profiles that can explain some of the most frequently occurring confusions between vegetation classes (Section IV-A). In Section IV-B and Section IV-C, the classification results based on the entire study area are presented in more detail. The first section shows the results as a function of the number of optical images. The second section shows the results that are achieved when incorporating all available images over the course of two years.

A. Exemplary temporal profiles causing confusion

The presented profiles are accompanied by field pictures that facilitate the explanation of specific behaviors of the observations. Per profile, one optical remote sensing image of the exact same field is added (in the lower left corner). The in-situ pictures were taken from similar fields in the study area and selected based on their association to the same observed phenomenon. We include profiles of annual crops, sugarcane and grassland, because between these classes the main confusions occur.

1) *Annual crop*: Generally, due to the relatively short duration of the growth cycle of annual crops and short time spans between crop cycles, optical observations suffering from cloud coverage can miss the registration of mature vegetation and harvest. In addition, SAR observations can miss the registration of bare ground between crop cycles due to precipitation events. The illustrative example of Figure 7 shows a corn field:

- A: the (approximate) harvest is not observed by optical due to cloud coverage and clearly observed by SAR.
- B: the harvest is clearly observed by optical and not clearly observed by SAR due to remaining stalks and vegetation (as was explained in Section II-A).
- C: the crop growth is not clearly observed by optical due to cloud coverage and clearly observed by SAR.
- D: the harvest is clearly observed by optical and not clearly observed by SAR due to a precipitation event.

As a result, for optical-only, the classification technique labels the pixel incorrectly as sugarcane when only taking into account the observations until June 2016. For SAR-only and for optical and SAR combined, the HMM classifies the pixel correctly as annual crop during the entire time sequence.

2) *Sugarcane*: Generally, due to the relatively long duration of the growth cycle of sugarcane, the presence of dry straw after harvest and due to precipitation events, optical and SAR observations can have difficulty with registering the

sugarcane growth trend and harvest events. With respect to annual crops, it may be expected that sugarcane is more often misclassified as grassland because sugarcane harvesting occurs less frequently. The illustrative example of Figure 8 shows a sugarcane field with severe heterogeneities in plant density, which are caused by old root systems (late ratoon cycle):

- A: the growth profile of sugarcane is observed by optical, including the senescence at the end of the cycle. The SAR observations register the presence of vegetation, but due to precipitation events and the sugarcane biomass heterogeneity in the field, the time series can be confused with a profile of grassland.
- B: the (approximate) harvest is observed by optical, but not clearly due to the relative high NDVI values. These are usually caused by weed infestation that remains after harvest of such an old ratoon field. Likewise, SAR observations do not register a harvest event, especially when affected by precipitation.
- C: the (approximate) harvest is first observed by optical, showing a clear renovation of the field with newly planted sprouts. The SAR observation taken 16 days is able to register this event.

For optical-only and for optical and SAR combined, the classification technique classifies the pixel correctly as two sugarcane cycles with the 'harvest' state at point A and C. For SAR-only, the technique classifies the pixel incorrectly as grassland during the entire time sequence.

3) *Grassland*: Since grassland fields experience a yearly cycle in plant greenness and biomass due to the climate conditions (and possibly grazing), the optical and SAR observations can lead to confusion with sugarcane. The illustrative example of Figure 9 shows a pasture field with fluctuating grass cover depending on the summer to winter conditions:

- A: the presence of healthy vegetation during summer months is clearly observed by optical and not clearly observed by SAR due to fluctuations caused by speckle and precipitation events.
- B: the field degradation, caused by plant senescence during winter months is clearly observed by optical and not clearly observed by SAR.
- C: compared to A, the increase in greenness is again clearly observed by optical, while SAR can give an indication of biomass increase.
- D: compared to B, the degradation is more clearly observed by optical and again not clearly observed by SAR.

For optical-only, the classification technique classifies the pixel incorrectly as sugarcane when only taking into account the first 6 months of observations. Afterwards, when more observations are digested, the entire sequence is correctly and continuously classified as grassland. For SAR-only and for optical and SAR combined, the entire sequence is correctly classified as grassland over the entire sequence. Occasionally, the sequence of SAR-only observations can also be confused with annual crops when dry grassland resembles a crop harvest event and a short time of wetness causes higher biomass and hence higher backscatter.

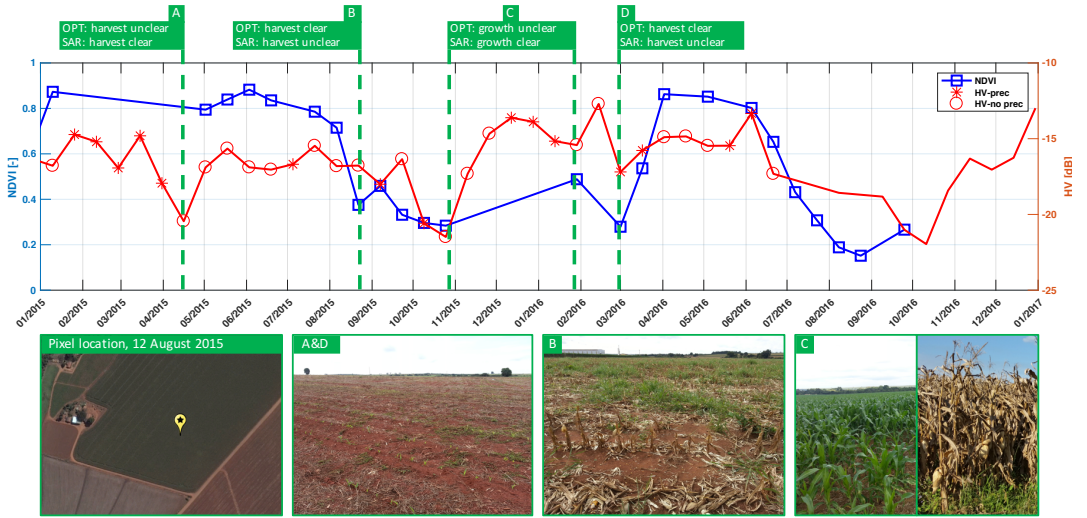


Figure 7. Optical (NDVI, left vertical axis) and SAR (HV, right vertical axis) time series over a corn field (latitude, longitude: -23.0188° , -47.2761°). The SAR observations were resampled to the optical acquisition scheme with 16-day interval. In addition, the SAR observations are differentiated between affected by precipitation (HV-prec) and not affected by precipitation (HV-no prec). The precipitation measurements are available until June 2016. The in-situ photos were taken from [14].

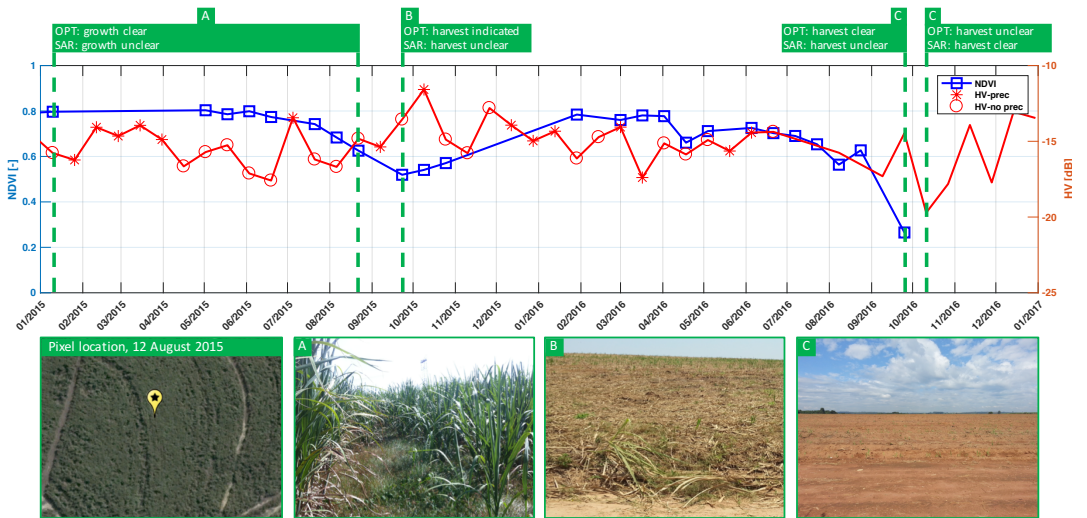


Figure 8. Optical (NDVI, left vertical axis) and SAR (HV, right vertical axis) time series over a sugarcane field (latitude, longitude: -22.9927° , -47.2834°). The SAR observations were resampled to the optical acquisition scheme with 16-day interval. In addition, the SAR observations are differentiated between affected by precipitation (HV-prec) and not affected by precipitation (HV-no prec). The precipitation measurements are available until June 2016. The in-situ photos were taken from [14].

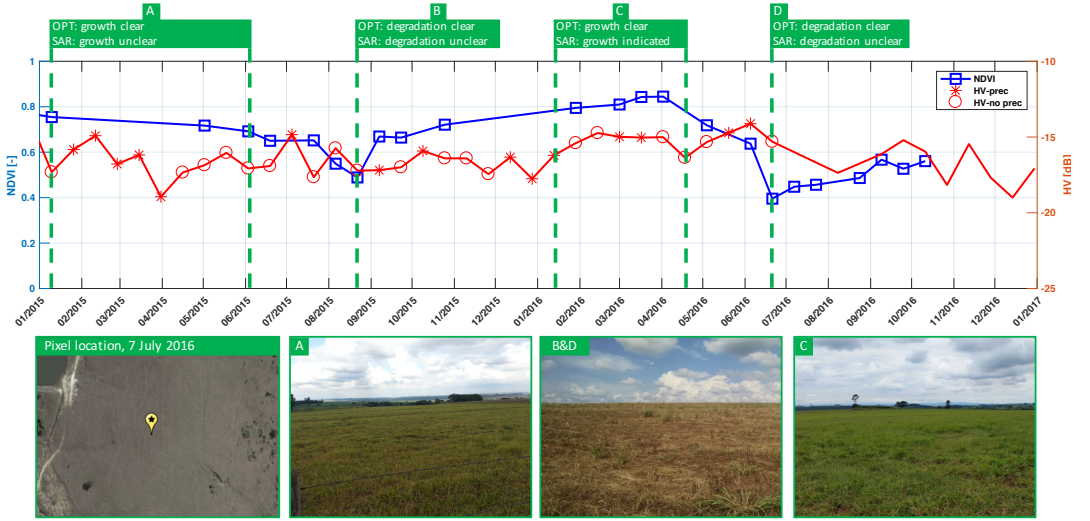


Figure 9. Optical (NDVI, left vertical axis) and SAR (HV, right vertical axis) time series over a grassland field (latitude, longitude: -22.91746° , -47.35105°). The SAR observations were resampled to the optical acquisition scheme with 16-day interval. In addition, the SAR observations are differentiated between affected by precipitation (HV-prec) and not affected by precipitation (HV-no prec). The precipitation measurements are available until June 2016. The in-situ photos were taken from [14].

B. Performances as a function of optical image availability

In this section, we simulate different availabilities of optical images, in order to show the sensitivity of the classification performance with cloud presence. As for SAR, we do continuously incorporate all available observations, as we assume they are provided regularly by the Sentinel-1 mission. The SAR-only classification performance remains constant. Furthermore, for the presentation of the accuracies, we only include the results of the two crop classes (annual crops and sugarcane) and grassland. The forest class will only be included in the confusion matrices afterwards.

The average accuracies of grassland and the two crop classes combined (see Figure 10) show that the combination of optical and SAR data always prevails over the single-sensor classification results. In addition, the optical-only accuracies are heavily dependent on the number of available images. This dependency is still existent but smaller when combining optical and SAR. Only when including the maximum availability of optical images, the optical-only kappa index approximates the SAR-only classification performance.

A detailed view on the accuracies of the individual classes (see Figure 11) reveals that the integration of both sensors results in higher performances with respect to optical-only for nearly all instances. This is a direct result from the SAR-only classification performing better than the optical-only results. Exceptions are particularly found for the PA of grassland and the UA of annual crop. The first one can be explained by the criticalities of SAR observations over grassland, see Section IV-A3. For SAR-only, grassland is more frequently classified as sugarcane and annual crop than for optical-only. As a result, the PA of grassland for the sensor combination is

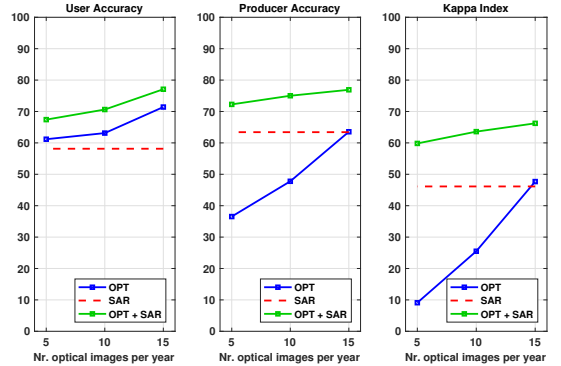


Figure 10. User accuracies, producer accuracies and kappa index versus number of available optical images, averaged over the grassland, sugarcane and annual crop classes, for optical-only (OPT), SAR-only (SAR) and optical and SAR combined (OPT + SAR). The overall accuracies are equal to the producer's accuracies. For SAR and OPT + SAR, all available SAR images were used.

deteriorated. The second exception is a result from the distinct temporal profile of annual crops (see Section IV-A1), leading to clear NDVI dynamics during the crop cycle. When optical observations are sufficiently available, the model classifies annual crops correctly. This is in contrary to SAR, where time series are more likely to be affected by fluctuations, leading to misclassifications. This results in lower UA for SAR and hence lower UA for the integration.

