

Propagating Sentinel-2 Top-of-Atmosphere Radiometric Uncertainty Into Land Surface Phenology Metrics Using a Monte Carlo Framework

Lukas Valentin Graf , Javier Gorroño , Andreas Hueni , Achim Walter , and Helge Aasen 

Abstract—Time series of optical imagery allow one to derive land surface phenology metrics. These metrics are only complete with a statement about their uncertainty. A source of uncertainty is the radiometry of the sensor. We propagated radiometric uncertainties within a Monte Carlo framework into phenological metrics using the TIMESAT approach based on time series of the normalized difference vegetation index (NDVI), three-band enhanced vegetation index (EVI), and green leaf area index (GLAI) derived from radiative transfer modeling. In addition, we studied the effect of propagated uncertainties on scene preclassification. We focused on Sentinel-2 multispectral imager top-of-atmosphere data since quantitative estimates of radiometric uncertainties are available. Propagation was carried out for a growing season over an agricultural region in Switzerland. Propagated uncertainties had little impact on the classification except for spectrally mixed pixels. Effects on the spectral indices and GLAI were more pronounced. In detail, the GLAI was more uncertain due to the ill-posedness of radiative transfer model inversion (median relative uncertainty for all crop pixels and Sentinel-2 scenes: 4.4%) than EVI (2.7%) and NDVI (1.1%). Regarding phenology, metrics exhibited largest uncertainties in the case of GLAI. The magnitude of uncertainty in the metrics depends on the interscene error correlation, which we assumed to be either zero (uncorrelated) or one (fully correlated) since the actual correlation is unknown. If uncertainties are fully correlated, uncertainties in metrics are small (two to three days) but take values up to greater ten days under the uncorrelated assumption. Thus, our work provides guidance for the interpretation of phenological metrics.

Manuscript received 23 February 2023; revised 19 April 2023, 7 July 2023, and 10 July 2023; accepted 18 July 2023. Date of publication 21 July 2023; date of current version 27 September 2023. This work was supported by the European Cooperation in Science and Technology (COST) through COST Action CA17134 “Optical Synergies for Spatiotemporal Sensing of Scalable Ecophysiological Traits.” The work of Lukas Valentin Graf was supported by the Swiss Science Foundation for the project “PhenomEn” under Grant IZCOZ0_198091. The work of Javier Gorroño was supported by the European Space Agency Living Planet Fellowship under Contract 4000130980/20/I-NS. (Corresponding author: Lukas Valentin Graf.)

Lukas Valentin Graf and Helge Aasen are with the Crop Science Group, Institute of Agricultural Science, ETH Zürich, 8092 Zürich, Switzerland, and also with the Earth Observation of Agroecosystems Team, Agroecology and Environment Division, Agroscope, 8046 Zürich, Switzerland (e-mail: lukasvalentin.graf@usys.ethz.ch; helge.aasen@agroscope.admin.ch).

Javier Gorroño is with the Research Institute of Water and Environmental Engineering, Universitat Politècnica de València, 46022 València, Spain (e-mail: javorvie@upv.es).

Andreas Hueni is with the Remote Sensing Laboratories, Department of Geography, University of Zürich, 8057 Zürich, Switzerland (e-mail: andreas.hueni@geo.uzh.ch).

Achim Walter is with the Crop Science Group, Institute of Agricultural Science, ETH Zürich, 8092 Zürich, Switzerland (e-mail: achim.walter@usys.ethz.ch).

Digital Object Identifier 10.1109/JSTARS.2023.3297713

Index Terms—Ecophysiological parameters, land surface phenology (LSP), Monte Carlo (MC) simulation, radiative transfer modeling, Sentinel-2 (S2), TIMESAT, uncertainty.

I. INTRODUCTION

SATELLITE remote sensing is of paramount importance for studying global environmental changes. More than four decades of optical remote sensing have generated a wealth of image and time-series data that have made a significant contribution to the understanding of vegetation dynamics [1], [2], [3]. From local to global scales, remote-sensing-based canopy greenness proxies [4] and derived functional traits [5], [6] provide the ability to quantify the effects of environmental factors on plant physiology over time [7], [8], [9], [10]. Seasonal patterns of remote sensing data of vegetation are mostly summarized under the term “land surface phenology” (LSP [11]). LSP has been extended to a set of metrics linking observed temporal changes in the spectral properties of vegetation to distinctive physiological transition phases [12]. Examples of these are start of season (SOS) and end of season (EOS) [13], [14], which mark the onset of the growing season after winter and its end, respectively. Both the metrics have, therefore, been used intensively in agricultural [15], [16], [17], [18], [19], forestry [7], [20], [21], [22], and ecological studies [23], [24], [25], [26]. For agriculture, such metrics are of increasing importance since different crops have strongly differing phenologies. Partly, the phenology depends on intrinsic properties of the crop. For example, winter wheat often is sown in October in the northern hemisphere, whereas maize or other summer crops are sown in spring. Moreover, field management and environmental conditions (drought, hail, etc.) affect the crops’ development, and the precise analysis of crop phenology is an important tool for insurance assessments or for decision support when it comes to management issues, such as finding the right time for fertilization.

In the past, sensors, such as MODerate Resolution Imaging Spectroradiometer (MODIS) or Advanced Very High Resolution Radiometer were mainly used for LSP studies. These have a high temporal but only a low spatial resolution (≥ 250 m), which leads to spectral mixing effects [27] and lack of spatial detail. With the Copernicus Sentinel-2 (S2) mission, in addition to enhanced spatial resolution (up to 10 m), the twin constellation of Sentinel-2A (S2A) and Sentinel-2B (S2B) has improved temporal resolution remarkably (up to three days at midlatitudes),

making S2 data a valuable data source for LSP studies and agricultural applications [8], [10], [26], [28], [29]. The multispectral imager (MSI) onboard the S2 satellites has 13 spectral bands, ten of which are suitable for remote sensing of land surfaces. By placing three spectral bands in the red edge, two bands in the near infrared, and two short-wave infrared bands, the S2-MSI instrument is suitable for vegetation studies [30], [31] and the accurate derivation of plant ecophysiological traits, such as the green leaf area index (GLAI). This extends vegetation mapping capabilities beyond widespread spectral indices of canopy greenness, such as the normalized difference vegetation index (NDVI [32]) or the enhanced vegetation index (EVI [33]). S2 data are, for instance, used operationally to derive the Copernicus Pan-European High-Resolution Vegetation Phenology and Productivity layer.¹

Despite the clear importance and widespread use of remote sensing data, our understanding of uncertainty in remotely sensed products, such as LSP, is still poor albeit White and Nemani [34] already stressed the lack of remote sensing uncertainty estimates more than one decade ago. Furthermore, the Quality Assurance Framework for Earth Observation endorsed by the Committee on Earth Observation Satellites (CEOS), which has recently been extended by a joint effort of ESA and NASA [35], highlights the need for uncertainty assessment of earth observation (EO)-derived products, such as LSP. Only with a quantification of the sources of uncertainty (uncertainty budget), traceable and complete data products can be generated that allow users to evaluate the products for their fitness for purpose and indicate the uncertainty of their own analyses. Here, we focus on a source of uncertainty in LSP metrics derived from S2-MSI data, which, to the best of our knowledge, have hardly been considered so far: uncertainty in the top-of-atmosphere (TOA) at-sensor-radiance values ($L_{\text{at_sensor}}$). Any remote sensing product necessarily builds on TOA radiances. Uncertainties in the radiometry are translated all the way down to the LSP metrics and define the limits of the achievable uncertainty. Mittaz et al. [36] concluded that no uncertainty propagation from raw data to higher level products has been conducted so far along an EO data processing chain. Moreover, the authors called for advancing EO practice using approaches from metrologia. A first step toward closing this gap is, therefore, taken with this article on the propagation of radiometric uncertainty.

Radiometric uncertainty is influenced by a set of uncertainty sources that derive from different contributions, including both the instrument and the ground processing chain. Some of these contributions include, for example, the instrument noise or the solar diffuser used for onboard calibration [37]. In turn, these contributions can affect pixel-based uncertainty with different levels of correlation in the spatial, temporal, and spectral domains [38]. Uncertainties of 0.1–1.5 K are known from remote sensing of sea surface temperature [39], for example, while relative radiometric uncertainties of up to 5% have been found in remote sensing of ocean color due to random effects [40], [41]. It is, therefore, reasonable to assume that uncertainties in

the radiometry of S2-MSI, which take about 1–2% [38], also have an impact on derived data products, such as LSP metrics.

The objective of this article is to propagate radiometric uncertainty from S2-MSI TOA (L1C) data into LSP metrics (L3) using widely used image processing resources focusing on an intensively farmed agricultural study area in Switzerland for a single growing season (see Section III). Swiss agriculture operates with small field sizes (average farm size 2020: 21 ha) of a larger number of crops next to each other. This high diversity is difficult to manage, and satellite-based solutions for management decision support are, therefore, urgently required. Spatial, temporal, and spectral uncertainty contributors are propagated using a Monte Carlo (MC) framework as recommended in Supplement 1 to the Guide to the Expression of Uncertainty in Measurement (GUM [42]) and implemented, for example, by Gorroño et al. [38], [43]. Starting with the L1C data, we propagate the radiometric uncertainty through the atmospheric correction (AC) into bottom-of-atmosphere (BOA, i.e., processing level L2A) reflectance factor values. From the BOA reflectance factors, we derive the two most widely used spectral vegetation indices (VIs): NDVI and EVI. Furthermore, we estimate the ecophysiological development of plants using GLAI, which we determine from the inversion of a radiative transfer model (RTM). Based on the uncertainties in NDVI, EVI, and GLAI, we construct vegetation time series and derive SOS and EOS and their uncertainty in days. In addition, we test the effect of propagated uncertainties on the scene classification layer (SCL) output by the L2A processor. The complete method outline is presented in Section IV followed by a presentation (see Section V) and discussion of the results obtained (see Section VI).