Additionally, the confusion matrices computed for the 5- and 15-image configurations, shown in Figure 12, convey that:

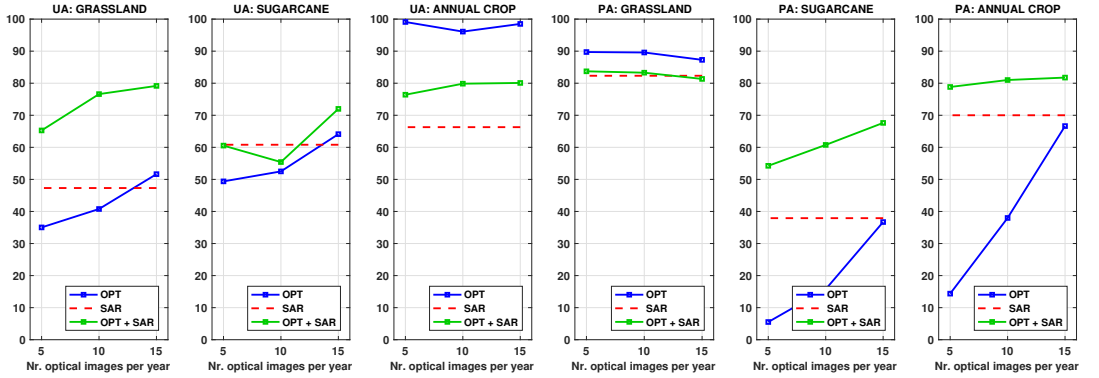


Figure 11. User accuracies (UA) and producer accuracies (PA) versus number of available optical images for optical-only (OPT), SAR-only (SAR) and optical and SAR combined (OPT + SAR). For SAR and OPT + SAR, all available SAR images were used.

- The main improvement for optical-only when incorporating 15 images instead of 5 images is the reduced misclassification of the crops classes as grassland. For SAR-only, this confusion is smaller, while the confusion between the crop classes is higher, which can be mainly ascribed to the fluctuations of the SAR observations, causing true harvest events to be missed and false harvest events to be observed. The combination of optical and SAR causes the crops-grassland confusion to be further decreased. As a result, the accuracies are generally higher, with smaller variation, despite the higher crop class confusion induced by SAR-only.
- The confusion of sugarcane with grassland indeed occurs more frequently than the confusion of annual crop with grassland, for all sensor configurations. This was earlier hypothesized in Section IV-A2 and related to the difference in harvest occurrence.
- The forest class generally performs best and most constantly with varying image availability and sensor incorporation. The only exception occurs for the SAR-only PA, where a large portion of forest pixels is classified as grassland. This is caused by saturation effects of C-band SAR, causing similar backscatter values over time.

C. Performances as a function of time

In this section, we incorporate all available optical and SAR images over the full course of two years. The average accuracies of grassland and the two crop classes (see Figure 13) show that:

- The combination of sensors performs better than the single sensors. The difference with optical-only increases over the timeframe, mainly due to the increase in SAR-only performance and the approximate stable performance of optical-only after six months.
- There are two periods during which the SAR-only accuracies, and hence the combination as well, increase rapidly; between 6 months and 1 year and between 1.5 year and 2 years. These can be attributed to the observations that

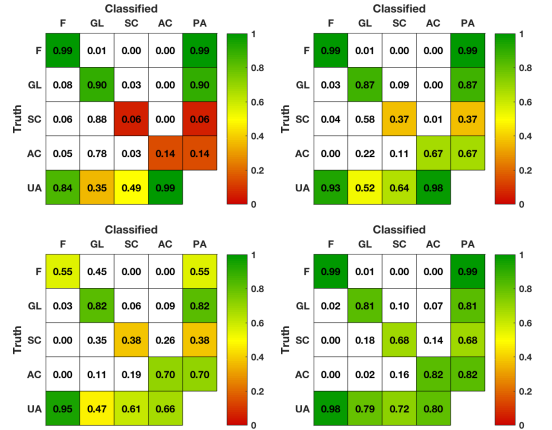


Figure 12. Confusion matrices for optical using respectively five and fifteen images (top), for SAR (bottom left) and for optical and SAR combined using fifteen optical images (bottom right). The classes refer to forest (F), grassland (GL), sugarcane (SC) and annual crop (AC). Per class combination, each ratio represents the number of pixels classified as the corresponding class ('classified') over the total number of pixels belonging to the true class (truth). In addition, the producer's accuracies (PAs) and user's accuracies (UAs) are presented.

are acquired during the harvesting period of sugarcane, which generally commences half-way the calendar year (after June) and ends at the end of the year (December), see Figure 3.

- The SAR-only accuracies reach the the optical-only accuracies after approximately two years, illustrating the superiority of optical-only observations over SAR-only observations during this time frame and image availability.

A detailed view on the confusion matrices after two years (see Figure 14) shows that:

- For optical-only, the misclassification of sugarcane and annual crop as grassland remains a challenge. As was

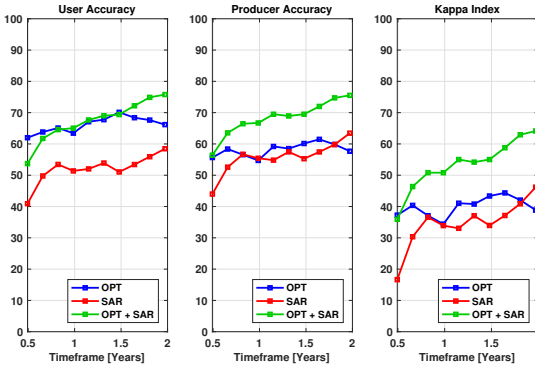


Figure 13. User accuracies, producer's accuracies and kappa index versus absolute time averaged over the grassland, sugarcane and annual crop classes, for optical-only (OPT), SAR-only (SAR) and optical and SAR combined (OPT + SAR). The overall accuracies are equal to the producer's accuracies. All available optical and SAR observations were used. The timeframe starts from January 2015 until December 2016; the first six months were omitted due to noisy behavior.

discussed in Section IV-A and Section IV-B, this is mainly caused by the missing of the growth dynamics, especially harvest events.

- For SAR-only, the main confusion occurs for sugarcane with grassland, which was mainly caused by the sparse occurrence of sugarcane harvest events and the remaining straw in combination with precipitation events. Especially during the early growth stage of sugarcane, the SAR observations resemble the observations of grasslands.
- For the combination of the sensors, the accuracies are more similar. The confusion of sugarcane with annual crops and grassland remains the most significant challenge.

V. CONCLUSIONS

This manuscript introduces a statistical framework based on Continuous Density Hidden Markov Models to assimilate the temporally irregular series from the optical and radar sensors over dynamic scenarios characterized by high crop pattern diversities. The proposed HMM adaptation is implemented with a 16-days sampling of the system state. Each 16-days epoch is in fact associated to a hidden state, or land cover state, and to its visible output, represented by the available radar and optical observations. Based on the available and on the historic observation, the whole hidden chain of states can be retrieved and recursively altered.

The work shows that the integration of optical and SAR data into the HMM classification technique introduces significant performance benefits when compared to using a single sensor, i.e. using optical-only and SAR-only observations. In addition, it was found that optical-only classification gives higher accuracies than SAR-only classification during (approximately) the first two years of mapping. However, the analysis also showed that the optical-only accuracies are heavily dependent on cloud coverage percentages. Only when 15 optical images per-year

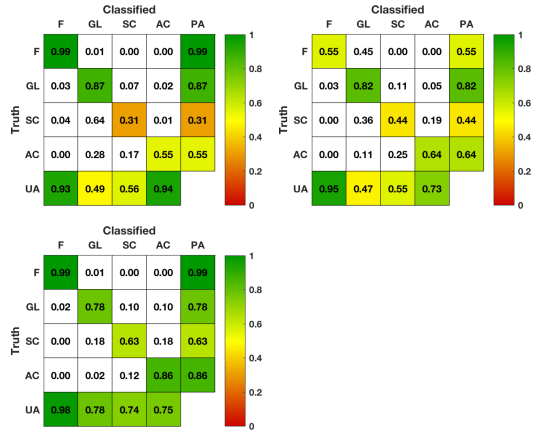


Figure 14. Confusion matrices based on the classification with 2 years of data for optical (top left), for SAR (top right) and for optical and SAR combined (bottom). The classes refer to forest (F), grassland (GL), sugarcane (SC) and annual crop (AC). Per class combination, each ratio represents the number of pixels classified as the corresponding class ('classified') over the total number of pixels belonging to the true class (truth). In addition, the producer's accuracies (PAs) and user's accuracies (UAs) are presented.

were incorporated, the optical-only system performs indeed better than the SAR-only system. This was mainly ascribed to the high confusion of annual crops and sugarcane with grassland for optical-only, which is caused by the missing of important crop cycle events (full vegetation cover and harvest). As expected, such availability factor is mitigated by the assimilation with SAR data. For SAR-only, the confusion is higher between annual crops and sugarcane, which can be related to the fluctuations of SAR observations caused by precipitation, crop residues after harvest and speckle noise. This leads to uncertainties in the registration of harvest events, which plays a key role in the performance economy. When the observations were ingested chronologically over the two year time frame, it was indeed shown that during the harvesting periods, the accuracies improve faster than outside of these periods. In this scenario, the optical-only classification results generally prevail over those from SAR-only observations, yet again outperformed by the integration of the datasets from the two sensors.

Finally, it should be noted that due to the relatively short time frame of the observations (two years), further improvements that can be achieved over longer time series has not been presented yet. In addition, we expect higher accuracies when the feature space is expanded, for example by ingesting other polarimetric features and optical indices. Nevertheless, the accuracies of the proposed integration of optical and SAR data already challenges reported accuracies of existing global land cover products. Hence, the proposed HMM technique shows promising potential for accurate land cover monitoring.

VI. ACKNOWLEDGEMENTS

The authors would like to acknowledge the European Space Agency (ESA) for providing the Radarsat-2 data under the

framework of project CIP.16849, Embrapa for access to MODIS time series through the SatVeg tool, and the Google Earth Engine team for providing access to the Sentinel-1 and Landsat-8 data. In addition, we like to thank colleagues for their assistance with taking the ground measurements, in particular Diego della Justina, Carlos Wachholz de Souza and Rubens Lamparelli from FEAGRI, Unicamp, as well as Arjan Tabak from TU Delft. Finally, we acknowledge Jefferson Rodrigo Batista de Mello, Alex Thiele Paulino and Ralph Guenther Hammer for supplying the precipitation data. The work was carried out in BE-Basic project FES0905 and partly carried out within the framework of the joint BE-Basic FAPESP project 2013/50943-9.

REFERENCES

- [1] P. G. Machado, N. A. M. Rampazo, M. C. A. Picoli, C. G. Miranda, D. G. Duft, and K. R. E. de Jesus, "Analysis of socioeconomic and environmental sensitivity of sugarcane cultivation using a Geographic Information System," *Land Use Policy*, vol. 69, no. August, pp. 64–74, 2017. [Online]. Available: <http://dx.doi.org/10.1016/j.landusepol.2017.08.039><http://linkinghub.elsevier.com/retrieve/pii/S0264837716303040>
- [2] INPE, "Canasat," 2017. [Online]. Available: <http://www.dsr.inpe.br/iaf/canasat/en/index.html>
- [3] A. Alkimim and K. C. Clarke, "Land use change and the carbon debt for sugarcane ethanol production in Brazil," *Land Use Policy*, vol. 72, no. August 2017, pp. 65–73, 2018. [Online]. Available: <https://doi.org/10.1016/j.landusepol.2017.12.039>
- [4] B. F. T. Rudorff, D. A. de Aguiar, W. F. da Silva, L. M. Sugawara, M. Adami, and M. A. Moreira, "Studies on the Rapid Expansion of Sugarcane for Ethanol Production in São Paulo State (Brazil) Using Landsat Data," *Remote Sensing*, vol. 2, no. 4, pp. 1057–1076, apr 2010. [Online]. Available: <http://www.mdpi.com/2072-4292/2/4/1057/>
- [5] A. Pérez-Hoyos, F. Rembold, H. Kerdiles, and J. Gallego, "Comparison of global land cover datasets for cropland monitoring," *Remote Sensing*, vol. 9, no. 11, 2017.
- [6] S. Fritz, L. See, I. McCallum, C. Schill, M. Obersteiner, M. Van Der Velde, H. Boettcher, P. Havlik, and F. Achard, "Highlighting continued uncertainty in global land cover maps for the user community," *Environmental Research Letters*, vol. 6, no. 4, 2011.
- [7] S. C. Steele-Dunne, H. McNairn, A. Monsivais-Huetero, J. Judge, P. W. Liu, and K. Papathanassiou, "Radar Remote Sensing of Agricultural Canopies: A Review," pp. 2249–2273, 2017.
- [8] C. Kontgis, M. S. Warren, S. W. Skillman, R. Chartrand, and D. I. Moody, "Leveraging Sentinel-1 time-series data for mapping agricultural land cover and land use in the tropics," *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, pp. 1–4, 2017. [Online]. Available: <http://ieeexplore.ieee.org/document/8035199/>
- [9] H. Balzer, B. Cole, C. Thiel, and C. Schmullius, "Mapping CORINE land cover from Sentinel-1A SAR and SRTM digital elevation model data using random forests," *Remote Sensing*, vol. 7, no. 11, pp. 14876–14898, 2015.
- [10] N. Torbick, D. Chowdhury, W. Salas, and J. Qi, "Monitoring rice agriculture across myanmar using time series Sentinel-1 assisted by Landsat-8 and PALSAR-2," *Remote Sensing*, vol. 9, no. 2, 2017.
- [11] D. Bargiel, "A new method for crop classification combining time series of radar images and crop phenology information," *Remote Sensing of Environment*, vol. 198, pp. 369–383, 2017. [Online]. Available: <http://dx.doi.org/10.1016/j.rse.2017.06.022>
- [12] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs/_all.jsp?arnumber=18626&tag=1
- [13] P. B. C. Leite, R. Q. Feitosa, A. R. Formaggio, G. A. O. P. Da Costa, K. Pakzad, and I. D. A. Sanches, "Hidden Markov Models for crop recognition in remote sensing image sequences," *Pattern Recognition Letters*, vol. 32, no. 1, pp. 19–26, 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.patrec.2010.02.008>
- [14] R. A. Molijn, "Land cover ground reference data in São Paulo state, Brazil, taken in 2015," 2015. [Online]. Available: <https://data.4tu.nl/repository/uuid:7b900822-4efe-42f1-9b6e-a099eda4ba02>
- [15] M. A. Vieira, A. R. Formaggio, C. D. Rennó, C. Atzberger, D. A. Aguiar, and M. P. Mello, "Object Based Image Analysis and Data Mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas," *Remote Sensing of Environment*, vol. 123, pp. 553–562, aug 2012.
- [16] MDA, "Radarsat-2 Product Description," MacDonald, Dettwiler and Associates, Tech. Rep., 2014.
- [17] Collecte Localisation Satellites (CLS), "Sentinel-1 Product Definition," European Space Agency (ESA), Tech. Rep., 2016. [Online]. Available: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/document-library/-/asset_publisher/1d07RF5fJMbdl/content/sentinel-1-product-definition
- [18] ESA, "ESA's radar observatory mission for GMES operational services," European Space Agency, Tech. Rep. sp-1322/1, 2012. [Online]. Available: [https://sentinel.esa.int/documents/247904/349449/S1\(_\)_SP-1322\(_\)_1.pdf](https://sentinel.esa.int/documents/247904/349449/S1(_)_SP-1322(_)_1.pdf)
- [19] JAXA, "PALSAR-2 Level 1.1/2.1/1.5/3.1 CEOS SAR Product Format Description," Japan Aerospace Exploration Agency, Tech. Rep., 2016.
- [20] USGS, "Landsat 8 (L8) Data Users Handbook," Tech. Rep., 2016.
- [21] DigitalGlobe, "World View- 2 Design and Specifications - Data sheet," DigitalGlobe, Tech. Rep., 2012.
- [22] —, "World View- 3 Design and Specifications - Data sheet," DigitalGlobe, Tech. Rep., 2013.
- [23] R. A. Molijn, L. Iannini, R. F. Hanssen, F. J. Van Leijen, R. A. C. Lamparelli, and A. Coutinho, "Integration of sar and optical dense time series for land cover monitoring," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2017-July, 2017, pp. 4330–4333.
- [24] L. Iannini, R. Molijn, A. Mousivand, R. Hanssen, and R. A. C. Lamparelli, "A HMM-based approach for historic and up-to-date land cover mapping through Landsat time-series in the state of São Paulo, Brazil," *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2016-Novem, pp. 5457–5460, 2016.