II. TERMS AND DEFINITIONS

To avoid ambiguity in terms, we provide definitions here: we define the term “uncertainty” as the degree of doubt about the reported TOA reflectance values, which is referred to as the best estimate. “Error,” in contrast, is the difference between the measured value and the true unknown TOA reflectance value and splits into random and systematic components. Systematic errors can be minimized by pre- and postlaunch calibration activities [44], while random errors can be suppressed by a sufficiently large sample size. Residual, i.e., uncorrected random and systematic errors, contributes to the overall radiometric uncertainty budget. Following the specification of the GUM, uncertainty estimates represent a confidence interval for a given probability distribution function (PDF) whose mean corresponds to the measured TOA reflectance value. Standard uncertainty is defined as the interval around the mean of the uncertainty PDF that provides a coverage of 68.27% ($k = 1$) of possible realizations. In equations, we denote the standard uncertainty with the Greek letter μ to be consistent with the GUM. All the findings in this article are reported as standard uncertainties. To obtain uncertainty values with a higher coverage factor k (i.e., multiple standard deviations), the uncertainties must be multiplied by k , for example, to obtain uncertainties for $k = 2$ used in instrument certification.

¹[Online]. Available: <https://land.copernicus.eu/pan-european/biophysical-parameters/high-resolution-vegetation-phenology-and-productivity>

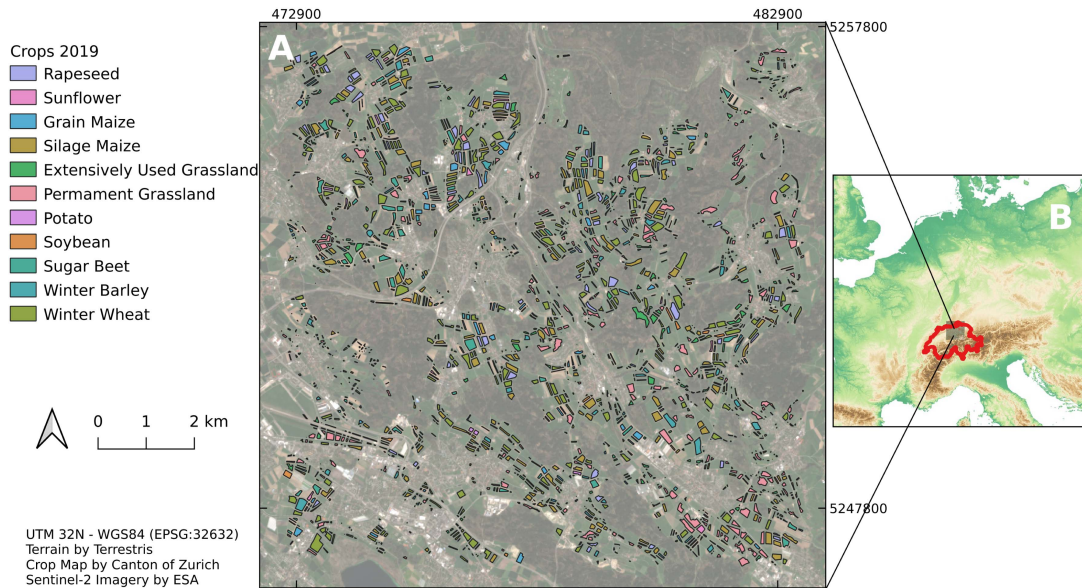


Fig. 1. (a) Map of the study area with the field parcel geometries buffered 20 m inward and their main crop types in 2019. (b) Location of the S2 tile T32TMT and the study region in Western Europe.

III. STUDY AREA AND DATA

A. Study Area

We selected an intensively farmed study area located around the agricultural research station at Eschikon (8.69E, 47.45N) operated by ETH Zürich, northeast of the city of Zürich, Switzerland (see Fig. 1). The study area is bound by a square with an approximated area of 100 km² [see Fig. 1(a)]. With an average annual precipitation of 1241 mm and an air temperature of 10.1 °C (reference period 2004–2022, based on a weather station² located at Strickhof-Lindau placed in the northwestern part of the study area), the study region is representative not only for cropping conditions in the Swiss Central Plain but also for geographically adjacent areas in Central and Western Europe.

A crop-type map showing the main crop per field parcel for the year 2019 was available from the administration of the canton of Zürich. The parcel boundaries and their crop type are shown in Fig. 1(a). To avoid spectral mixing effects at the parcel boundaries, all the geometries were buffered 20 m inwards. Parcels too small for the buffering operation were dropped from the database. In total, 2021 field parcels with nine different crop types and two types of grassland were available. Most parcels are small; the median parcel size was about 0.19 ha (0.44 ha on average). Table I shows the approximate area per crop type in hectares with winter wheat (228.8 ha) occupying the largest area.

B. Data

We acquired 34 S2-MSI TOA scenes (LIC product) over the study area [S2 granule 32TMT, Fig. 1(b)] from the Data and

TABLE I

AREA PER CROP TYPE IN HECTARES IN THIS STUDY AFTER BUFFERING ALL THE FIELD PARCELS 20 M INWARD TO AVOID SPECTRAL MIXING EFFECTS

Crop type	Area [ha]
Rapeseed	70.4
Grain maize	41.3
Extensively used grasland	62.9
Permanent grasland	168.9
Potato	4.7
Silage maize	175.6
Soybean	7.8
Sugar beet	65.4
Sunflower	7.9
Winter barley	63.4
Winter wheat	228.8

Information Access Service (DIAS) CREODIAS³ covering a time period from February 1 to October 15 2019. All the scenes had a scene-wide cloud coverage of $\leq 20\%$, which was used as a filtering criterion. This threshold is lower than reported in the literature for phenology retrieval. It was chosen to ensure that cloud and shadow contamination did not affect uncertainty propagation. Fig. 2 shows the S2 scenes used and their cloudy pixel percentage derived from the metadata. The spacecraft (S2A or S2B) are denoted as blue and red dots, respectively. The PDGS (S2 Payload Data Ground Segment) processing baseline of the S2 scenes was 02.07 (N0207) for all the scenes before July 16 2019, and 02.08 (N0208) after that date. According to the

²[Online]. Available: <https://www.agrometeo.ch/>

³[Online]. Available: <https://finder.creodias.eu/>

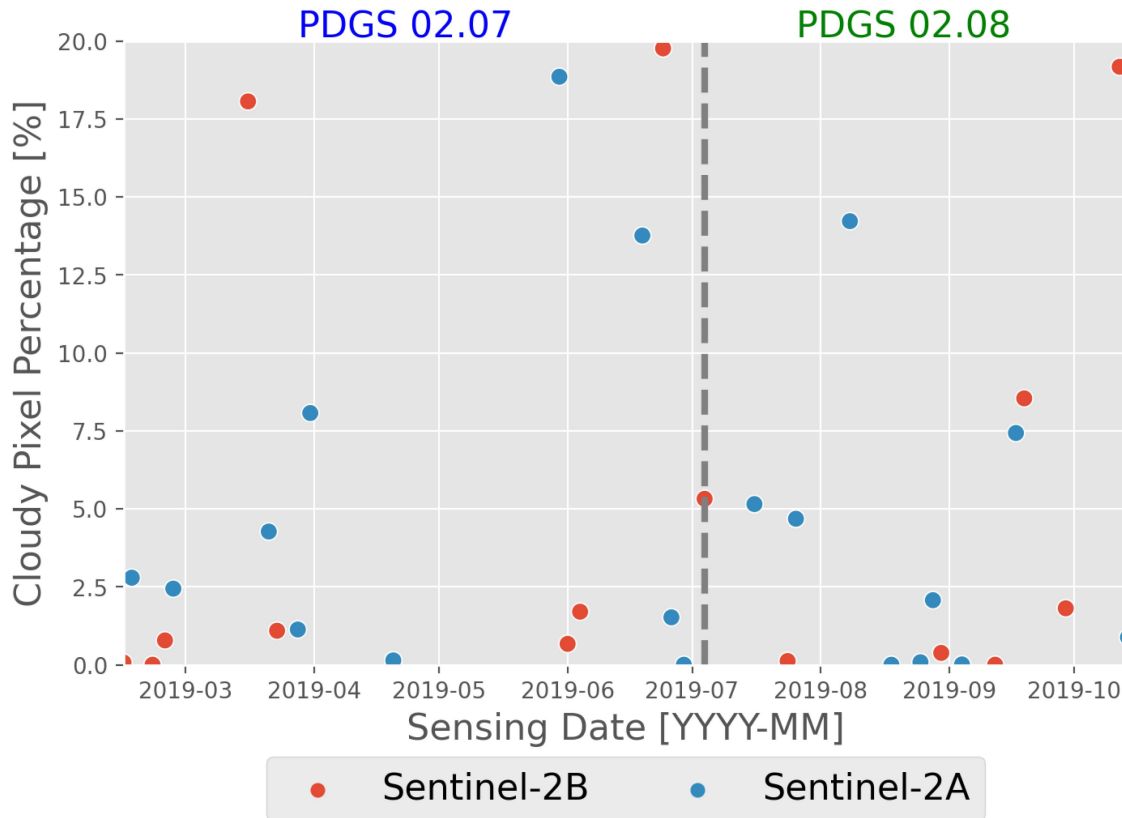


Fig. 2. S2 scenes used plotted against their scene-wide cloudy pixel percentage. Red dots denote observations made by S2B, and blue dots represent S2A scenes. The two different PDGS processing baselines are separated by the gray dashed line.

technical guide,⁴ the change in baseline only affected the formatting of the instrument telemetry metadata. Therefore, we do not assume an impact of the baseline change on radiance values.