5

SPACE-BASED SAR REMOTE SENSING SIGNALS AFFECTED BY PRECIPITATION FOR VEGETATION CHARACTERIZATION

It took nearly ten years to see my dream come true. It was indeed worth the effort and the wait. [...] No longer in space and now an "earthling" once again, I feel that my good fortune yet continues. I welcome the opportunity to express my heartfelt gratitude to many people who have encouraged me to believe that if you can dream it, you can do it. And I strongly believe that education enables us to envision and to pursue our dreams.

Chiaki Mukai

THIS chapter has been published in Remote Sensing, Remote Sensing in Agriculture and Vegetation, in 2018 [1]. The article demonstrates a technique that improves vegetation classification based on changes in wetness condition of vegetation through the use of SAR-only remote sensing signals. More specifically, it focuses on precipitation conditions as a proxy for vegetation wetness and how the resulting SAR signal fluctuations can be used advantageously for improved discrimination between vegetation stages. Additionally, the research includes the further improvement in classification performance through incorporating soil type information. The findings provide new insights into use of a-priori information on wetness and soil conditions for land cover monitoring applications. The proposed techniques were not integrated with the techniques that were analyzed in the previous chapters. For the study, both the sugarcane biophysical ground reference dataset [2] as well as the land cover ground reference dataset [3] were used.

REFERENCES

- [1] R. A. Molijn, L. Iannini, P. L. Dekker, P. S. Magalhães, and R. F. Hanssen, *Vegetation characterization through the use of precipitation-affected SAR signals*, *Remote Sensing* **10**, 1 (2018).
- [2] R. A. Molijn, *Sugarcane ground reference data over four fields in São Paulo state, Brazil, taken in 2014 and 2015*, (2015), dataset, doi.org/10.4121/uuid:37112e18-f794-4d66-a8cd-7f1e92af09fc.
- [3] R. A. Molijn, *Land cover ground reference data in São Paulo state, Brazil, taken in 2015*, (2015), dataset, doi.org/10.4121/uuid:7b900822-4efe-42f1-9b6e-a099eda4ba02.

Article

Vegetation Characterization through the Use of Precipitation-Affected SAR Signals

Ramses A. Molijn^{1,*}, Lorenzo Iannini¹, Paco López Dekker¹, Paulo S.G. Magalhães²
and Ramon F. Hanssen¹

¹ Geoscience and Remote Sensing, Delft University of Technology, 2628 CN Delft, The Netherlands; L.Iannini@tudelft.nl (L.I.); F.LopezDekker@tudelft.nl (P.L.D.); R.F.Hanssen@tudelft.nl (R.F.H.)

² Faculdade de Engenharia Agrícola (FEAGRI), Unicamp, Campinas 13083-875, Brazil; graziano@g.unicamp.br

* Correspondence: R.A.Molijn@tudelft.nl

Received: 31 August 2018; Accepted: 13 October 2018; Published: 16 October 2018



Abstract: Current space-based SAR offers unique opportunities to classify vegetation types and to monitor vegetation growth due to its frequent acquisitions and its sensitivity to vegetation geometry. However, SAR signals also experience frequent temporal fluctuations caused by precipitation events, complicating the mapping and monitoring of vegetation. In this paper, we show that the influence of a priori known precipitation events on the signals can be used advantageously for the classification of vegetation conditions. For this, we exploit the change in Sentinel-1 backscatter response between consecutive acquisitions under varying wetness conditions, which we show is dependent on the state of vegetation. The performance further improves when a priori information on the soil type is taken into account.

Keywords: SAR signals; precipitation; vegetation classification; soil type; incidence angle

1. Introduction

Land cover classification offers baseline information for several land monitoring applications such as food security monitoring and early warning, water use efficiency studies or fraud detection, to cite a few. Over the last decade, the spatial and temporal resolution of land cover monitoring products underwent significant enhancements, mainly due to advances in the provision of Analysis-Ready Data (ARD) from existing satellite platforms and due to the increasing computational resources for processing large amounts of remote sensing data. Global land cover maps such as Global Food Security-Support Analysis Data at 30-m (GFSAD30) and GlobeLand30 at 30-m resolution [1–3] and Global Land Cover Project 2000 (GLC 2000), MODIS and GlobCover at 300-m to one-kilometer resolution [4–7] are available and are mainly based on optical imagery produced by the Landsat, SPOT, MODIS and MERIS missions. These land cover products commonly feature broad land cover classes and generally do not distinguish between crop types.

Although the information provided by these products can be valuable for generating regional overviews, important shortcomings were identified, as well. A study comparing the maps of GLC 2000, MODIS and GlobCover produced until 2011 [8] focused specifically on the thematic accuracies between these products. It was reported that the overall accuracies of these maps are 0.68, 0.75 and 0.67, respectively, and even higher disagreements are found for cropland and forest classes. This is critical for studies where precise spatial and temporal discrimination of vegetation classes is required. In the same study, it was stressed that the errors in the maps were higher than the detected changes, making them ineffective for land cover change detection. Although maps produced since 2011 show accuracy improvements, these were found not to be significant given the advances in satellite data

acquisition and algorithms [9]. In addition, all of the mentioned land cover products provide the land cover type, but usually lack phenological information.

The use of Synthetic Aperture Radar (SAR) for land cover classification, and in particular crop classification, is only successful when time series are taken into account [10]. Since the backscatter in a single SAR image can be practically identical for different crops, the focus must lie on the temporal backscatter differences, when vegetation structure changes, especially during the seed and development phase [10]. Recognizing that different wavelengths feature different interactions with vegetation and hence give different classification performances, for C-band SAR specifically, it was found that saturation effects, occurring generally from a vegetation height of 1–1.5 m [10–12], cause classification to be most accurate for lower biomass crops (e.g., wheat, hay, pasture) [10]. Generally, in these studies, it is mentioned that signals in cross-polarization are favored over signals in co-polarization for effective crop classification, since these are more sensitive to volume scattering, as well as to soil moisture [13].

Even though the introduction of Sentinel-1, providing free and temporally-dense C-band SAR imagery, yields a unique opportunity for incorporating these signals into the generation process of land cover monitoring products, surprisingly few works were found that promote such efforts along with descriptions on the SAR-related methodologies and consequential improvements. Limiting to publications that describe the utilization of temporal Sentinel-1 backscatter (intensity) signals for producing land cover products, the studies by [14–16] are illustrative, all applying machine learning algorithms (random forest). None of the works mention the incorporation of contextual information with daily precipitation events and incidence angle. Presumably, the applied models translate the corresponding signal variations into uncertainty in the classification results. An exception is the study by [16], which found that the effect of the incidence angle on the resulting classification performance was minimal. A work that explicitly mentioned the removal of precipitation-affected SAR signals was the study by [17], which evaluated the radiometric stability and quality of Sentinel-1. In addition, the studies by [13,18] emphasized the importance of incorporating daily irrigation and precipitation information when monitoring crop fields using SAR signals, and the latter showed that the signals are most sensitive to soil moisture when taken over bare ground as compared to full vegetation cover. However, none of the above-mentioned works proposed methods on how to treat precipitation-affected signals for improved vegetation classification.

Based on these literature findings, we underline the importance of incorporating precipitation information when performing temporal land cover classification, and in particular vegetation classification, based on SAR signals. In this paper, we focus on the variations in SAR signals that arise from precipitation events and how these could be used for advanced discrimination of vegetation conditions. The objective is to manifest the effect of these variations on the classification performance and the improvements that can be achieved when taking into account precipitation and backscatter information with temporally-consecutive SAR observations. In addition, we quantify the effect on the classification performance when taking into account information with SAR incidence angle and soil type. The aim is to propose an approach on how to handle precipitation-affected SAR signals and the advantages of incorporating such information, along with available knowledge on incidence angle and soil type, regardless of the SAR-based land cover monitoring or classification model. As such, it is not intended that the presented methodologies be used as stand-alone techniques for SAR-based vegetation characterization.

2. Study Area and Data Description

2.1. Study Area

The study area is São Paulo state, which covers nearly 250,000 km², 3% of Brazil. It is inhabited by 46 million people, 20% of Brazil's population, most of them living in the eastern part close to São Paulo city. The state territory is located in the sub-tropics and covers various climatic types, mainly

influenced by extratropical and tropical weather systems. The average monthly temperature generally reaches close to 30 °C during summer and close to 20 °C during winter. Most of the precipitation occurs during the rainy season in the summer months; from November until January, almost 40% of the annual precipitation falls. The winter months are usually dry, occasionally with no precipitation during several consecutive weeks. The state's interior receives 1000–1600 mm annual rainfall, whereas the forest strip along the coast receives more than 3000 mm annual rainfall. When precipitation occurs, the average of daily cumulative rainfall is slightly more than 10 mm. The majority of the state is elevated between 300 m and 900 m above sea level.

The forest strip along the coast consists of protected native (mainly Atlantic) forest and divides the coastline areas from the state's interior. The interior is covered by a mix of sugarcane, annual crops including mainly corn, soybean and bean, as well as citrus trees and forest crops including mainly eucalyptus and pine. The state also hosts numerous watersheds and three major river basins. An indication of the coverage of the land cover types is provided in Table 1, and their spatial distribution is illustrated by Figure 1. For more details on the origin of this dataset, see Section 2.2.3. Mid-vegetation includes mainly shrubland and citrus; grassland includes mainly pasture; and forest includes high bushes, native forests and production forests. There are two dominant soil types in São Paulo state, which are different mainly in clay content, as will be further described in Section 2.2.4.

Table 1. Coverage of land cover types in São Paulo state based on the 2015 classification map produced by Delft University of Technology (TU Delft) [19,20].

Land Cover	Coverage
Sugarcane	19%
Annual crop	4%
Grassland	33%
Mid-vegetation	15%
Native forest	23%
Production forest	1%
Water	3%
Built-up	2%

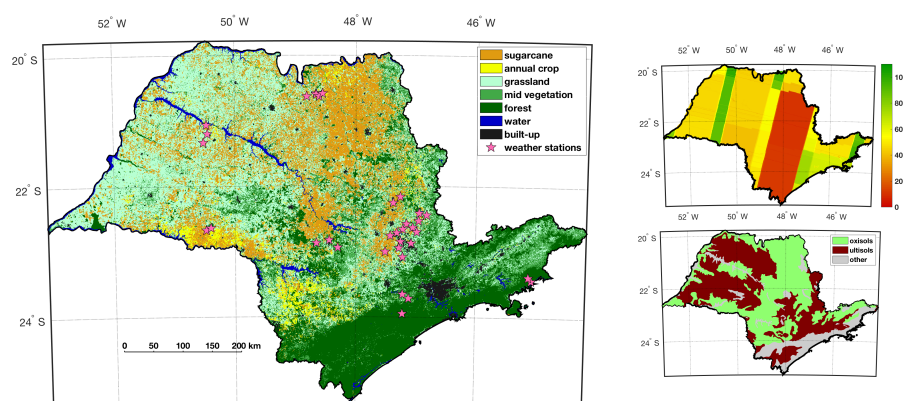


Figure 1. São Paulo state land cover map for 2015 based on Landsat data, produced by TU Delft [19,20] (left), the number of Sentinel-1 Interferometric-Wide acquisitions (top right) and the location of the two soil types (bottom right).

The São Paulo state sugarcane sector is important to Brazil, producing more than half of the national volume of bio-ethanol, which makes Brazil the largest sugarcane and second largest bio-ethanol producer in the world. Sugarcane acreage more than doubled over the last fifteen years [21,22]. This expansion, dominantly over grasslands, causes pasture fields to displace to areas

outside of the state and may indirectly affect the preservation of natural vegetation [23,24]. Sugarcane is a semi-perennial crop with a crop cycle of 12–18 months. After 5–7 cycles of growth from the same root system, new shoots are planted due to decreased yield. Usually, the plants reach up to 4 m high [25]. The annual crops in São Paulo state are usually rain-fed, but occasionally irrigated during growth and usually have a growth cycle of between three and five months, depending on the season and crop type. Different harvest practices are applied, especially in the case of corn, whereby the plant is sometimes left for drying of the corn kernels, after which the stalks are cut at 0.5 m above ground, and the field is tilled at a later stage, sometimes months later [26]. The grasslands are usually used as pastures for cattle grazing and can differ considerably in the state of vegetation. Overgrazing results in degraded pasture fields that are dry and with little or no grass cover, occasionally causing severe soil erosion [27]. Healthy pastures usually contain more grass cover, with grass height usually between 0.1 and 0.5 m [26]. When it grows beyond this point, the vegetation approaches shrubland, which belongs to the mid-vegetation class together with citrus trees up to 4 m in height. The forest crop class includes eucalyptus and pine trees, and both have a growth cycle of about seven and eleven years, respectively, and are planted mainly for paper, pulp and timber production. The native forest class mainly comprises Atlantic forest, a moist tropical forest with multiple tree canopies and plants ranging up to 30 m. Native forest is mainly present at the coastal zone, but it can also be found inland, usually covering water drainage channels between crop fields, which are protected by governmental regulations to preserve water presence in the soil, as well as in national parks and isolated patches scattered throughout the state. Built-up is defined as all man-made structures with an impervious surface and covers a broad range of objects, from roads and railways to buildings and industrial factories. Water is defined as any surface covered by a water layer, both natural and artificial water bodies.