IV. METHODS

In the following, the manuscript is structured along the workflow in Fig. 3. Fig. 3 shows the entire radiometric uncertainty propagation chain—starting from the L1C TOA S2 scenes to uncertainties in the LSP metrics (CEOS processing level L3). The individual steps from Fig. 3 are explained in detail in the following Sections IV-B–IV-E. Complementary to the uncertainty propagation workflow, the total uncertainty budget of the LSP metrics is shown in an uncertainty tree diagram in Fig. 4. The diagram visualizes the context the radiometric uncertainty propagation chain is embedded into, as explained in Section IV-A.

A. Uncertainty Tree Diagram

Uncertainty tree diagrams represent the uncertainty budget of a measurement and trace back sources of uncertainty to

⁴[Online]. Available: <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/processing-baseline>

their causing effects [36], [45]. Uncertainty tree diagrams are complementary to workflow illustrations, such as the one in Fig. 3 depicting the origin and entanglement of uncertainties. Fig. 4 shows an uncertainty tree diagram for the LSP metrics. Starting from the final product (LSP metrics in the large blue box in Fig. 4), sources of uncertainty are plotted along paths. Partial derivatives represent sensitivity coefficients. For example, $\frac{\partial LSP}{\partial NDVI}$ represents the sensitivity of uncertainty in the LSP metrics to uncertainty in the NDVI. Sources of uncertainty that cannot be quantified at this time are included in the total budget with $\mu(0)$, i.e., these contributions are set to zero, although their actual contribution is most likely > 0 .

Uncertainty in the LSP metrics depends not only on the parameterization of the time-series model (light blue colored parts in Fig. 4) but also on uncertainties in the spectral indexes (EVI or NDVI) or biophysical parameters (GLAI, pink colored parts in Fig. 4) used. Here, it should be noted that EVI, NDVI, and GLAI are alternative ways to derive the LSP metrics. All these, in turn, depend on the uncertainty in the L2A product, which includes uncertainties in the L2A BOA reflectance factors, as well as the resulting uncertainties in the scene preclassification (green colored parts in Fig. 4). The uncertainty in the L2A product ultimately results from the uncertainty in the L1C TOA reflectance factors (dark blue parts in Fig. 4). Their uncertainty can be determined by means of the S2 Radiometric Uncertainty

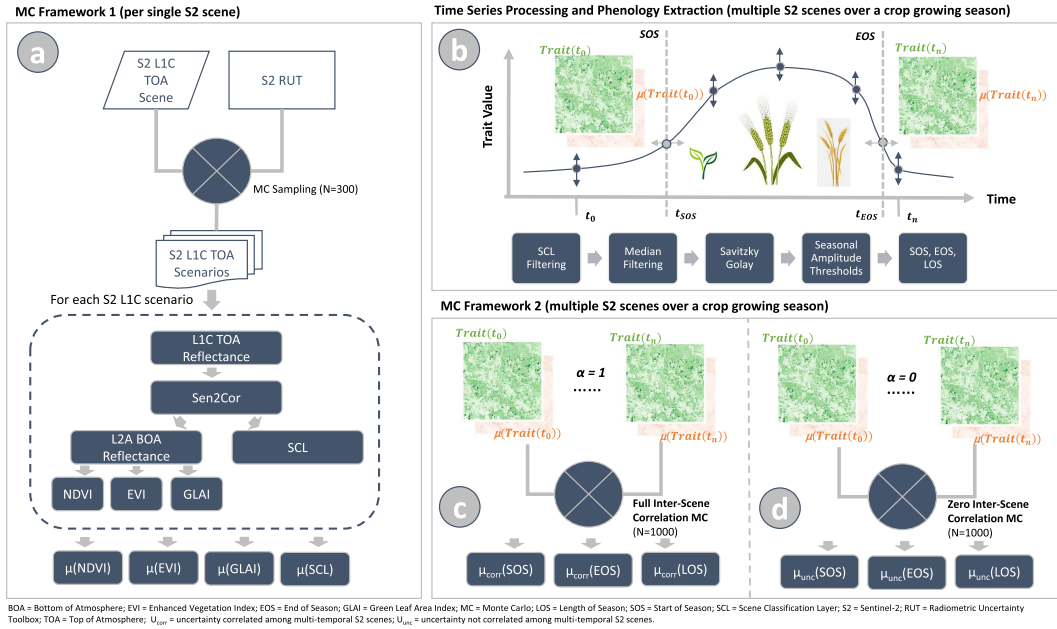


Fig. 3. Overview of the workflow for uncertainty (μ) propagation. (a) Using the S2 RUT, radiometric uncertainties are propagated from TOA reflectance factors through AC into NDVI, EVI, and GLAI as well as SCL uncertainty in a first MC framework working on single S2 scenes. (b) Multiple (n) S2 scenes acquired over an entire growing season are processed to obtain LSP metrics from a trait (e.g., NDVI) time series. Uncertainties obtained in the first MC framework are fed into (b) for a second MC sampling to derive LSP uncertainty. Interscene (i.e., multitemporal) error correlation is set to (c) full correlation and (d) zero correlation to include two possible extreme cases for LSP uncertainty retrieval.

Toolbox (S2-RUT, [37], dashed border in Fig. 4, upper left). The uncertainty sources shown in the S2-RUT include effects whose uncertainties are, thus, propagated through the entire EO processing pipeline.

Besides the radiometric uncertainty propagation path, there are other effects that contribute uncertainties to the uncertainty budget of the LSP metrics. Examples are uncertainties in the GLAI retrieval due to uncertainty in the RTM used or the uncertainty of the AC in the L2A data. Therefore, all the contributions except the radiometric uncertainty derived from the S2-RUT are set to zero, since the quantification of the other uncertainties is far beyond the scope of this article (see also Section VI-C).

B. S2 Radiometric Uncertainty

We calculated the pixel-based uncertainty in the S2 TOA reflectance factor values using the S2-RUT, which is an extension to the Sentinels Application Platform. The S2-RUT takes an S2 L1C product as input and returns the relative standard uncertainty of each implemented contributor for each pixel in each spectral band in its native spatial resolution [see Fig. 3(a)]. Overall, 11 uncertainty contributors were considered (see Fig. 4, dashed box in the top left corner) based on [38]. Other sources of error in the PDGS process chain, at the end of which is the L1C product, were not considered, such as polarization error (denoted by $\mu(0)$ above the “PDGS” box in Fig. 4). For a detailed explanation of the error sources used in the S2-RUT, as well as the effects of those not considered, the reader is referred to [37] and [38].

C. MC Simulation of S2 TOA Reflectance Factors

To convert the radiometric uncertainty obtained per spectral band and contributor from the S2-RUT into higher level products, we implemented an MC framework for error propagation working on single S2 scenes [see Fig. 3(a)]. The goal is to generate L1C TOA reflectance factor scenarios from random samples, which include the determined radiometric uncertainty and the known error correlations in the spectral, spatial, and temporal domains. The domains differ in their definition from the common remote sensing terminology, except for the spectral domain. The spatial component comprises the pixels along a scan line of the MSI detector array. This domain corresponds to the across-flight direction, which is approximately east–west direction. The temporal domain, in turn, denotes consecutively scanned lines along the sensor flight direction, i.e., approximately north–south direction (see also Section II).

The MC-derived error in a single pixel i from uncertainty contributor j in MC iteration (scenario) t ($\delta_{t,j,i}$) is a linear combination of uncorrelated ($\delta_{uncorr,i}$) and correlated ($\delta_{corr,i}$) error terms

$$\delta_{t,j,i} = (1 - \alpha) \cdot \mu_{j,i}^{RUT} \cdot x_{t,j,i}^{uncorr} + \alpha \cdot \mu_{j,i}^{RUT} * x_{t,j,i}^{corr}. \quad (1)$$

The left-hand side of (1) denotes the uncorrelated error term, where $\mu_{j,i}^{RUT}$ is the pixel-based uncorrelated radiometric uncertainty (see Section IV-B) and $x_{t,j,i}^{uncorr}$ is a sample drawn from a Gaussian $\mathcal{N}(0, 1)$ or uniform $\mathcal{U}(-1, 1)$ distribution for each contributor j . The right-hand side denotes the correlated error term, where a single sample $x_{t,j}$ is drawn for all the pixels and scaled by the S2-RUT derived uncertainty. The correlation