2.2. Study Data

2.2.1. Sentinel-1 SAR

São Paulo state is covered by Sentinel-1 SAR products in three different modes: Interferometric-Wide (IW), Extra-Wide (EW) and StripMap (SM). The acquisitions in EW and SM modes occur west of São Paulo city and over the city, respectively. Since these modes feature different resolutions and since IW is the common mode, only the Sentinel-1 acquisitions in IW mode were used in this study. Due to the presence of the three different modes, the availability of IW acquisitions varies over São Paulo state. The upper right figure in Figure 1 illustrates the heterogeneous temporal coverage of images over São Paulo state. The wide swath of the EW acquisitions, resulting in very few IW acquisition, is clearly visible in the middle of the state.

The analysis is based on Ground Range-Detected (GRD) products, which were pre-processed applying thermal noise removal, radiometric calibration and terrain correction using the SRTM 30-m resolution Digital Elevation Model (DEM). The nominal IW resolution (range by azimuth) of 20 m by 22 m was spatially averaged to 30 by 30 m grid resolution. All acquisitions were taken during the descending orbit and occurred between 08:20 and 09:02 UTC (−2—3 h local time). The time frame ranges from October 2014 until December 2016, resulting in 739 images and 13 billion pixels in time and space. Together with the cross-polarization backscatter values, the corresponding incidence angles (ranging from 29.1°–46.0°) were extracted, as well.

In this study, we have restricted the analyses to backscatter data from the cross-polarized (VH) channel only, for two reasons. First of all, VH is known to be more sensitive to variations of the vegetation cover than the co-polarized channels and is therefore typically used for vegetation monitoring (see Section 1). Secondly, processing and storage capacities limited the data from one channel only, considering the area of interest and time frame.

2.2.2. Precipitation Condition

The precipitation data were extracted from products provided by the Global Precipitation Measurement (GPM) mission operated by NASA and JAXA. GPM offers three-hourly images gridded over 0.1° by 0.1° , which corresponds to approximately 11 km by 11 km in São Paulo state. The specific variable used, precipitationCal, is the multi-satellite precipitation estimate with gauge calibration. It is produced through the assimilation of observations from the Core Observatory satellite, launched in February 2014 and hosting a microwave imager and dual-frequency radar, with sensor data from various other precipitation-relevant passive microwave satellites, and at a later, stage post-processed with data from ground weather stations [28]. For this study, the daily cumulations were taken over the same time frame as the Sentinel-1 data.

In order to assess the reliability of the GPM data, we compared the precipitation measurements against in situ weather stations spread over São Paulo state (indicated by the pink stars in Figure 1). The distance between weather stations ranges from several kilometers to more than 600 km. For each site, at least two weather stations within 10 km were selected and cross-checked to ensure that each station actually records precipitation and is in acceptable agreement with other weather stations on short distances. It was concluded that the level of agreement of detecting precipitation between weather stations themselves is more than 0.85 on short distances (smaller than 5 km), after which, it rapidly drops below 0.75 (around 100 km), and over long distances, the rate reduces to approximately 0.6. Given the scattered positions of available weather stations in São Paulo state, this illustrates the need for precipitation data that are regularly sampled over large areas, as offered by GPM.

When comparing the detected precipitation by GPM with the weather stations, the rate of agreement is 0.70. This discrepancy can be explained by several reasons. First of all, there could be precipitation detected by GPM that is not detected by weather stations, due to the coverage difference of GPM pixels with respect to point measurements of weather stations and due to precipitation evaporation before reaching the ground level. Vice versa, the weather stations may be dependent on a minimum detectable amount of precipitation due to the instrument characteristics. Since accurate precipitation flagging of Sentinel-1 is essential for this study, we analyzed the effects of varying the threshold of measured precipitation on the agreement between the two data sources. From this, it followed that increasing the threshold of the weather stations rapidly decreases the rate of agreement to chance agreement (0.5). Conversely, when increasing the GPM threshold, the agreement increases to a maximum of 0.78 at 10 mm, after which the agreement declines. Hence, in order to increase the confidence in the GPM-derived rain-flagging, we adopted the following criteria for the labeling of Sentinel-1 pixels:

- All Sentinel-1 pixels within a GPM resolution cell were labeled as ‘affected by precipitation’ if GPM measured more than 10 mm of precipitation for that GPM cell and the eight adjacent GPM cells. This should furthermore hold for the day preceding the Sentinel-1 acquisition, as well as for the day of acquisition.
- All Sentinel-1 pixels within a GPM resolution cell were labeled as ‘not affected by precipitation’ if GPM measured zero precipitation for that GPM cell and the eight adjacent GPM cells. This should furthermore hold for all three days preceding the Sentinel-1 acquisition, as well as for the day of acquisition.

It should be noted that this labeling approach is conservative and that, consequently, numerous data-points were not rain-labeled and, therefore, not used in the analysis.

As an illustration of the effect of vegetation wetness on Sentinel-1 VH backscatter, Figure 2 shows high backscatter gains from the dry condition to the wet condition for marginal vegetation (low backscatter, colored as blue), approximating the backscatter from high vegetation (high backscatter, colored as red) for both conditions.

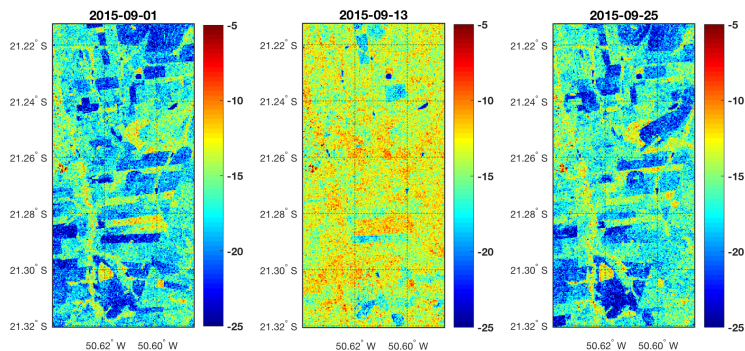


Figure 2. Three consecutively-acquired Sentinel-1 Interferometric-Wide (IW) images over the same area illustrating the effects of precipitation on VH backscatter (in dB). The area spans approximately 5 by 10 km. The right and left images show the backscatter, average -15.8 dB and -15.5 dB, respectively, when at least two days of dryness occurred. The middle image, average -13.0 dB, shows the same location with 40 mm of precipitation on the day before the acquisition and 15 mm on the day of acquisition. The blue patches in the dry images are most of all marginally vegetated (harvested fields or early stage crops), and the yellow and red patches are vegetated (mature crops or forests).

2.2.3. State of Vegetation

Since the publicly available land cover products do not offer information on the state of vegetation (such as phenology) or vegetation height on a similar time interval as the Sentinel-1 acquisitions (see Section 1), we use land cover maps that were generated by the TU Delft [19,20] (see Figure 1 for the map of 2015). In short, these maps are based on NDVI and NDWI derived from optical Landsat-5 and Landsat-8 data. Based on an adopted version of the Hidden Markov Model (HMM), the state of the land cover is estimated with a regular interval of 16 days (i.e., the temporal revisit of the Landsat satellites). For temporally-dynamic land cover types, specifically annual crops and sugarcane, the growth cycles are divided into respectively seven and 14 states. These states can be roughly linked to phenological stages, ranging from the first two states that can be regarded as bare ground or little vegetation to the last states that can be regarded as senescent plants just before harvest. For grassland, mid-vegetation and forest, no state distinctions are made throughout the season; hence, they are treated as state-stationary in time. The overall accuracies of these maps for the considered years, 2015 and 2016, are between 85% and 90%. The training and validation datasets for accuracy assessments were based on ground reference datasets acquired in São Paulo state, Brazil, including [25,26]. The most notable confusions were found between grasslands and mid-vegetation (high grasslands classified as mid-vegetation and young citrus orchards belonging to mid-vegetation classified as grasslands) and between grasslands, sugarcane and annual crops when gaps in optical observations due to clouds caused misregistration of harvest and early growth events.

The land cover maps, including the temporal states estimated on a 16-day interval, allow us to group states together into discrete vegetation categories. The following acronyms are used throughout this article for concise referencing to the states of vegetation:

- NV: No to marginal Vegetation (indicatively 0 m–0.1 m high): states of annual crop associated with bare ground and the germination stage.
- LV: Low Vegetation (indicatively 0.1 m–1 m high): states of sugarcane associated with early growth stages and grasslands.
- MV: Medium Vegetation (indicatively 2 m–4 m high): states of sugarcane associated with the maturation and senescence stages.
- HV: High Vegetation (indicatively 5–30 m high): forest.

The reason for taking only the first states of annual crops (and not sugarcane) for NV is because these fields are usually cleared or tilled before the growth cycle commences, consequently generally exposing bare ground. The first states of sugarcane on the other hand usually represent the land cover condition where straw material produced during the past sugarcane harvest event covers the remaining sugarcane sprouts for protection against plant dehydration. As such, in order to minimize the plant material present on the ground in this category, these sugarcane states are excluded. Conversely, LV does not include states of annual crop, predominantly because the annual crops may be irrigated during their vegetative stage, which is not captured by the GPM observations. In addition, the annual crops class can include corn and soybean, which differ in plant height and geometry. For MV, the mid-vegetation class was not included in order to avoid the erroneous reported inclusion of low to medium vegetation, such as high grasslands, and medium to high vegetation, such as citrus orchards.

2.2.4. Soil Type

For this study, two dominant soil types present in São Paulo state were taken into account, latossolos and argissolos, according to the Brazilian taxonomy convention, and oxisols and ultisols, respectively, according to the Soil Taxonomy of the USDA [29–31]. The location of these two soil types is depicted in the lower right plot of Figure 1. Each of these soil types covers approximately 40% of the state's area and in total supports more than 95% of the state's cultivated land.

Both soil types consist of mineral material, highly weathered by high temperatures and precipitation, and are typical for tropical regions [30]. The main difference is that the oxisols generally have higher clay content and are uniform in clay content, approximately 380 g/kg in the top one-meter layer, whereas for ultisols, a gradual increase in clay content in deeper soil layers is observed (158 g/kg from 0–0.2 m, 201 g/kg from 0.4–0.6 m and 334 g/kg from 0.8–1 m) [32].

This difference also has a direct effect on the SAR microwave interaction through the change in the medium's dielectric constant. As is explained in more detail in [33,34], the dielectric constant is a major contributor to microwave backscatter and is directly related to the soil chemical and physical properties. One of the main variables is soil moisture; the effect of water content on the medium's dielectric constant is dependent on the ability of the water molecule's dipole moments to align along an applied field. In soil, the first added water molecules are tightly bound to the soil particle's surface due to matric and osmotic forces, preventing the free movement of the dipole moments and hence causing a small increase in the dielectric constant. When the water content surpasses the transition (moisture) point, the water molecules become free particles, and the dielectric constant becomes much larger than that of soil. The well-known semi-empirical mixing dielectric model (SMDM) proposed by [35] allows for the computation of the permittivity of microwaves in soil as a function of its frequency, soil moisture and soil texture properties. Specifically, for the two São Paulo soils of interest, from the application of the model, it can be concluded that soil moisture has a major effect on the backscatter magnitude (roughly 5 dB increase). The difference in backscatter magnitude between the two soils themselves is negligible when both are dry and slightly higher backscatter for oxisols (higher clay content) when both are wet (smaller than 1 dB).

Specifically regarding the dry state of these specific soils types, the oxisols are characterized by the presence of deep, well-drained soils [32]. This leads to the additional assumption that the moisture content in soils that are assumed dry under the criteria reported in Section 2.2.2 is lesser for oxisols (i.e., resulting in lower backscatter) than for ultisols. Hence, from the SMDM and the characteristics of the soils, we expect that, on average, the backscatter from oxisols with respect to ultisols is slightly lower when both are dry and is slightly higher when both are wet. Finally, as a general note, C-band microwaves penetrate in the order of a few centimeters for wet soil and a few tens of centimeters for dry soil.

3. Methodology

3.1. Feature Extraction

In order to label each Sentinel-1 gridded observation with the incidence angle, precipitation condition, state of vegetation and soil type, these data sources were all projected onto the geographical Sentinel-1 grid with a pixel spacing of approximately 30 by 30 m, for each acquisition. Subsequently, only those samples were selected for which the precipitation condition of the consecutive acquisitions were known. From these data, the combined training and validation subset was randomly sampled per land cover type. For each of these samples, the backscatter and backscatter difference between consecutive epochs were taken, referred to as σ_0 and $\Delta\sigma_0$, respectively. Here, σ_0 is taken from the first of the two acquisitions, and $\Delta\sigma_0$ is the increase in σ_0 of the first acquisition to the second acquisition, i.e., $\Delta\sigma_0 = \sigma_0^{t+1} - \sigma_0^t$.

3.2. Configurations and Precipitation Information Scenarios

We define seven scenarios with different precipitation conditions (Table 2). We introduce acronyms for these scenarios that are referred to throughout this article. The table also shows the percentual occurrences of the precipitation conditions with respect to the total amount of extracted samples. This illustrates the effect of the strict precipitation detection criteria, as was delineated in Section 2.2.2. It shows that less than 25% of the samples were labeled as precipitated or non-precipitated and that less than 10% of the samples were labeled as condition pairs for consecutive acquisitions.

Table 2. Precipitation information scenarios, with acronyms used as references. The occurrences show the percentages of detected precipitation condition with respect to the total number of extracted samples. The occurrences between parentheses show the same, but normalized relative to their grouped sum. The variate analysis indicates which variables are used for classification.