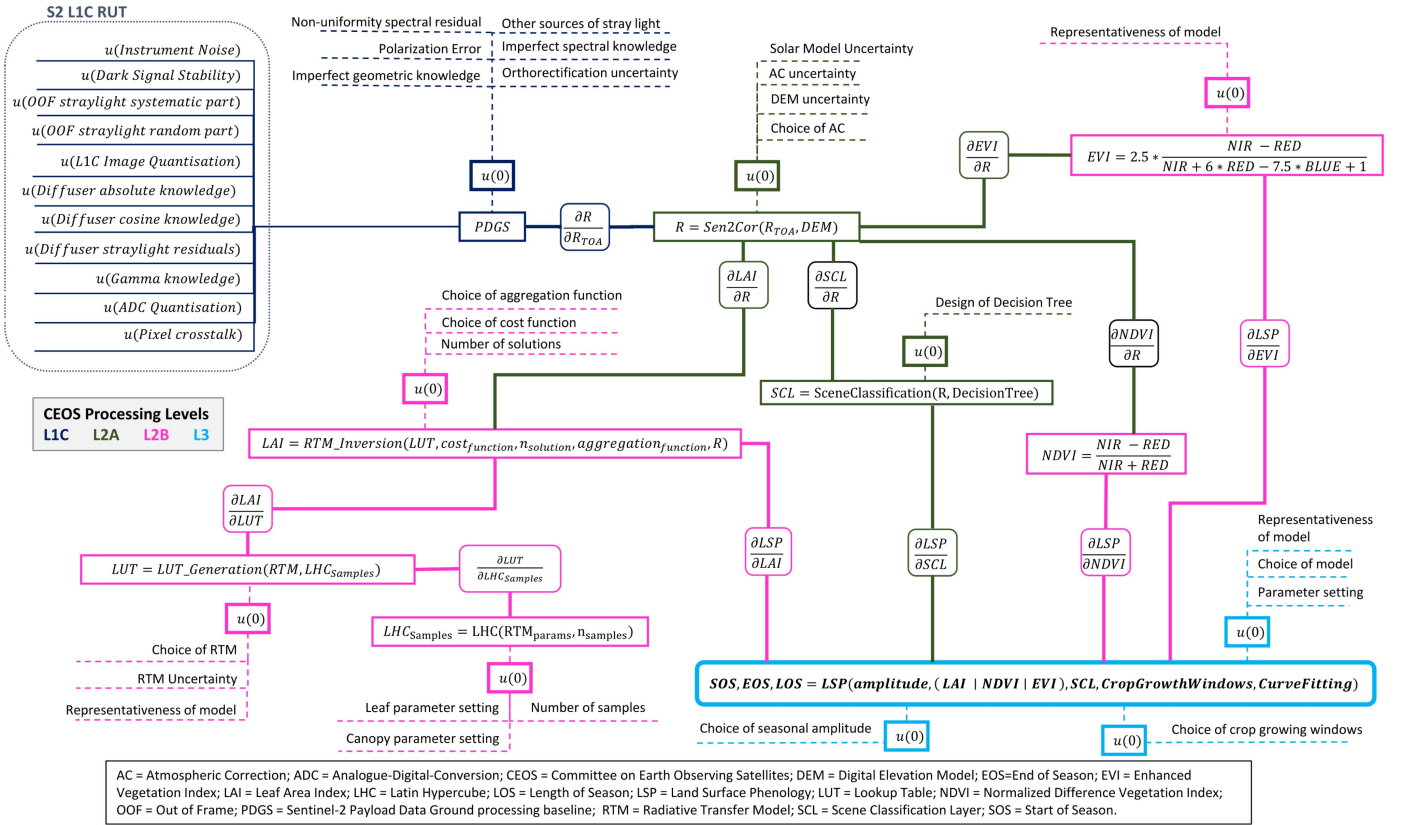


Fig. 4. Uncertainty tree diagram of uncertainty in LSP metrics derived from multitemporal S2 data (blue box). Processing levels defined by the CEOS are color-coded starting at L1C (dark blue) and ending at the LSP metrics (L3). Partial derivatives represent sensitivity coefficients showing how uncertainty propagates from certain source effects into the end product. Uncertainty contributors not considered in this study are indicated by $\mu(0)$.

coefficient α takes values between 0 and 1, where $\alpha = 0$ means that the error is completely independent. When α is 1, then an error is fully correlated between pixels and spectral bands. The values for α within the three domains for the single uncertainty contributors are based on correlation coefficients reported by [38, Table I].

While the MC framework is rather easy to implement, a sufficiently high number of realizations (scenarios) are essential to obtain reliable results. Running Sen2Cor on a single scene took approximately 15 min on a powerful workstation running under Linux Fedora 34 (16 Core AMD Ryzen Threadripper PRO 3955WX 3.9 GHz, 128-GB DDR4 3200 MHz, SSD drives). To determine the necessary number of scenarios in the MC framework, we selected three S2 scenes at different times of the year and created 1000 L1C scenarios for each of the scenes using the aforementioned MC framework. Then, we applied the image processing chain (see Section IV-D). Next, we calculated the uncertainty in each of these variables. We started calculating the uncertainty using two scenarios and iteratively increased the number of scenarios up to 1000. Then, we plotted the retrieved absolute uncertainty values (per target variable and different crop types) against the number of scenarios considered. We determined a number of 300 scenarios as an optimum threshold after which the derived uncertainty converged against a fixed value.

D. S2 Processing Chain

Each L1C TOA scenario is processed using classical remote sensing image processing, which underlies most LSP studies. First, AC is performed to minimize atmospheric influences. The atmospheric-corrected data are then used in a second step to determine the NDVI, EVI, and GLAI [see Fig. 3(a)]. The standard deviation of these across all the scenarios of an S2 scene gives its standard uncertainty ($k = 1$). This is determined from all the scenes and used to generate time-series scenarios from which the LSP metrics are calculated. Besides AC, all the processing steps were carried out using the open-source Python Earth Observation Data Analysis Library [46]. The code required to rerun the uncertainty propagation chain and subsequent is available publicly under GNU 3.0 license.⁵

1) *Atmospheric Correction*: All the scenes and their MC scenarios in L1C processing level were converted to BOA (L2A) reflectance factors using Sen2Cor v2.9 [47], [48]. While there are many software packages available for AC of S2 scenes, we selected Sen2Cor because it is integrated by the operational Copernicus S2 ground segment and some of the DIAS platforms.

⁵[Online]. Available: <https://doi.org/10.5281/zenodo.6669854>

TABLE II
KEY CROP GROWTH PERIODS DEFINED BASED ON AGRONOMIC KNOWLEDGE TO CONSTRAIN THE TEMPORAL RANGE CONSIDERED PER CROP TYPE BETWEEN A START AND END DATE (YYYY-MM-DD)

Crop type	Start	End
Winter wheat	2019-02-01	2019-08-15
Winter barley	2019-02-01	2019-08-15
Grain maize	2019-05-01	2019-10-15
Silage maize	2019-05-01	2019-10-15
Soybean	2019-04-01	2019-10-15
Potato	2019-04-01	2019-10-15
Rapeseed	2019-02-01	2019-08-15
Permanent grassland	2019-02-01	2019-10-15
Extensively used grassland	2019-02-01	2019-10-15
Sugar beet	2019-04-01	2019-10-15
Sunflower	2019-04-01	2019-10-15

In addition, the widely used Google Earth Engine [49] platforms provide Sen2Cor-processed S2 L2A data.

It is important to note that we used the default configuration of Sen2Cor for all runs. Thus, we propagated the radiometric uncertainty through the AC without introducing uncertainty of the AC itself (see Fig. 4). Therefore, we do not further address the uncertainties of the AC and its parameters (water vapor and aerosol optical depth). Quantifying the uncertainty of the AC itself is a complex task far beyond the scope of this article. Currently, a follow-up paper describing AC uncertainty quantification is in preparation by Gorrion et al. The source code is already available online.⁶

2) *Uncertainty of the SCL*: For each LIC MC scenario, an SCL result is available after AC (see Section IV-D1). Again, we focus on the effect of propagated radiometric uncertainty and do not consider, e.g., the design and parameterization of the classifier.

For each MC scenario of an S2 scene, the percentage of pixels belonging to one of the 12 SCL classes was computed. Then, the mean and standard deviation across all MC scenarios of a scene was calculated per SCL class giving a quantitative estimate how radiometric uncertainty influenced the number of pixels belonging to a certain class. In addition, for each pixel, a confidence score C_i was provided for the SCL class i the majority of the MC runs ($N = 300$) voted for. A confidence score of 100% for SCL class i of a pixel indicates that all N MC scenarios had the same SCL class, whereas lower percentages indicate that a certain share of j ($0 \leq j \leq N$) MC scenarios resulted in different SCL classification outputs

$$C_i = \frac{N - j}{N}. \quad (2)$$

3) *VIs and Parameters*: Using the Sen2Cor-derived L2A BOA reflectance factors, we calculated two widely used spectral VIs, which were recently reported to have been used in more than 80% of 445 LSP studies [2]. Using the SCL product,

pixels were masked if not classified as Class 4 (vegetation) or 5 (nonvegetated, linked to bare soil and brown vegetation).

The most common index is the NDVI [32], which is sensitive to canopy greenness. We calculated the NDVI using S2 bands B04 (red) and B08 (near-infrared), which both have a spatial resolution of 10 m

$$\text{NDVI} = \frac{B08 - B04}{B08 + B04}. \quad (3)$$

In addition, we calculated the three-band EVI [33] that utilizes the blue band (B02) in addition (spatial resolution 10 m). In contrast to the NDVI, the EVI is less prone to saturation effects for high biomass values and has reduced sensitivity to background and atmospheric effects. However, this comes at the price of empirical coefficients, which are an additional source of uncertainty (see Fig. 4, denoted as “representativeness of model”)

$$\text{EVI} = 2.5 \frac{B08 - B04}{B08 + 6B04 - 7.5B02 + 1}. \quad (4)$$

Both the NDVI and the EVI are canopy greenness proxies whose development is linked to changes in plant physiology [8], [50]. The NDVI and EVI take values between -1 and $+1$. The higher the value, the more green (i.e., healthy) vegetation is.

For GLAI retrieval, we developed a lookup table (LUT)-based inversion approach of the four-stream RTM PROSAIL [51] using PROSPECT-D as leaf model [52]. We created a large LUT for each S2 scene with 50000 entries using a Latin hypercube sampling scheme. The range of input leaf and canopy parameters were limited to upper and lower bounds suggested by Danner et al. [53]. Sun and observer angles were set to scene-specific values. By comparing the PROSAIL-simulated S2 BOA spectra in the LUT with satellite observations by means of the root-mean-square error (RMSE), we retrieved the LAI as the median of the 100 best-fitting solutions.