Scenario	Precipitation Information	Occurrence	Variate Analysis
None	None	100% (100%)	Uni (σ_0)
NP	Non-Precipitated	18% (73%)	Uni (σ_0)
P	Precipitated	6% (27%)	Uni (σ_0)
P2NP	Precipitated to Non-Precipitated	0.9% (12%)	Bi ($\Delta\sigma_0, \sigma_0$)
NP2P	Non-Precipitated to Precipitated	0.8% (10%)	Bi ($\Delta\sigma_0, \sigma_0$)
P2P	Precipitated to Precipitated	0.7% (9%)	Bi ($\Delta\sigma_0, \sigma_0$)
NP2NP	Non-Precipitated to Non-Precipitated	5% (69%)	Bi ($\Delta\sigma_0, \sigma_0$)

The resulting amount of training samples for each scenario (Table 3) shows that for NV and MV, significantly lower amounts of samples are available than for LV and HV. The relative differences between the precipitation scenarios do not necessarily correspond to those of Table 2 because of the non-uniform occurrence of precipitation and land cover type in São Paulo state.

Finally, in order to evaluate the effect of incidence angle and soil type on the classification performance, we consider five configurations. These differ in ranges of incidence angles and soil types. The ranges of incidence angles were selected such that they were sufficiently disjoint, while still containing enough samples. The configurations overview is given by Table 4, which is placed just before the results and discussion (Section 4) for quick reference.

Table 3. Number of training samples for each vegetation state and precipitation scenario. NV, No to marginal Vegetation; LV, Low V; MV, Medium V; HV, High V.

Scenario	NV	LV	MV	HV
None	376,809	2,885,966	210,071	4,078,977
NP	352,150	2,177,407	180,758	3,260,595
P	24,659	708,559	29,313	818,382
P2NP	12,093	416,469	8469	489,782
NP2P	70,982	283,664	4452	284,219
P2P	12,566	292,090	20,844	328,600
NP2NP	281,168	1,893,743	176,306	2,976,376

3.3. Parametrization of Distributions and Classification

Based on the univariate and bivariate histograms for each configuration and each scenario, the Probability Density Functions (PDFs) were fitted using maximum likelihood estimation. As an illustrative example, Figure 3 shows the bivariate histogram of NV per precipitation condition. The center location of the distribution ($\Delta\sigma_0$, σ_0) changes from (−15 dB, −13 dB) for P2NP to (+7 dB, −21 dB) for NP2P. For the condition pairs with no precipitation condition change, P2P and NP2NP, the center σ_0 location is similar to the P2NP and NP2P, respectively, and the center $\Delta\sigma_0$ location is approximately zero.

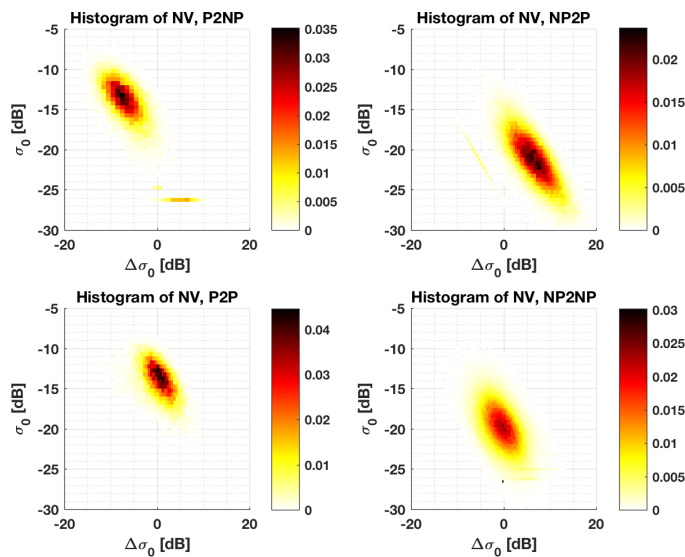


Figure 3. Bivariate histograms for NV per precipitation condition pair. The histograms are normalized by the probability density function estimate, i.e., the sum of the bin volumes approximates one.

The estimated PDFs from the NV bivariate histograms are also visualized as the contour lines, $f_{75\%}$, of the 75% Highest Density Region (HDR), defined as $R(f_{75\%}) = \{x : f(x) \geq f_{75\%}\}$, such that $P(X \in R(f_{75\%})) \geq 0.75$ in Figure 4. The differences in location of the distributions for the four precipitation condition pairs is now more clearly illustrated. In addition, the percentile plots of HV illustrate that the $\Delta\sigma_0$ location of the distribution is centered around zero for all precipitation condition pairs. In other words, the effect of precipitation on cross-polarized σ_0 is not observed for high vegetation, which can be explained by the dominant contribution of microwave volume scattering. Finally, the σ_0 location of HV is similar to the σ_0 location of NV for the P2NP and P2P conditions.

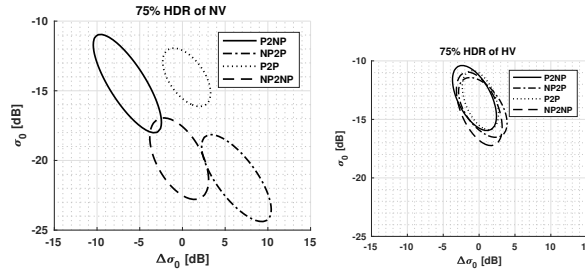


Figure 4. The 75% Highest Density Region (HDR) plots for NV and HV for all precipitation condition pairs.

For classification, the log-likelihoods are computed based on the univariate and bivariate distributions given the configuration and precipitation scenario. The vegetation state yielding the maximum log-likelihood is consequently compared to the true state for validation. Resulting confusion matrices are accompanied by the following:

- Overall accuracy: percentage of total number of correctly classified samples with respect to all classified samples.
- Producer's accuracy: fraction of correctly classified samples with respect to all samples of the truth, which is directly related to the omission error and can hence be interpreted as the accuracy from the classification map maker's perspective.
- User's accuracy: fraction of correctly classified samples with respect to all samples classified as this class, which is directly related to the commission error and can hence be interpreted as the accuracy from the classification map user's perspective.

All accuracies are corrected for the differences in the numbers of classified samples per class. Because of this, the overall accuracy is equal to the average of the producer's accuracies. In addition, the classification procedure was carried out ten times with randomly sampled sets of training data (70%) and validation data (30%), allowing for the computation of the averages and standard deviations of the accuracies. The differences between Configurations 2, 3, 4 and 5 with respect to Configuration 1 are presented as the corresponding accuracy improvements with a significance level of 0.05.

3.4. Hellinger Distances

To quantitatively describe the similarity of the PDFs for each precipitation scenario, we employ the Hellinger distance statistic, which is used for various applications related to classification techniques [36–40]. This metric H , which is closely related to the Bhattacharyya distance [41,42], is dependent on two bivariate normal distributions, $P \sim \mathcal{N}(\mu_1, \Sigma_1)$ and $Q \sim \mathcal{N}(\mu_2, \Sigma_2)$, where μ and Σ are the mean and covariance matrix of the distributions, respectively. Its generic expression [36,41,43] is:

$$\begin{aligned} H(P, Q) &= \sqrt{\frac{1}{2} \int \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx} \\ &= \sqrt{1 - \int \sqrt{f(x)g(x)} dx} \end{aligned} \quad (1)$$

Here, $H(P, Q) \in [0, 1]$ and a value of $H(P, Q) \rightarrow 0$ indicates high similarity, and $H(P, Q) \rightarrow 1$ indicates low similarity between P and Q [36].

Table 4. Configurations with varying incidence angle and soil type used for performance comparisons.

Configuration	Incidence Angle	Soil Types
1	All	All
2	29.1°–35.9°	All
3	39.2°–46.0°	All
4	All	Oxisols
5	All	Ultisols

4. Results and Discussion

4.1. Hellinger Distances

Under Configuration 1, based on Hellinger distances as a metric for the differences between the PDFs of the vegetation states per precipitation scenario (Figure 5), the following observations may be listed:

1. Generally, the PDFs of P and P2P are less discriminative than the PDFs of the other scenarios. When precipitation occurs, the backscatter of all vegetation states approximate the backscatter of HV, as was also illustrated by Figure 4. For both P and P2P, the distance between NV and HV is smaller than the distances between the other vegetation states and HV. This illustrates the severe effect of precipitation on backscatter when the ground is practically bare, seriously complicating accurate discrimination under such conditions. In addition, the distances between the vegetation states is generally smaller for P than for P2P, implying that the incorporation of two consecutive precipitation-affected acquisitions improves the discriminatory power as compared to a single precipitation-affected acquisition.
2. The distances for the scenarios with no information (None) and NP are similar to each other due to the relative underrepresentation of samples associated with scenario P (also see Table 3). As for the precipitation condition pairs, NP2P and NP2NP are similar to each other, but differ from P2NP. The latter mainly has larger differences between the PDF of NV and the PDFs of other vegetated states (LV, MV and HV). These latter states also show smaller differences between each other. This may be explained by P2NP causing higher remaining moisture content in vegetation (i.e., for LV MV and HV) after the first precipitation event as compared to the remaining moisture in soil when NV due to faster evaporation for bare ground.
3. Overall, there is relatively high confusion between LV and MV and to a lesser extent between LV and NV. This can be ascribed to the broad range of grasslands in the LV class and the confusion of higher grasslands with bushland; also see Section 2.2.3.

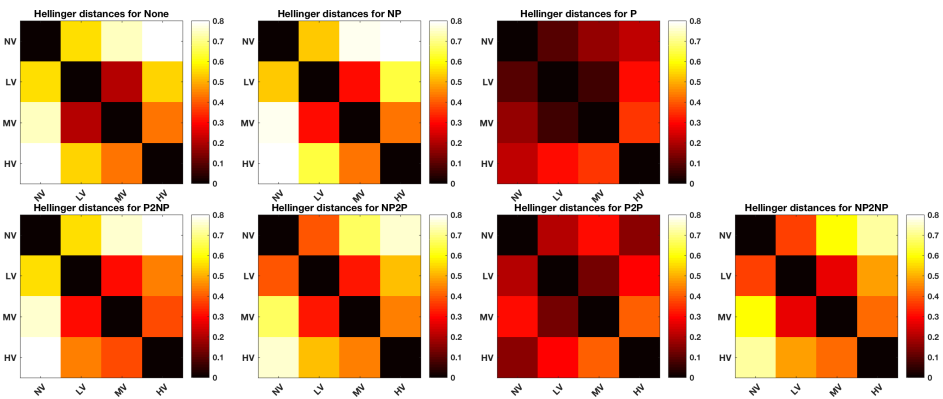


Figure 5. Hellinger distances between the PDFs of all vegetation states and for all precipitation scenarios.

4.2. Classification Results with Precipitation Conditions

The overall accuracies from the classification under Configuration 1 (Figure 6) first of all show that omitting P-related observations results in a marginal improvement of the classification performance from None to NP. Conversely, only taking P-related data significantly deteriorates the performance. This asymmetric behavior, which also applies to the producer's and user's accuracies, can be ascribed, similarly to before, to the difference in the number of data samples, i.e., their corresponding occurrences. The producer's accuracies for NV for both P and P2P are relatively low with respect to the user's accuracy due to the classification confusion with the other vegetation states (as was described in Discussion Point 1 in Section 4.1), particularly with HV for P2P.

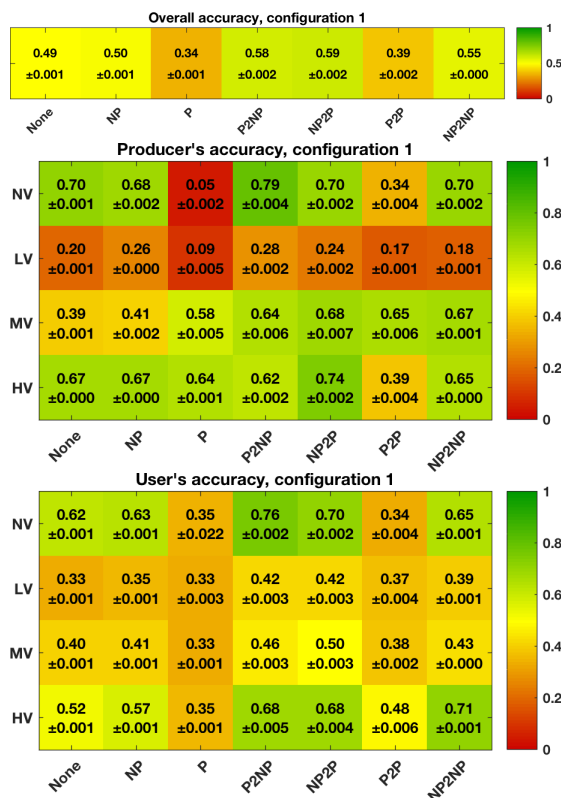


Figure 6. The averages and standard deviations of the overall (top), producer's (middle) and user's (bottom) accuracies for Configuration 1.

A greater classification improvement can be achieved when the bivariate information is taken into account, firstly demonstrated by the NP2NP-related accuracies with respect to the None-related accuracies. Hence, even though the center $\triangle_0$ location is close to zero (as indicated in Section 3.3), the second observation still adds valuable information to the classifier. In addition, the absolute difference in overall accuracies between NP2NP and P2P is similar to the difference between NP and P.

The best classification performances can be reached when changing precipitation conditions occur for consecutive acquisitions, as demonstrated by the increase in overall accuracies for P2NP and NP2P. With respect to NP2NP, the increases in the overall accuracies of P2NP and NP2P are approximately equal. Nevertheless, the producer's and user's accuracies show differences, both commonly for NV.

This is reflected by the larger Hellinger distances between NV and the other states for P2NP with respect to NP2P, as was described and explained by Discussion Point 2 in Section 4.1.

Finally, the general confusion of LV with the other states, as was laid out by Discussion Point 3 in Section 4.1, consequently results in relatively low producer's accuracies for all precipitation scenarios.

4.3. Effect of Incidence Angle on Classification Results

Restricting the training and evaluation to a narrower range of incident angles (Configurations 2 and 3) has a marginal effect on the classification accuracy (Figure 7). The slight deterioration could be ascribed to the non-uniform geographical distribution of the land cover classes and incidence angles, causing lower number of samples per subset and potentially an over-representation of clusters within one land cover type (especially for LV).

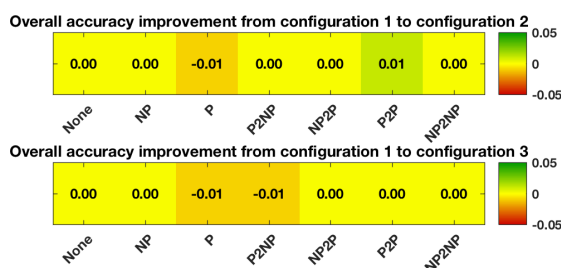


Figure 7. Overall accuracy improvements from Configuration 1 to Configuration 2 (top) and from Configuration 1 to Configuration 3 (bottom). Only the improvements with a significance level of 0.05 are non-zero.

4.4. Effect of Soil Type Information on Classification Results

The soil-related differences, on the other hand, show significant improvements, particularly for Configuration 5 (ultisols, lower clay content) for scenarios P, NP2P and P2P. These can be mainly attributed to less confusion of NV with the other vegetation states. A comparison of the PDF of NV with the PDF of MV for NP2P and the PDF of NV with the PDF of HV for P2P for Scenarios 4 and 5 with Scenario 1 (Figure 8) first of all indicates that the PDFs of Scenario 4 (oxisols, higher clay content) closely resemble the PDFs of Scenario 1 (all soil types combined), which is evidently based on data where the oxisols soil is over-represented.

Furthermore, as was earlier conveyed in Section 2.2.4, a slightly lower σ_0 when dry and slightly higher σ_0 when wet for Configuration 4 (oxisols, higher clay content) with respect to Scenario 5 (ultisols, lower clay content) was expected and is confirmed by the NP2P and P2P plots, respectively. For NP2P, this consequently leads to a higher $\Delta\sigma_0$ for Scenario 4. The combination of the differences in σ_0 and $\Delta\sigma_0$ between the two scenarios results in a smaller confusion with particularly MV for NP2P. For P2P the smaller confusion can mostly be attributed to the difference in σ_0 for NV with respect to HV. Both result in the main accuracy improvements, as is shown in Figure 9.

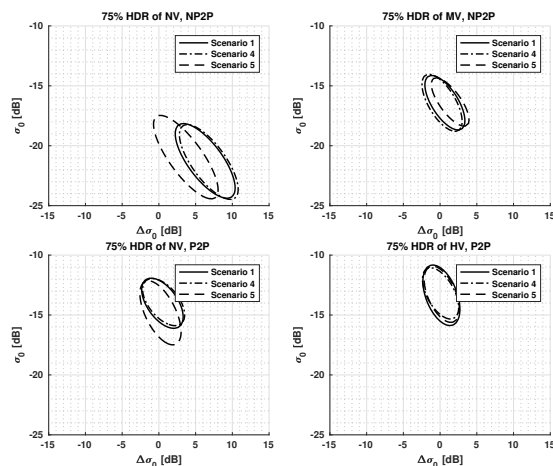


Figure 8. The 75% HDR plots for NV for NP2P and P2P for Scenarios 1, 4 and 5.

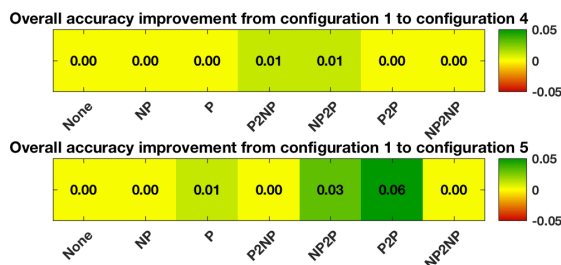


Figure 9. Overall accuracy improvements from Configuration 1 to Configuration 4 (top) and from Configuration 1 to Configuration 5 (bottom). Only the improvements with a significance level of 0.05 are non-zero.

5. Conclusions

Incorporating information on surface wetness combined with the signal from the consecutive SAR observation can significantly improve the characterization of vegetation. Moreover, a change in wetness condition between acquisitions further increases the characterization accuracy. This is caused by the non-uniform increase in SAR backscatter for the different vegetation states from dry to wet condition and vice versa. In order to maximize the accuracy even further, the classification model may be trained for different soil types as well, which, depending on their conductivity properties, increases the difference between probability density functions and hence enhances the separability between NV and other vegetation states.

Since the effect of precipitation on backscatter differences is dependent on the amount of vegetation, other gradations in vegetation coverage than used in this article may be investigated. As an example, improved discrimination may be possible between degraded grasslands and healthy grasslands, due to their difference in soil contribution on the SAR signals. Furthermore, the potential of co-polarized backscatter (VV in the case of Sentinel-1 IW in this area) for vegetation discrimination based on precipitation information may be studied, as well.

The methodologies as laid out in this article are not recommended for utilization as stand-alone classification techniques. Rather, we stress the concept that variations in surface wetness condition can actually be used advantageously in SAR-based land cover monitoring or classification models.

Author Contributions: R.A.M. conceived of and designed the study, designed the methodologies, analyzed and interpreted the data and results and wrote the manuscript. L.I. and P.L.D. assisted in the design of the study, the design of the methodologies, interpreting the data and results and reviewed the manuscript. L.I. furthermore should be acknowledged for the production of the land cover maps. P.S.G.M. assisted in the soil characterization and interpretation of the soil-related results and reviewed the manuscript. R.F.H. initiated the overall project and funding, endorsed the design and approved the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to acknowledge the ESA for providing the Sentinel-1 data and NASA and JAXA for providing the GPM data. In addition, we would like to express our gratitude to the Google Earth Engine team for providing access to these datasets. For access to the precipitation data from the weather stations in São Paulo state, we are grateful to the Department of Water and Electric Energy. Access to the land cover maps of São Paulo state provided by the TU Delft was made possible with the aid of Everton Valiati Hemerly. This work was carried out on the Dutch national e-infrastructure with the support of the SURF Cooperative. We are grateful for the support from the SURFsara team and for the allocation of resources on the Dutch national supercomputer. Finally, we would like to thank the colleagues at FEAGRI, Unicamp, for their assistance with taking the ground measurements, in particular Diego della Justina, Carlos Wachholz de Souza, Walter Rossi Cervi, Rubens Lamparelli and Jansle Vieira Rocha. The work was carried out in BE-Basic Project FES0905 and partly carried out within the framework of the joint BE-Basic FAPESP Project 2013/50943-9.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xiong, J.; Thenkabail, P.S.; Tilton, J.C.; Gumma, M.K.; Teluguntla, P.; Oliphant, A.; Congalton, R.G.; Yadav, K.; Gorelick, N. Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on google earth engine. *Remote Sens.* **2017**, *9*, 1065. [\[CrossRef\]](#)
2. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* **2014**, *103*, 7–27. [\[CrossRef\]](#)
3. Jokar Arsanjani, J.; Tayyebi, A.; Vaz, E. GlobeLand30 as an alternative fine-scale global land cover map: Challenges, possibilities, and implications for developing countries. *Habitat Int.* **2016**, *55*, 25–31. [\[CrossRef\]](#)
4. Bartholomé, E.; Belward, A.S. GLC2000: A new approach to global land cover mapping from Earth observation data. *Int. J. Remote Sens.* **2005**, *26*, 1959–1977. [\[CrossRef\]](#)
5. Friedl, M.A.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **2010**, *114*, 168–182. [\[CrossRef\]](#)
6. Defourny, P.; Schouten, L.; Bartalev, S.; Bontemps, S.; Caccetta, P.; Bella, C.D.; Gond, V.; Hazeu, G.W.; Heinimann, A.; Herold, M.; et al. Accuracy Assessment of a 300 m Global Land Cover Map: The GlobCover Experience. In Proceedings of the 33rd International Symposium on Remote Sensing of Environment, Sustaining the Millennium Development Goals, Stresa, Italy, 4–8 May 2009; pp. 1–5.
7. Bontemps, S.; Defourny, P.; Bogaert, E.V.; Kalogirou, V.; Perez, J.R. GLOBCOVER 2009 Products Description and Validation Report. *ESA Bull.* **2011**, *136*, 53.
8. Fritz, S.; See, L.; McCallum, I.; Schill, C.; Obersteiner, M.; Van Der Velde, M.; Boettcher, H.; Havlík, P.; Achard, F. Highlighting continued uncertainty in global land cover maps for the user community. *Environ. Res. Lett.* **2011**, *6*, 1–6. [\[CrossRef\]](#)
9. Herold, M.; See, L.; Tsensbazar, N.E.; Fritz, S. Towards an integrated global land cover monitoring and mapping system. *Remote Sens.* **2016**, *8*, 1036. [\[CrossRef\]](#)
10. Steele-Dunne, S.C.; McNairn, H.; Monsivais-Huertero, A.; Judge, J.; Liu, P.W.; Papathanassiou, K. Radar Remote Sensing of Agricultural Canopies: A Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 2249–2273. [\[CrossRef\]](#)
11. Baghdadi, N.; Boyer, N.; Todoroff, P.; El Hajj, M.; Bégué, A. Potential of SAR sensors TerraSAR-X, ASAR/ENVISAT and PALSAR/ALOS for monitoring sugarcane crops on Reunion Island. *Remote Sens. Environ.* **2009**, *113*, 1724–1738. [\[CrossRef\]](#)
12. McNairn, H.; Brisco, B. The application of C-band polarimetric SAR for agriculture: A review. *Can. J. Remote Sens.* **2004**, *30*, 525–542. [\[CrossRef\]](#)

13. Moran, M.S.; Alonso, L.; Moreno, J.F.; Cendrero Mateo, M.P.; de la Cruz, D.F.; Montoro, A. A RADARSAT-2 Quad-Polarized Time Series for Monitoring Crop and Soil Conditions in Barrax, Spain. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1057–1070. [[CrossRef](#)]
14. Kontgis, C.; Warren, M.S.; Skillman, S.W.; Chartrand, R.; Moody, D.I. Leveraging Sentinel-1 time-series data for mapping agricultural land cover and land use in the tropics. In Proceedings of the 2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Brugge, Belgium, 27–29 June 2017; pp. 1–4.
15. Balzter, H.; Cole, B.; Thiel, C.; Schmullius, C. Mapping CORINE land cover from Sentinel-1A SAR and SRTM digital elevation model data using random forests. *Remote Sens.* **2015**, *7*, 14876–14898. [[CrossRef](#)]
16. Torbick, N.; Chowdhury, D.; Salas, W.; Qi, J. Monitoring rice agriculture across myanmar using time series Sentinel-1 assisted by Landsat-8 and PALSAR-2. *Remote Sens.* **2017**, *9*, 119. [[CrossRef](#)]
17. El Hajj, M.; Baghdadi, N.; Zribi, M.; Angelliaume, S. Analysis of Sentinel-1 radiometric stability and quality for land surface applications. *Remote Sens.* **2016**, *8*, 406. [[CrossRef](#)]
18. Satalino, G.; Balenzano, A.; Mattia, F.; Davidson, M.W. C-band SAR data for mapping crops dominated by surface or volume scattering. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 384–388. [[CrossRef](#)]
19. Iannini, L.; Molijn, R.; Mousivand, A.; Hanssen, R. Monitoring LULC dynamics in the Sao Paulo region through landsat and C-band SAR time series. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 90–93.
20. Iannini, L.; Molijn, R.; Mousivand, A.; Hanssen, R.; Lamparelli, R.A.C. A HMM-based approach for historic and up-to-date land cover mapping through Landsat time-series in the state of Sao Paulo, Brazil. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 5457–5460.
21. Machado, P.G.; Rampazo, N.A.M.; Picoli, M.C.A.; Miranda, C.G.; Duft, D.G.; de Jesus, K.R.E. Analysis of socioeconomic and environmental sensitivity of sugarcane cultivation using a Geographic Information System. *Land Use Policy* **2017**, *69*, 64–74. [[CrossRef](#)]
22. INPE. Canasat. Available online: <http://www.dsr.inpe.br/canasat/> (accessed on 16 October 2018).
23. Alkimim, A.; Clarke, K.C. Land use change and the carbon debt for sugarcane ethanol production in Brazil. *Land Use Policy* **2018**, *72*, 65–73. [[CrossRef](#)]
24. Rudorff, B.F.T.; de Aguiar, D.A.; da Silva, W.F.; Sugawara, L.M.; Adami, M.; Moreira, M.A. Studies on the Rapid Expansion of Sugarcane for Ethanol Production in São Paulo State (Brazil) Using Landsat Data. *Remote Sens.* **2010**, *2*, 1057–1076. [[CrossRef](#)]
25. Molijn, R.A.; Iannini, L.; Rocha, J.V.; Hanssen, R.F. Detailed ground reference data for sugarcane biomass estimation in São Paulo state, Brazil. *Sci. Data* **2018**, *5*, 1–18. [[CrossRef](#)] [[PubMed](#)]
26. Molijn, R.A. Land cover ground reference data in São Paulo state, Brazil, taken in 2015. TU Delft, 2015. Dataset.
27. Merten, G.H.; Minella, J.P. The expansion of Brazilian agriculture: Soil erosion scenarios. *Int. Soil Water Conserv. Res.* **2013**, *1*, 37–48. [[CrossRef](#)]
28. Huffman, G.J.; Bolvin, D.T.; Nelkin, E.J. *Integrated Multi-satellite Retrievals for GPM (IMERG) Technical Documentation*; Technical Report; NASA: Washington, DC, USA, 2017.
29. USDA. *Keys to Soil Taxonomy*; Technical Report; USDA: Washington, DC, USA, 2014.
30. Eswaran, H.; Reich, P. World Soil Map. In *Encyclopedia of Soils in the Environment*; Elsevier: Amsterdam, The Netherlands, 2005; pp. 352–365.
31. Embrapa. *Sistema Brasileiro de Classificação de Solos*, 2nd ed.; Embrapa Informação Tecnológica: Brasília, Brazil, 2006; p. 306, ISBN 85-85864-19-2.
32. Alves, M.R.; Demattê, J.A.M.; Barros, P.P.S. Multiple Geotechnological Tools Applied to Digital Mapping of Tropical Soils. *Rev. Bras. Ciênc. Solo* **2015**, *39*, 1261–1274. [[CrossRef](#)]
33. Behari, J. *Microwave Dielectric Behavior of Wet Soils*; Springer: Dordrecht, The Netherlands, 2005; ISBN 978-1-4020-3271-4.
34. Gadani, D.H.; Vyas, A.D. Dielectric mixing model for the estimation of complex permittivity of wet soils at C and X band microwave frequencies. *Indian J. Pure Appl. Phys.* **2015**, *53*, 190–198.