As with the previous steps in the processing chains, there are further sources of uncertainty that cannot be addressed within the scope of this work: PROSAIL, for instance, has an inherent uncertainty, while the size of the LUT as well as the underlying sampling strategy most likely has an influence on the inversion

⁶[Online]. Available: https://www.researchgate.net/publication/371653323_A_software_tool_for_the_estimation_of_uncertainties_and_spectral_error_correlation_in_Sentinel-2_Level-2A_data_products

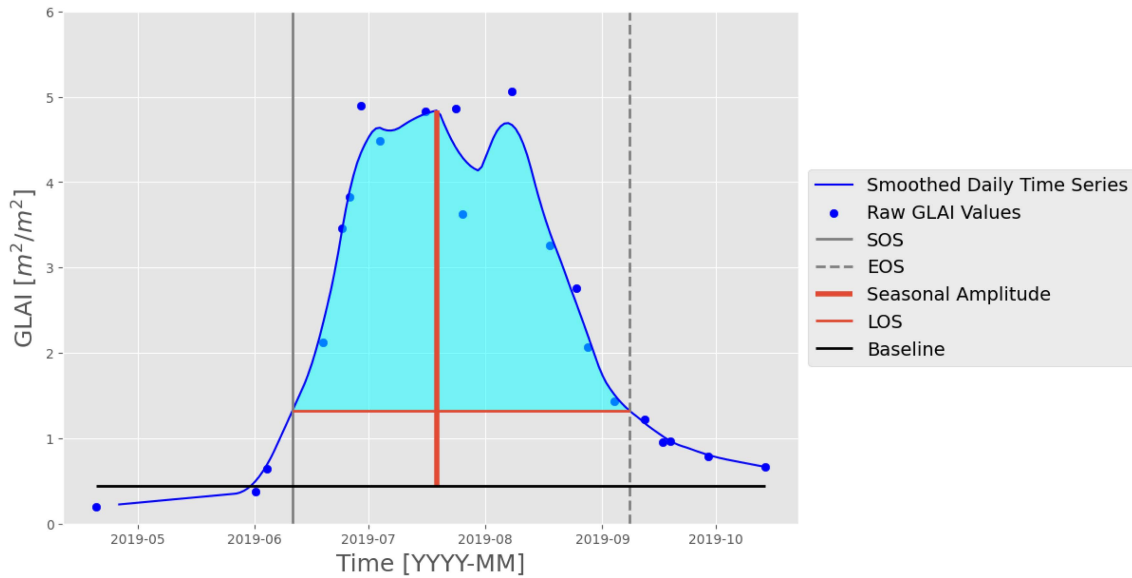


Fig. 5. Example GLAI time series from a single pixel in daily resolution (blue solid line) derived from raw GLAI values (blue dots) showing SOS (gray solid line), EOS (gray dashed line), and LOS (red horizontal line) for a single sunflower pixel using a threshold of 20% seasonal amplitude (bold vertical red line). The area colored in cyan between the GLAI curve and the LOS line represents the actual growing season.

result like the inversion strategy itself. Likewise, another RTM could have been used instead of PROSAIL. These effects are denoted with $\mu(0)$ in Fig. 4.

4) *Time-Series Generation and Phenology Extraction:* For each crop, growing periods were defined based on agronomic knowledge, since the period studied (February to October) is longer than the growing period of most of the crops considered. For example, winter wheat and barley are harvested in Switzerland by mid to late July at the latest, whereas maize or soybean are not sown before mid-April. Therefore, S2 observations before or after the main crop growth period were discarded. Table II shows the chosen crop growing periods for all crop types available. In addition to agronomic knowledge, start and end dates were adjusted manually based on S2 data availability (see Fig. 2) to ensure that sufficient data points were available at the beginning and end of the growing season.

For the time-series analysis, an approach based on the widely used TIMESAT software [54] (v3.3) was chosen, which ensures the comparability of this work with other LSP studies [see Fig. 3(b)]. We used the Python package “phenology” available under Apache-2.0 license that implements the TIMESAT v3.3 approach in Python [55]. Subsequently, a simple outlier detection was performed with a moving median with a window width of five subsequent observations. Data points that deviate more than the user-defined cutoff value from the median of the window and are also smaller than the mean of the immediate two neighbors minus the cutoff value are identified as outliers. In this case, the cutoff value was set to two standard deviations. The remaining data points were interpolated linearly to obtain daily values. The outlier removal was applied after filtering by SCL. Overall, the number of outliers can be considered low ($\leq 10\%$ of

all data points), since mainly cloud-free scenes were used. In a further step, the time series was smoothed with a Savitzky–Golay filter [56], using a window size of 11 days and a first degree polynomial. Savitzky–Golay is a widely used method in remote sensing studies for smoothing time series [57], [58], [59]. Again, the uncertainty arising from the outlier removal and curve fitting is denoted with $\mu(0)$ (see Fig. 4).

The LSP metrics SOS and EOS were calculated using a 20% threshold in seasonal amplitude. The seasonal amplitude is defined as the difference between the maximum value of the time series and a baseline. The baseline is given by the mean value of the minimum before and after the maximum of the time series. The SOS is the date when the time series reaches 20% of the amplitude and the EOS is the date when the time series falls below this value after the maximum for the first time [2], [60]. The length of season (LOS) can be calculated as the difference between the timing of SOS and EOS in days. Of course, the choice of seasonal amplitude is also a source of uncertainty which we set to zero in this study (see Fig. 4).

Fig. 5 exemplifies the concept of the seasonal amplitude for a single sunflower pixel in the study area. The blue dots denote the raw GLAI values derived from ProSAIL, whereas the blue solid line shows the result after outlier removal and smoothing using Savitzky–Golay in daily resolution.

E. Time-Series Scenarios

The methodology described in Section IV-D4 was embedded into a second MC framework [see Fig. 3(c) and (d)] to generate time-series scenarios. MC sampling was performed to create the time-series scenarios ($N = 1000$) for propagating radiometric

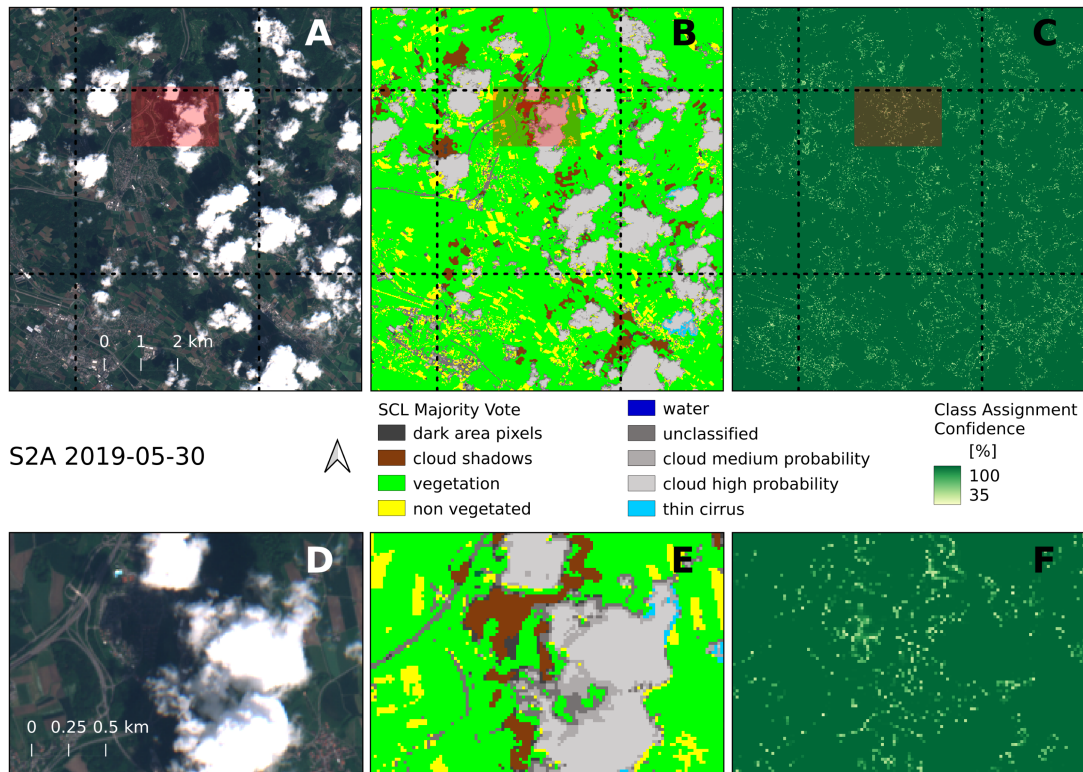


Fig. 6. (a) RGB L1C TOA true-color composite, (b) SCL majority vote, and (c) SCL class assignment confidence scores for an S2A scene acquired on May 30, 2019. The more greenish the color in (c), the higher the confidence score and the lower the uncertainty. A small-scale detail plot [red rectangle in (a)–(c)] highlighting the higher uncertainties at class boundaries is depicted in (d)–(f).

uncertainty into LSP metrics. Pixel time series were generated using the original EVI, NDVI, and GLAI values and their uncertainty derived from Section IV-D3.

For scenario generation, two different types of error correlation (see also Section IV-C) were assumed. Since it is currently unknown to what extent systematic effects have an impact between the individual S2 scenes, two different approaches were run, reflecting the two possible extreme cases:

- 1) full correlation ($\alpha = 1$) between the S2 scenes, i.e., the determined uncertainty in the VIs or the GLAI affects all dates equally per pixel [see Fig. 3(c)];
- 2) zero interscene correlation ($\alpha = 0$) where the single S2 scenes are completely uncorrelated and the uncertainties of the dates of a pixel are, thus, stochastically independent [see Fig. 3(d)].