35. Dobson, M.C.; Ulaby, F.T.; Hallikainen, M.T.; El-Rayes, M.A. Microwave Dielectric Behavior of Wet Soil-Part II: Dielectric Mixing Models. *IEEE Trans. Geosci. Remote Sens.* **1985**, *GE-23*, 35–46. [[CrossRef](#)]
36. Salmon, B.P.; Kleynhans, W.; Van Den Bergh, F.; Olivier, J.C.; Marais, W.J.; Grobler, T.L.; Wessels, K.J. A search algorithm to meta-optimize the parameters for an Extended Kalman filter to improve classification on hyper-temporal images. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012; pp. 4974–4977.
37. Silva, W.B.; Freitas, C.C.; Sant’Anna, S.J.; Frery, A.C. Classification of segments in PolSAR imagery by minimum stochastic distances between wishart distributions. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 1263–1273. [[CrossRef](#)]
38. Salmon, B.P.; Kleynhans, W.; Bergh, F.V.D.; Olivier, J.C.; Marais, W.J.; Wessels, K.J. Meta-optimization of the extended Kalman filter’s parameters for improved feature extraction on hyper-temporal images. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011; pp. 2543–2546.
39. Erol, H.; Akdeniz, F. A new supervised classification method for quantitative analysis of remotely-sensed multi-spectral data. *Int. J. Remote Sens.* **1998**, *19*, 775–782. [[CrossRef](#)]
40. Genovez, P.C.; Freitas, C.C.; Santanna, S.J.; Bentz, C.M.; Lorenzetti, J.A. Oil Slicks Detection from Polarimetric Data Using Stochastic Distances between Complex Wishart Distributions. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 463–477. [[CrossRef](#)]
41. Subasi, Y.; Demirekler, M. Quantitative measure of observability for stochastic systems. *IFAC Proc. Vol. (IFAC-PapersOnline)* **2011**, *18*, 4244–4249. [[CrossRef](#)]
42. Girish, V. Hellinger distance. *Commun. Complex.* **2011**, *2011*, 12.
43. Frasca, M.; Liberati, R. Riemann manifolds from Hellinger distance. In Proceedings of the 2012 Tyrrhenian Workshop on Advances in Radar and Remote Sensing: From Earth Observation to Homeland Security, TyWRRS 2012, Naples, Italy, 12–14 September 2012; pp. 59–61.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

6

CONCLUSIONS AND RECOMMENDATIONS

Je moet je afvragen: blijkt ergens uit dat leven een bedoeling heeft? Ik denk eerder dat het tegenovergestelde het geval is. Darwin en Wallace hebben laten zien hoe belangrijk het toeval is, dus juist het ontbreken van een doelstelling. Dat is het bepalende principe van de evolutie. Later hebben biologen de vraag opgeworpen: als leven een doel heeft, waarom zit het dan zo slordig in elkaar? Dat spreekt me zeer aan. Neem de blinde vlek in het menselijk oog. De zenuwen die de signalen naar je hersenen overbrengen, gaan door de retina heen. Dus is er een plekje in je oog dat niet gevoelig is voor licht, omdat die zenuwbanen erdoorheen moeten. Welke sukkel heeft dat ontworpen? Dat had helemaal niet gehoeven, het netvlies van de octopus bijvoorbeeld is onbelemmerd. Als God een Delftenaar geweest zou zijn, zou hij hier nooit op zijn afgestudeerd.

Vincent Icke

IN this concluding chapter, the findings from the studies will be discussed in light of the research questions that were introduced in Section 1, including a presentation of preliminary results that demonstrates the potential of produced land cover maps. Subsequently, a set of recommendations is provided.

CONCLUSIONS

The main goal of this study is to advance the state-of-the-art on sugarcane productivity monitoring and land cover monitoring through SAR and optical remote sensing data. Below, we address the four main research questions:

1. *How can ground reference measurements that describe the biophysical state of sugarcane be acquired to support the analysis of space-based SAR and optical measurements for productivity monitoring?*

Chapter 2 presented a methodology that is based on a set of techniques and allometric models that allows for estimating biomass based on non-destructive bio-

physical field measurements. Combining the collected biomass and biophysical parameters is important for the interpretation and analysis of SAR and optical remote sensing signals due to their sensitivity to volume (change) and geometric characteristics of the plants. Sound understanding of the growth developments, based on the measured biophysical parameters and photographs, proved to be indispensable to analyze the field situations in this study. Moreover, the non-destructive collection of field measurements minimized the introduction of biases to the remote sensing signals. The demanding conditions of such an endeavor in terms of time and effort were adequate for the sugarcane growth development analyses based on remote sensing, but limited the investigation of the signals' full potential with respect to the intra-field variability mapping, as will be discussed in the following research question item.

In terms of quality of the measurements, the comparison between the directly field-measured biomass and biomass values that were estimated based on field-measured biophysical parameters (sugarcane stalk diameter, height, stalk mass density and leaf mass) at the same locations overall brought satisfyingly high correlations for further use of the models. The more detailed quality assessments also demonstrated which uncertainties should be taken into account when estimating biomass from biophysical measurements. These were signified as a function of the instrument operation precision (dependent on instrument errors), idealization precision (dependent on positioning errors) and modeling precision (dependent on function fitting errors), whereby the idealization precision generally proved to be the most dominant. The explanations on (i) the setup of the measurement campaign, (ii) the methodologies for collecting and combining the measurements and (iii) the outcomes of the quality assessments can help to adapt and optimize the design of new sugarcane ground reference campaigns. Altogether, the collected ground reference dataset offers useful knowledge for supporting the analysis of SAR and optical satellite measurements for productivity monitoring, whereby the collection of non-destructive biophysical field-measurements is a major asset. Yet, the specific limitations of the dataset in terms of spatial coverage, temporal coverage and quality should be taken into account when using the dataset or setting up similar acquisition campaigns.

2. *How can measurements from space-based SAR and optical sensors be exploited to effectively monitor sugarcane productivity?*

We presented several insights into the techniques to monitor sugarcane growth (Chapter 3). First of all, through coupling the ground reference dataset to remote sensing imagery from various SAR and optical sensors, it was shown that the monitoring of sugarcane growth is limited by (mainly) saturation, precipitation and signal characteristics. For both optical and SAR remote sensing signals, sensor-specific time windows were identified that constrain the effectiveness of monitoring sugarcane growth. Specifically to SAR, the signals proved to be particularly sensitive to surface wetness (e.g. caused by precipitation events), affecting the saturation point. For optical, the saturation point occurs approximately halfway the growth period, whereby C-band SAR signals saturate approximately at around

one-fourth of the period. For L-band SAR, the signals did not reveal saturation points during growth.

Whereas the ground reference dataset was adequate for analysis on remote sensing for growth monitoring over the span of sugarcane growth, it should also be noted that its spatial coverage limited the examination of the full potential of the signals for detecting sugarcane growth anomalies in space. Nevertheless, the comparison of the estimated ground reference biomass with the remote sensing signals indicated that only high-resolution optical imagery has potential to detect growth variations in the field. We also found that, based on only remote sensing imagery, the temporal consistency of spatial patterns is highly dependent on the growth stage and differs for SAR and optical. For SAR signals, the consistency also depends on environmental conditions, such as precipitation and beam geometry conditions. As a result, sets of remote sensing images can be effectively selected for carrying out sugarcane productivity monitoring based on saturation-dependent time windows, beam configurations and sensor type.

The second study in this chapter, on the identification of spatio-temporal patterns, showed that the two-fold hypothesis testing technique can be used to discover field locations that continuously emerge in time. The technique gives an indication on how to detect these locations and subsequently to analyze their correspondence to sugarcane growth anomalies. The third study demonstrated that spatial gaps in sugarcane fields influence the SAR signals as a function of vegetation height, surface wetness and averaging windows. These findings contribute to further understanding the considered remote sensing signals (in particular L-band and high resolution C-band imagery) in relation to sugarcane gaps and growth conditions, and furthermore provide indications on the signals' potential for detecting sugarcane growth anomalies.

Altogether, the results showed that remote sensing images can be selected based on specific time windows and environmental conditions under which the remote sensing imagery is acquired, as is specified in more detail in Chapter 3, for effective monitoring sugarcane growth. Furthermore, to assess (more) effectively the potential of remote sensing imagery for intra-field biomass variability estimation, the features of the ground reference dataset and the satellite measurements ideally match, including their spatial and temporal resolutions, their measured and estimated parameters and their levels of precision (as will be discussed further in [Recommendations](#)).

3. *How can the combined use of space-based SAR and optical measurements improve their singular use for land cover monitoring with a focus on sugarcane classification?*

Chapter 4 demonstrated how an adapted Hidden Markov Model (HMM) that models the sequence of system states based on stochastic mechanisms allows for the consistent progress monitoring of states during vegetation growth while effectively combining SAR and optical measurements. We found that missing certain crop cycle events (such as harvest events) and the occurrence of environmental events (such as precipitation and drought) during the acquisitions of satellite measure-

ments create the most common classification confusions between the various land cover types. Specific examples of such confusions include those between sugarcane and grassland due to growth anomalies and between annual crops and grasslands due to remaining (corn) stalks after harvest and tall grassland.

In terms of overall model performance, the classification accuracies showed that optical-only observations generally outperform SAR-only observations, whereby the model based on optical-only data is particularly sensitive to the missing of growth events such as early growth and harvest. Insights into these characteristics are of particular importance when such conditions are more likely to occur, e.g. in severely cloudy regions. When combining SAR and optical measurements, the kappa accuracy nearly doubles in comparison to the singular use of SAR and optical measurements.

Altogether, the combination of the HMM's ability to (i) track states, to (ii) cope with temporally irregular (optical) data, to (iii) digest data from both sensors and to (iv) improve the result when combining these sensor data, makes it a particularly attractive alternative to commonly used land cover models. Yet, the model's sensitivity to gaps in (optical) observational during important crop events and the limited ability of SAR measurements to compensate for these are important limitations of the model when using these datasets. The presented research can help to assess the potential of the model in light of the model's tracking capabilities of vegetation states, and to further explore the advantages of combining the data from both sensors for improved land cover maps.

6

4. *How can a-priori knowledge on precipitation and soil type be used advantageously for the characterization of vegetation states through space-based SAR measurements?*

Chapter 5 quantified the classification performance improvements that can be achieved when information on precipitation events and soil type are taken into account. It was shown that the results already improve when excluding SAR observations that were acquired under surface wetness conditions and further improve when changes in wetness condition occur between consecutive acquisitions (i.e. from non-precipitation to precipitation or vice versa). The incorporation of information on soil type (differentiated by their clay content) proved to be additionally beneficial for the results, though to a lesser extent. The incorporation of both types of information proved to be effective for enhanced discrimination between certain growth stages, particularly between bare soil and other states since the SAR signals from bare soil are more sensitive to wetness than other vegetation states. The incorporation of the incidence angle did not reveal noteworthy improvements.

Altogether, the analysis was mainly carried out in order to demonstrate the classification improvements that can be achieved rather than the provision of stand-alone classification techniques. These findings show a method that effectively combines SAR acquisitions and precipitation and soil information rather than discarding precipitation-affected signals or underestimating their potential when including these acquisitions for vegetation mapping.

The following preliminary and exemplary results were produced by applying the land cover monitoring technique that was presented in Chapter 4, based on optical images

only. It should be noted that these results have not been further applied and published, and serve here only to illustrate new research opportunities, to (partly) support the recommendations and to demonstrate possible applications concerning research on a biobased society.

First of all, the produced land cover maps can visualize the effect of drought on the water level of a Brazilian reservoir used for hydro-electric power generation (Figure 6.1); after one of the most severe droughts in the region since 1960 that happened during 2014 and 2015 [1, 2]. It shows that the main body of the reservoir as well as the side bulges contracted. In addition, the maps show considerable conversions of pasture land into sugarcane land, as well as the expansion of urban areas. Hence, such detailed maps can facilitate in detecting temporal variations in water reservoirs and in pinpointing land cover changes.

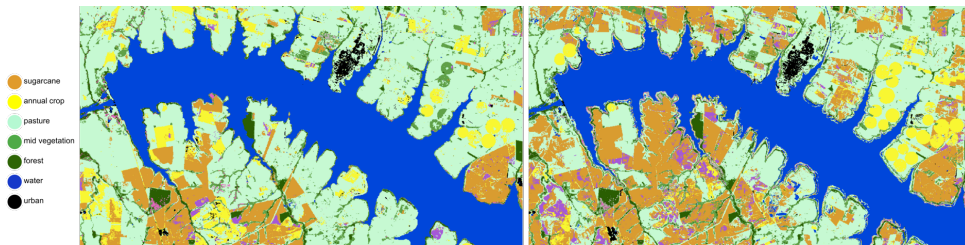


Figure 6.1: Water reservoir for hydro-electric power in the Tietê river in the west of São Paulo state, in 2003 (left) and after severe drought 2015 (right). Main differences are the reduced water area and the replaced pasture fields by sugarcane fields. North is up.

The maps can also show the extensive expansion of sugarcane over pasture fields over the course of approximately a decade on municipality-level, in this case a municipality in São Paulo state named Palestina (Figure 6.2). In terms of acreage, sugarcane expanded by approximately 65% between 2005 and 2014, most dominantly along the municipality's edges, bordering local rivers. The expansion mostly occurred over pasture land, which decreased by approximately 80%.