Errors can be correlated due to systematic effects such as straylight during the sensor (re)calibration. Instrument noise or analog-to-digital quantization, in turn, are uncorrelated between scenes and, therefore, favor the uncorrelated assumption.

V. RESULTS

A. SCL Uncertainty

Radiometric uncertainty imposed uncertainties in SCL class assignment. Fig. 6 shows a true color representation of the study

area taken by S2A on May 30, 2019 [see Fig. 6(a)] and the corresponding SCL calculated from the majority vote of all 300 MC realizations [see Fig. 6(b)]. Fig. 6(c) shows that the SCL class membership was mostly close or equal to 100%, i.e., the pixel-based classification was the same in all 300 scenarios. In particular, areas with homogeneous spectral properties, such as vegetated areas, but also the central regions of the cumulus clouds and their shadows had a high confidence (100%), thus a low classification uncertainty. At the transition area of two classes, e.g., at the transition from cloud to cloud shadow or cloud shadow to vegetation, lower confidence values (smaller than 48% in some cases) and thus higher uncertainty due to spectral mixing effects caused by the spatial resolution of 20 m of the SCL product and adjacency effects were shown. This is clearly evident from the small-scale detail plot in Fig. 6(d)–(f). Depending on the cloudiness of a scene the relative number of pixels where the class assignment confidence was $<100\%$ was 88.2% (2019-05-30, cloudy pixel percentage: 17%) to 94.6% (2019-06-19, cloudy pixel percentage: 0.6%).

Throughout the entire season, average class confidence was highest for vegetation ($99.3 \pm 0.8\%$), clouds with high probability ($96.4 \pm 1.7\%$), nonvegetated pixels ($95.9 \pm 1.7\%$), and thin cirrus ($95.6 \pm 3.8\%$) followed by clouds with medium probability ($92.0 \pm 2.3\%$). Also, the water ($95.1 \pm 1.8\%$), unclassified ($91 \pm 2.4\%$), and dark area pixel class ($93.3 \pm 2.2\%$)

TABLE III
MEAN PERCENTAGE OF PIXELS FOR SELECTED SCL CLASSES ± STANDARD DEVIATION DERIVED FROM 300 MC REALIZATIONS FOR FIVE S2 SCENES

sensing date	cloud shadows	vegetation	non vegetated	cloud medium probability	cloud high probability
2019-02-14	6 ± 4.0	52.1 ± 0.4	6.8 ± 1.5	1.0 ± 4.0	0.0 ± 5.0
2019-04-20	0.0 ± 25.0	83.3 ± 0.1	13.0 ± 0.6	0.4 ± 3.3	0.0 ± 4.0
2019-05-30	3.5 ± 0.9	65.2 ± 0.2	7.7 ± 1.0	4.3 ± 1.4	12.4 ± 0.4
2019-07-16	2.0 ± 0.8	76.8 ± 0.1	11.7 ± 0.7	2.3 ± 2.0	3.8 ± 1.2
2019-09-19	(0.0 ± 1.1)e+02	85.0 ± 0.1	11.1 ± 0.6	0.4 ± 3.0	0.0 ± 5.0

The overall number of pixels per scene is 360 000 (20 × 20 m spatial resolution). The complete table including all SCL classes and S2 scenes can be found in Table IV in the Appendix.

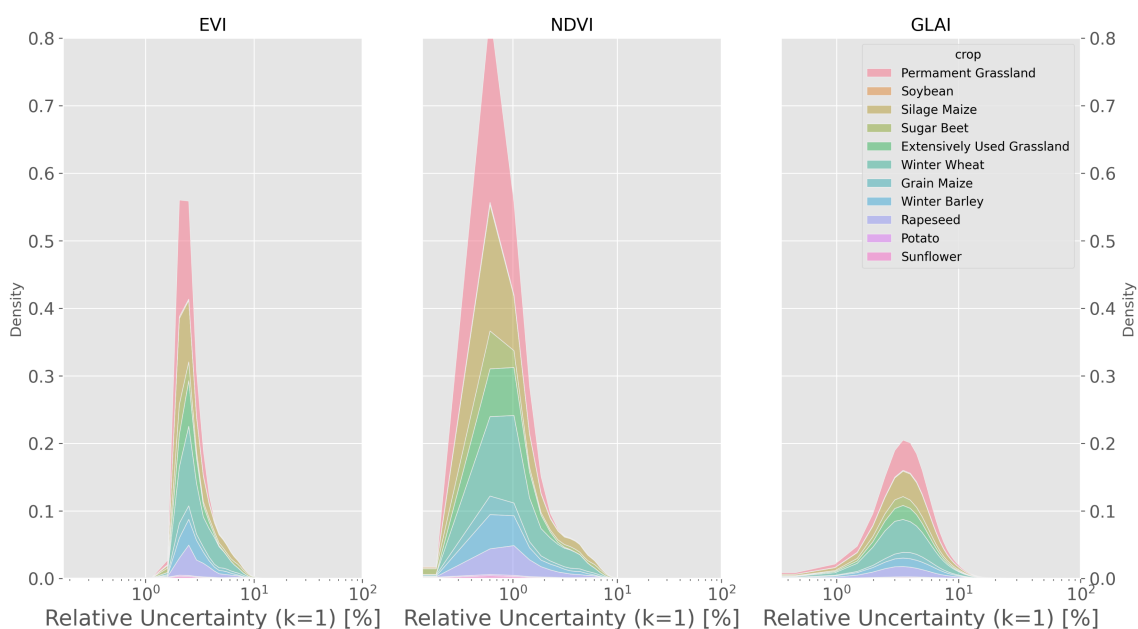


Fig. 7. Kernel density based distributions of MC-derived relative uncertainty values in EVI (left), NDVI (middle) and GLAI (right) considering all available crop pixels and S2 scenes excluding pixel observations not classified as “vegetated” (SCL class 4) or “nonvegetated” (SCL class 5). Contributions to relative uncertainty per crop type are color coded. To improve readability, the x-axis has been log-scaled.

scored high on average. These classes, however, were less abundant in the S2 scenes (see, for instance, Fig. 6). Cloud shadows (84.9 ± 2.6%) had the lowest average confidence score.

Looking at the relative uncertainty in the percentage of pixels (%) assigned to an SCL class per S2 scene, we noted cloud classes (medium and high probability) to have a higher relative uncertainty of up to 4% than vegetation (≤ 0.1%) and bare soil (≤ 1.5%). Table III shows the uncertainty in the relative number of pixels per class in relation to all S2 pixels (360 000 pixels, 20-m spatial resolution) for five selected scenes including the scene from May 30 (see Fig. 6). The complete table can be found in Table IV in the appendix. The scenes were selected to represent different important times in vegetation development, such as the end of winter dormancy, advanced spring, early and

mid-summer, and incipient fall. Pixels from one of the two cloud classes often only change from medium to high cloud probability and vice versa, whereas vegetation pixels are classified as cloud shadows in some scenarios. This is significant because unrecognized cloud shadows are among the main causes of outliers in vegetation time series.

B. NDVI, EVI, and GLAI Uncertainty

Fig. 7 shows the kernel-density-based distributions of MC-derived relative uncertainty values for EVI (left), NDVI (middle), and GLAI (right) per crop type (color-coded). All pixels marked as SCL class 4 (vegetated) or 5 (nonvegetated) have been included. To improve readability, the x-axis in Fig. 7 has been log-scaled.

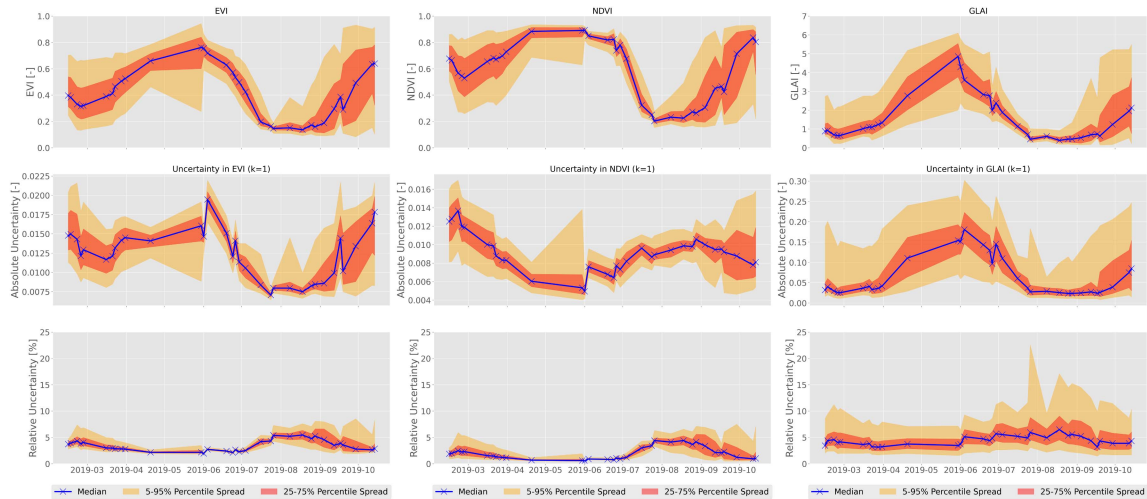


Fig. 8. Spatiotemporal variability of EVI (left column), NDVI (mid column), and GLAI (right column) for all pixels annotated as winter wheat (228.8 ha). For each S2 scene, the median value (blue line), central 50% (red), and 90% spread (orange) across all pixels is shown not classified as cloud, shadow, or snow based on the SCL data of the original S2 outputs after Sen2Cor. The top row shows the spatiotemporal variability in the actual EVI, NDVI, and GLAI derived from the S2 scenes. In the mid row, the absolute uncertainties derived from the MC simulations are depicted. The bottom row shows the uncertainties in relative terms.

Overall, NDVI was less uncertain than EVI. The spectral indices exhibited lower relative uncertainties than GLAI. Moreover, uncertainty distributions of EVI and NDVI revealed a narrow spike and delta-like shape with little dispersion, while the maximum of the uncertainty distribution in GLAI was flatter and less pronounced. The distribution of NDVI relative uncertainties (see Fig. 7, middle) showed a sharp peak for all crop types between 0% and 1%. Values greater than 10% did not occur. Median uncertainty in NDVI for all crops was about 1.1% ranging from 0.8% for permanent grassland to 2.4% for potato. In contrast, the uncertainties in EVI were slightly higher (see Fig. 7, left) as the spike was shifted in positive x -direction. In detail, the maximum of the EVI uncertainty distribution laid between 1% and 10%. Median uncertainty in EVI for all crops was about 2.7%, ranging from 2.4% for permanent grassland to 3.8% for potato. Isolated outliers, however, assumed uncertainties larger than 10%. GLAI (see Fig. 7, right) exhibited a higher share of uncertainty values greater than 10%, while the peak of the distribution was located between 1% and 10%. The median of the uncertainty distributions for all crops was about 4.4% ranging from 4.1% for extensively used grassland to 5.0% for potato, which is again the crop with highest uncertainties.

For all crop types, the relative uncertainty was lowest when the canopy was greenest (high EVI or NDVI values) or reached the largest leaf area values (GLAI). Fig. 8 shows the spatiotemporal variability in observed EVI, NDVI, and GLAI values and their uncertainties for winter wheat (228.8 ha). In the top row in Fig. 8, the median (blue line), central 50% (red fill), and central 90% percentile (orange) of all pixels not classified as cloud, cirrus, cloud shadow or snow was shown for each S2 scene. In the same way, the MC-derived absolute uncertainties were displayed in the mid-row of Fig. 8 alongside relative uncertainty values in the bottom row. Plots for the other crops were available

in the Appendix (see Figs. 12–21) as well as pixel time series from randomly selected pixels (see Fig. 13).

Throughout the growing season, median EVI (see Fig. 8, upper left) showed a moderate increase through spring until maximum EVI values were reached in early June. This increase was followed by a steeper decline to median EVI values below 0.2 at the end of July. Subsequently, the median EVI increased again. A similar picture was shown in median NDVI (see Fig. 8, top center), which started with higher values (between 0.7 and 0.5) and remained almost constant at a high level (median NDVI > 0.8) indicating saturation between mid-April and the end of June. GLAI (see Fig. 8, upper right) showed low values at the beginning of the growth period (median GLAI around $1 \text{ m}^2 \text{ m}^{-2}$) and a maximum ($4.5 \text{ m}^2 \text{ m}^{-2}$) at the beginning of June. The gradient before and after this maximum was more symmetrical than for EVI and NDVI.

Absolute uncertainties of the winter wheat pixels (see Fig. 8, middle row) showed a temporal pattern, which partly reflected vegetation development. In the case of EVI (left), the absolute uncertainties decreased first and increased moderately until the end of June (median uncertainty at the beginning of June about 0.019). Consistent with the EVI values, the uncertainty reached a minimum at the end of July (median about 0.0075) and then increased again. In NDVI (middle), absolute uncertainties decreased steadily, reaching a minimum in early June (median uncertainty around 0.005) and then increasing again, reaching lower values than at the beginning of the growing season (around 0.009). The GLAI (right) showed a different behavior: The absolute uncertainty curve increased similarly to the median GLAI values from $0.05 \text{ m}^2 \text{ m}^{-2}$ in February to the beginning of June (around $0.17 \text{ m}^2 \text{ m}^{-2}$ median uncertainty) and then decreased to values lower than $0.05 \text{ m}^2 \text{ m}^{-2}$ in August. From September on-wards, the uncertainties increased again. The spatial variability of the absolute uncertainties was also

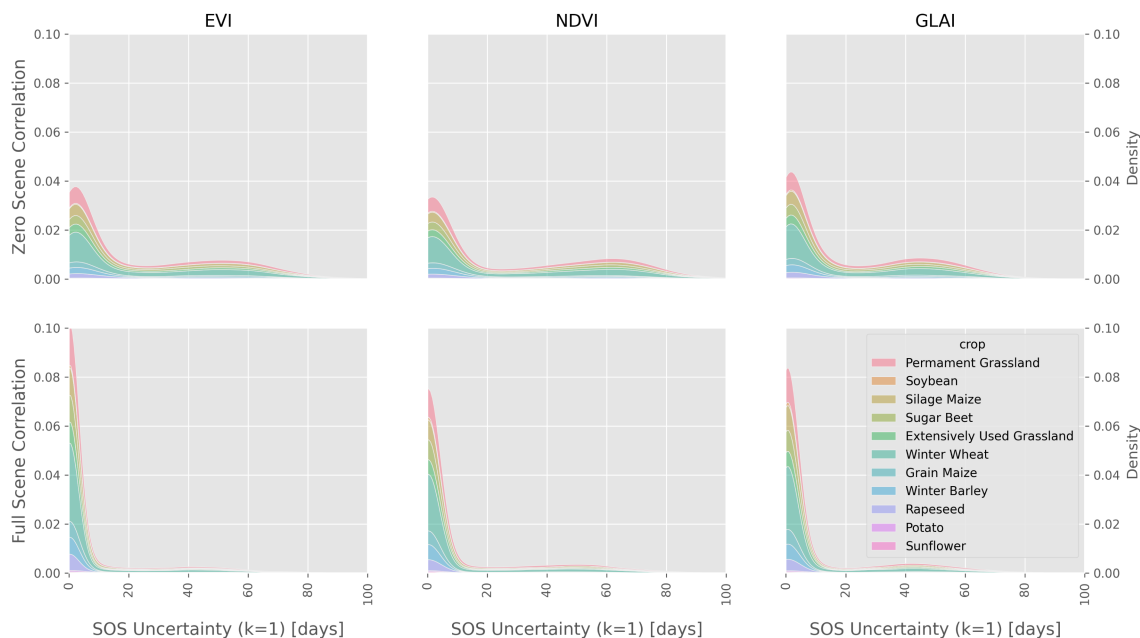


Fig. 9. Kernel-based relative uncertainty distributions in SOS for EVI (left column), NDVI (middle column), and GLAI (right column) color-coded by crop type. The top row shows the results of the MC runs with zero interscene correlation; the bottom row the results assuming full interscene correlation.

variable over time. Moreover, the spatial variability in absolute uncertainty was lowest in NDVI and EVI, followed by GLAI.

Relative uncertainties (see Fig. 8, bottom row) were lowest in NDVI, followed by EVI. Regarding NDVI, the median relative uncertainty was constantly below 5% and reached the 1% mark at the end of May. At this stage, the S2 band B08 dominated the ratio [see (3) and (4)]. The spatial variability of the relative uncertainty of NDVI was largest at the beginning of the growth period (February to the end of March) and declined to less than 1% until the beginning of July. The spatiotemporal pattern of relative uncertainty in EVI was similar, but relative uncertainties were higher on median than in NDVI and exceeded the 5% mark in August. In GLAI, the relative uncertainties were consistently higher than in the spectral indices, and median values were around the 5% mark. The spatiotemporal pattern was less evident in GLAI than in the spectral indices and did not show a clear decrease as the growth period progresses.

C. LSP Metrics Uncertainty

1) *Start of Season*: Relative standard uncertainty distributions for SOS are shown in Fig. 9 color-coded by crop type. The top row denotes the results when the interscene correlation is set to $\alpha = 0$, whereas the bottom row shows the results for full interscene correlation ($\alpha = 1$). Overall, the assumption of zero interscene correlations causes SOS uncertainties to be clearly higher and more dispersed than in the case of full interscene correlation.

In zero interscene correlation (see Fig. 9, top row), the uncertainty distributions showed two distinct properties for all crop

types and time-series sources. First, the uncertainty distributions showed a peak shifted from zero in the magnitude of two to five days. Median uncertainties ranged from one (NDVI in Soybean and GLAI in Potato) to eight days (EVI in rapeseed and sunflower). Second, a secondary, although less pronounced, peak with a maximum around 60 days was apparent in all three plots. In contrast to the more pronounced main peak, which had a steep right flank, the secondary peak had very flat shoulders that taper off to uncertainties of 80 days and greater.

In full interscene correlation (see Fig. 9, bottom row), only a single peak at 0 was evident, which was accompanied by a very steep right slope. The median uncertainty was less than one day except for the grasslands and sunflower in the GLAI, where the median was one day. 95% of all uncertainty values were between 24 days (EVI in potato) and 59 days (NDVI in extensive grassland, winter wheat, and silage maize). The second peak, which was evident in the uncorrelated case, was also evident here, although it was much less pronounced and had a maximum between 40 and 50 days.

2) *End of Season*: Analogous to SOS, the propagated uncertainty in EOS is shown in Fig. 10. As for SOS, the uncertainty in the fully correlated case (see Fig. 10, bottom row) was clearly smaller than in the uncorrelated case (see Fig. 10, top row).