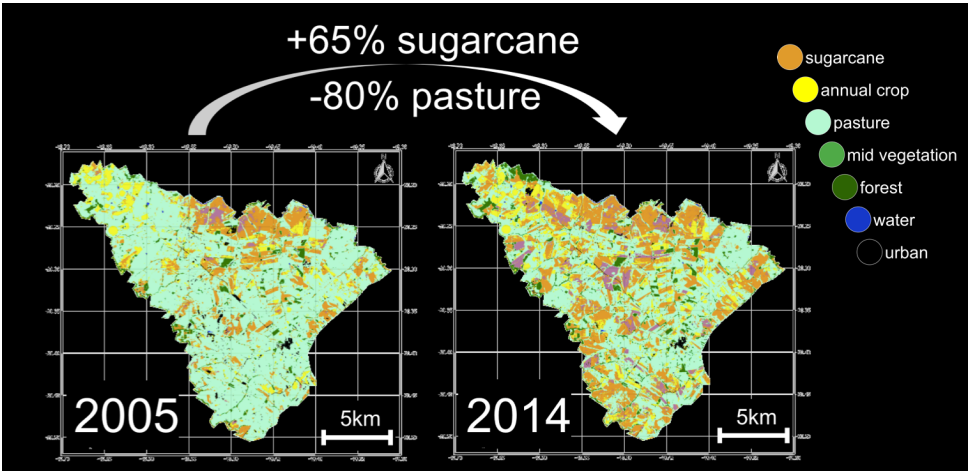


Figure 6.2: Land cover in municipality Palestina in the north of São Paulo state, in 2005 (left) and 2014 (right). North is up.

6

Furthermore, through the land cover maps of the entire São Paulo state, the dynamics of land cover types in time (Figure 6.3) can be observed. The loss map shows that between 2005 and 2015 the main land cover loss occurred for pasture, particularly in the middle and west of the state. The majority of these fields were replaced by sugarcane. In fact, though not shown here, the intermediate maps reveal that over the course of the period, the substitution of pasture by sugarcane gradually shifted in northwest direction, pushing out pasture to the neighboring states. The protected (native) Atlantic forest in the east of the state has virtually not changed.

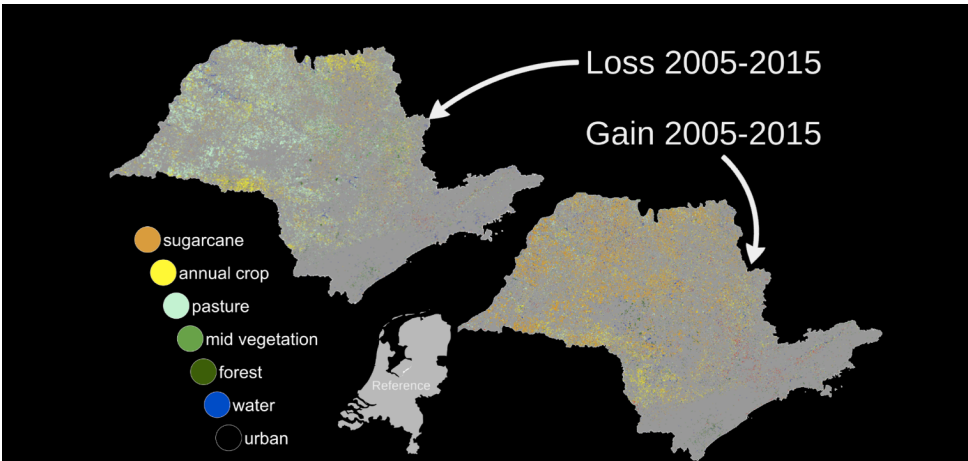


Figure 6.3: Loss map (left) and gain map (right) of São Paulo state between 2005 and 2015. It shows wide-scale replacement of sugarcane with pasture fields, most particularly in the western part of the state. The map of the Netherlands puts the size of the state in perspective. North is up.

Finally, the aggregate of land cover area over the entire state from 2003 to 2015 (Figure 6.4) shows that sugarcane increased by more than half, while pasture and annual crops decreased by more than 20%. In addition, the small deviation of water in 2014 and 2015 is most likely related to the earlier mentioned severe drought period.

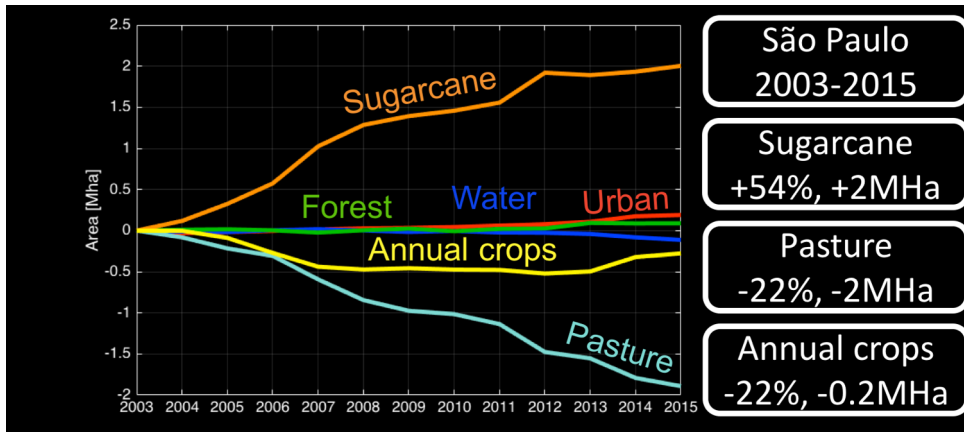


Figure 6.4: Change in land cover area in São Paulo from 2003 to 2015 in million hectares (Mha).

Based on these examples, we advocate for the application of the (land cover) techniques that are proposed in this study to São Paulo state as well as to neighboring states. In particular, the spatial trend in terms of land cover migration can give further insights into the (indirect) land cover changes that are induced by the expansion of bio-energy crops such as sugarcane. As pasture fields move westward and northward and out of the state, it is of particular interest to assess what the effect is on land cover types elsewhere, such as native vegetation and annual crops.

RECOMMENDATIONS

Since the majority of the presented research relies on collected ground reference data, it is recommended to enrich such ground reference datasets with dense plant and soil moisture measurements along with plant failure and plant substitution (e.g. weed infestations) measurements. Especially for SAR satellite imagery, these moisture and plant measurements are of interest due to the sensitivity of SAR signals to wetness conditions. The methodology for setting up such ground campaigns as was presented in Chapter 2 may be adopted and adjusted according to the specific needs, whereby the non-destructive estimation of biomass may be of particular importance when undesired alteration of space-based observations needs to be minimized. In addition, in order to match the spatial resolution of the field-collected biomass measurements with the resolution of the analysed satellite imagery, yield sensors may be mounted on the harvester, shedding more light on the value of the spatial features in the satellite images that are closest to harvesting. Ideally, the spatial and temporal acquisition density, the parameters and levels of precision of the ground measurements should match the effective spatial and temporal resolution, parameters to be estimated and levels of precision of

the satellite sensors, respectively. The resolutions could be approximated by applying variable scales, whereby the spatial density of field measurements is dependent on the (biomass) variability, and the temporal resolution of field measurements is dependent on the closeness to signal saturation and environmental conditions (such as precipitation conditions and growth stage).

Secondly, the consistency analysis of the remote sensing imagery as was presented in Chapter 3 should be extended through the incorporation of Sentinel-1 Interferometric Wide Swath images, as only Sentinel-1 Extra Wide Swath images were available for this specific research. This also holds for the inclusion of more ALOS-2 images, which could shed more light on saturation effects and the understanding of spatial patterns. Additionally, short-wave SAR imagery from satellites such as TerraSAR-X and COSMO-SkyMed may be examined, whereby it should be acknowledged that saturation effects will probably limit the effectiveness to early crop stages due to their signal's wavelength. Regarding optical imagery, it is recommended to incorporate larger sets of high-resolution images from satellites, for example from WorldView, RapidEye and Planet satellites. Alternatively, as a quick-win solution, freely available images from Sentinel-2 may give easily attainable insights into the effect of the use of higher-resolution imagery (10 meter resolution) than the already analyzed Landsat imagery (30 meter resolution).

Specifically for SAR remote sensing, it is recommended to analyse the potential of InSAR coherence and PolSAR for monitoring vegetation growth, particularly for detecting agricultural events (such as crop emergence, crop closure and harvest events) as well as for detecting environmental conditions (such as precipitation and drought events) and integrate these into models such as the proposed version of the HMM. Successful detection of agricultural events could substantially improve the tracking of crop stages and classification of land cover types when employing land cover monitoring techniques that facilitate the incorporation of state probabilities. The inclusion of environmental conditions could further improve such a technique by means of attributing reliability estimates to the observations, giving, for example, different weights to SAR observations than to optical observations when affected by precipitation. Additionally, the methodology that was presented in Chapter 5 may be reversed in order to detect precipitation events. Also, the sensitivity of SAR signals to soil moisture may be exploited for the discrimination of healthy and degraded grasslands. Such analysis is particularly relevant for a region like São Paulo state, in light of the threat of pasture degradation to the fitness of arable land and its relation to overgrazing, as was earlier touched upon in Chapter 5.

Finally, but most importantly, it is recommended to apply the presented HMM-based land cover monitoring model on the state of São Paulo first and subsequently observe the trends in spatial expansion, reduction and migration of certain land cover types in time, in particular of sugarcane, grasslands, annual crops and forest. As was carried out for this study, it is recommended to utilize Google Earth Engine (GEE) to acquire and preprocess the remote sensing data, followed by the transfer of the data to a high-performance computing facility (e.g. SURFsara) in order to apply the proposed models to the data in tiles and finally combine these tiles in time and space. Alternatively, all steps (preprocessing, models and map production) can be incorporated into GEE for direct processing on GEE's facilities. The model may, furthermore, be trained, validated and applied to Brazil in its entirety. Such analysis may give highly valuable insights into potential relation-

ships between the expansion of sugarcane, the migration of grasslands that are used as pastures toward the Amazon and the loss of forests in those regions.

REFERENCES

- [1] R. Zhang, L. A. Cuartas, L. Valerio, D. C. Carvalho, S. Birkinshaw, G. S. Mohor, C. A. Nobre, and M. E. Seluchi, *Season-based rainfall–runoff modelling using the probability-distributed model (PDM) for large basins in southeastern Brazil*, [Hydrological Processes](#) **32**, 2217 (2018).
- [2] C. A. S. Coelho, D. H. F. Cardoso, and M. A. F. Firpo, *Precipitation diagnostics of an exceptionally dry event*, [Theoretical and Applied Climatology](#) **125**, 769 (2016).

CURRICULUM VITÆ

Ramses Amadeus MOLIJN

05-08-1983 Born in Puttershoek, the Netherlands.

EDUCATION

1995–2001	School Goois Lyceum, Bussum
2001–2009	University, MSc. & BSc. Bachelor of Science: Aerospace Engineering Master of Science: Geomatics
2013–2020	University, PhD. Delft University of Technology & University of Campinas <i>Thesis:</i> Space-based SAR and optical remote sensing for pro- ductivity monitoring and mapping of sugarcane <i>Promotor:</i> Prof. dr. ir. R.F. Hanssen <i>Promotor:</i> Prof. dr. J. Vieira Rocha

LIST OF PUBLICATIONS

INCLUDED IN DISSERTATION

- [1] R. A. Molijn, L. Iannini, J. V. Rocha, and R. F. Hanssen, *Ground reference data for sugarcane biomass estimation in São Paulo state, Brazil*, [Scientific Data](#) **5**, 1 (2018).
- [2] R. A. Molijn, L. Iannini, J. V. Rocha, and R. F. Hanssen, *Sugarcane productivity mapping through C-band and L-band SAR and optical satellite imagery*, [Remote Sensing](#) **11** (2019), 10.3390/rs11091109.
- [3] R. A. Molijn, L. Iannini, R. F. Hanssen, and J. Vieira Rocha, *Sugarcane growth monitoring through spatial cluster and temporal trend analysis of radar and optical remote sensing images*, in [International Geoscience and Remote Sensing Symposium \(IGARSS\)](#), Vol. 2016-November (2016) pp. 7141–7144.
- [4] R. A. Molijn, L. Iannini, C. W. De Souza, D. D. Justina, J. V. Rocha, and R. F. Hanssen, *The effects of sugarcane productivity anomalies on L-band and C-band SAR signals*, in [International Geoscience and Remote Sensing Symposium \(IGARSS\)](#), Vol. 2018-July (IEEE, 2018) pp. 5371–5374.
- [5] R. A. Molijn, L. Iannini, P. L. Dekker, J. V. Rocha, and R. F. Hanssen, *Integration of space-borne SAR and optical data into a Hidden Markov Model-based approach for land cover monitoring (under review)*, (2019).
- [6] R. A. Molijn, L. Iannini, P. L. Dekker, P. S. Magalhães, and R. F. Hanssen, *Vegetation characterization through the use of precipitation-affected SAR signals*, [Remote Sensing](#) **10**, 1 (2018).

NOT IN DISSERTATION

- [1] R. A. Molijn, L. Iannini, R. F. Hanssen, F. J. Van Leijen, R. A. C. Lamparelli, and A. Coutinho, *Integration of SAR and optical dense time series for land cover monitoring*, in [International Geoscience and Remote Sensing Symposium \(IGARSS\)](#), Vol. 2017-July (2017) pp. 4330–4333.
- [2] L. Iannini, R. Molijn, A. Mousivand, R. Hanssen, and R. A. C. Lamparelli, *A HMM-based approach for historic and up-to-date land cover mapping through Landsat time-series in the state of Sao Paulo, Brazil*, [International Geoscience and Remote Sensing Symposium \(IGARSS\) 2016-Novem](#), 5457 (2016).
- [3] L. Iannini, R. Molijn, A. Mousivand, and R. Hanssen, *Monitoring LULC dynamics in the Sao Paulo region through Landsat and C-band SAR time series*, [International Geoscience and Remote Sensing Symposium \(IGARSS\) 2015-Novem](#), 90 (2015).

- [4] R. A. Molijn, L. Iannini, A. Mousivand, and R. F. Hanssen, *Analyzing C-band SAR polarimetric information for LAI and crop yield estimations*, *Remote Sensing for Agriculture, Ecosystems, and Hydrology XVI* **9239**, 92390V (2014).

DATASETS

- [1] R. A. Molijn, *Sugarcane ground reference data over four fields in São Paulo state, Brazil, taken in 2014 and 2015*, (2015), dataset, doi.org/10.4121/uuid:37112e18-f794-4d66-a8cd-7f1e92af09fc.
- [2] R. A. Molijn, *Land cover ground reference data in São Paulo state, Brazil, taken in 2015*, (2015), dataset, doi.org/10.4121/uuid:7b900822-4efe-42f1-9b6e-a099eda4ba02.

WEBSITE

- [1] R. A. Molijn, L. Iannini, and E. Valiati Hemerly, *TU Delft - BE-Basic WebGIS*, (2018), <http://be-basic.grs.tudelft.nl/maps/316/view>.