The uncertainty distribution in the uncorrelated case (see Fig. 10, upper row) showed a singular peak, which had a maximum between two and 20 days depending on crop type and time-series source. The median uncertainty varied between two days (GLAI in winter barley) and 16 days (EVI in soybean). The secondary peak evident in SOS (see Fig. 9) was not present. It follows that 95% of the uncertainty values ranged

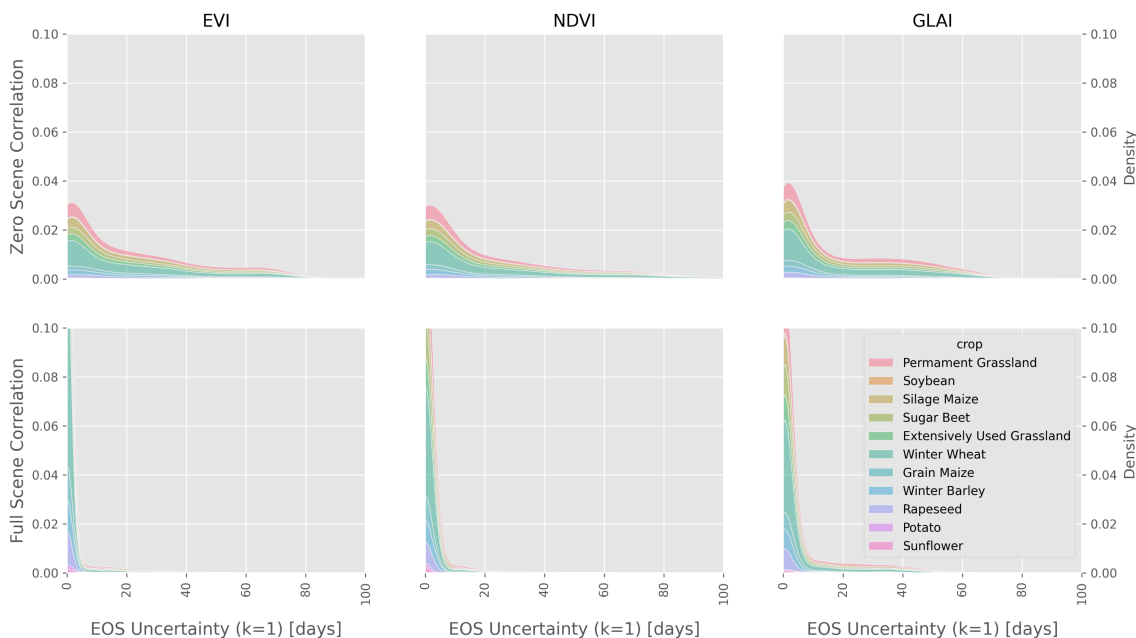


Fig. 10. Kernel-based relative uncertainty distributions in EOS for EVI (left column), NDVI (middle column), and GLAI (right column) color-coded by crop type. The top row shows the results of the MC runs with zero interscene correlation; the bottom row the results assuming full interscene correlation.

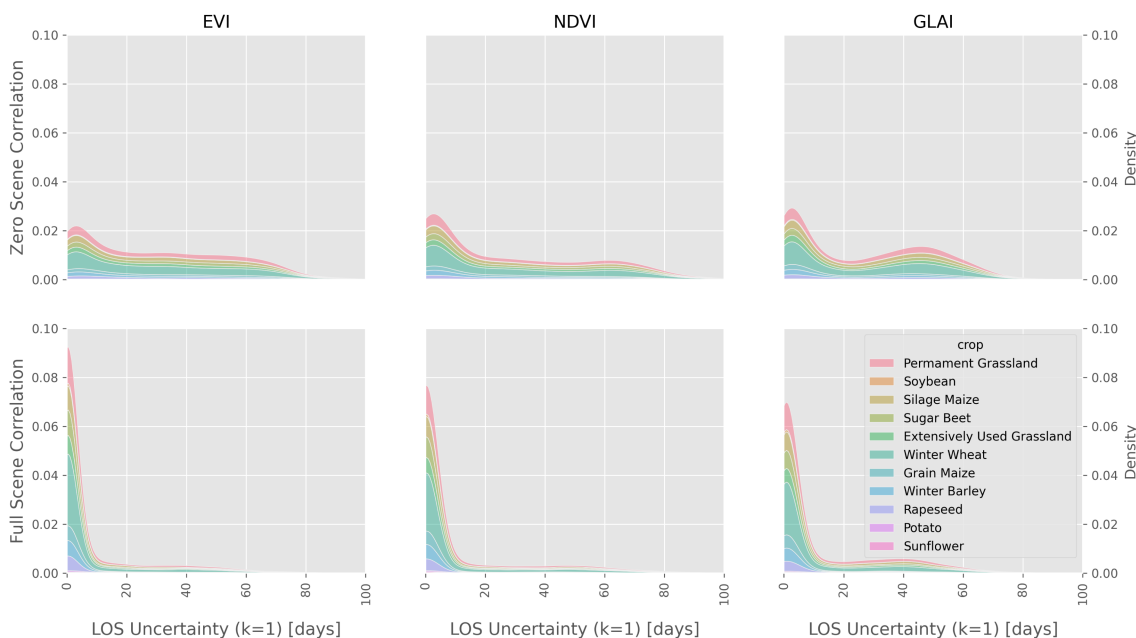


Fig. 11. Kernel-based relative uncertainty distributions in LOS for EVI (left column), NDVI (middle column), and GLAI (right column) color-coded by crop type. The top row shows the results of the MC runs with zero interscene correlation; the bottom row the results assuming full interscene correlation.

from 45 days (GLAI in sunflower) to 101 days (NDVI in silage maize).

Under the fully correlated assumption (see Fig. 10, bottom row), the uncertainty distributions showed a sharp peak at zero with a very steep right shoulder. EOS uncertainties were on median smaller than one day in all cases. The 95% percentile values ranged from zero days in the case of potato from EVI to

51 days for the same crop in GLAI. GLAI 95% percentiles were significantly higher for all crops than in the case of NDVI and EVI.

3) *Length of Season*: Uncertainty in LOS (see Fig. 11) reflected shifts in SOS and EOS due to propagated uncertainties. Again, in the fully correlated case (see Fig. 11, bottom row), uncertainties were low, and hence, the number of days the

seasons gets longer or shorter was small, whereas the larger uncertainties in case of zero interscene correlation caused larger shifts in LOS (see Fig. 11, top row).

In the uncorrelated case (see Fig. 11, top row), all three time-series sources showed a peak between one and 20 days with median uncertainties ranging from nine days (NDVI in potatoes) to 28 days (EVI and GLAI in the grasslands). The right shoulder of the uncertainty distributions was broad and declined only smoothly toward zero. Thus, 95% of the uncertainties were between 60 days (EVI in sunflower) and 78 days (EVI in permanent grassland and sugar beet). The secondary peak evident in SOS and EOS uncertainty distributions was pronounced in GLAI but not in EVI and NDVI.

LOS uncertainties were low in case of full interscene error correlation. The uncertainty distributions (see Fig. 11, bottom row) showed a peak close to zero with a broader dispersion of values toward higher uncertainties in the case of GLAI. Median uncertainties were less than one day for all crops for the two spectral indices and one day in the case of GLAI. The values of the 95% percentile ranged from 24 days (EVI in potatoes) to 59 days (NDVI in permanent grassland).

VI. DISCUSSION

A. Interpretation of Results

1) *SCL Uncertainty*: The SCL product is robust to radiometric uncertainty. High uncertainties only occur at class boundaries (see Fig. 6). Namely, the cloud shadow class showed the lowest confidence score, i.e., the highest class assignment uncertainties. We assume that the spectral indices and thresholds used were selected so that a broad number of different spectra are assigned to a class [47]. The class “Vegetation” (SCL class 4), for instance, includes contrasting vegetation types such as coniferous forests or crop canopies. The modification due to the radiometric uncertainty is smaller than the spectral within-class variability. Consequently, radiometric uncertainty has little influence. Only at the transition between classes, where spectral mixing effects are expected or vegetation is overlaid by translucent clouds does uncertainty cause changes in class assignments. Although the uncertainty analysis does not provide an indication of the accuracy of the SCL product, the areas identified as uncertain in the S2 scenes are consistent with findings from validation studies: Louis et al. [61] attested only limited accuracy of Sen2Cor in more complex scenarios, especially for cloud shadow detection, and thus under conditions where uncertainty is also higher. Liu et al. [62] compared different cloud detection algorithms in S2 data and found that the SCL product gave inadequate results when it comes to cloud delineation. For example, translucent cloud edges were misclassified, which is consistent with the finding that pixels in these regions have high SCL uncertainty. This suggests that applying a spatial buffer around cloud and shadow pixels might lower the occurrence of undetected atmospheric artifacts.

2) *VIs and GLAI Uncertainty*: EVI and NDVI (see Figs. 7 and 8) have clearly lower relative uncertainty and show smaller spatiotemporal variability compared to GLAI.

In the NDVI, we assume band rationing [see (3)] to cancel out random errors, such as those from sensor noise, to a large degree. Still, normalization is accompanied by saturation at high biomass levels [63]. Nonrationing indices like EVI avoid saturation and have an increased sensitivity at high biomass values [33]. However, a basic assumption of the EVI is that the satellite data are mostly noise-free. Due to inherent radiometric uncertainties in the blue, red, and infrared channels, this assumption is not fully met. Over vegetated areas, random sources of uncertainty dominate in the blue and red bands, where systematic effects play a smaller role due to the low radiance level. Systematic effects originate, for example, from straylight during sensor calibration. In the NIR band, systematic and random effects occur nearly equal due to the higher radiance level of green vegetation. Therefore, the EVI is influenced by both systematic and random sources of uncertainty. Consequently, the higher relative uncertainties in EVI compared to the NDVI can be explained. In the NDVI, the saturation effect at high biomass values causes the uncertainty of the NIR band alone to dominate during this phase. For low biomass levels, the red band additionally contributes to the systematic uncertainty component. Thus, relative uncertainty in the NDVI reaches a minimum and maximum at the time of maximum and minimum green biomass accumulation, respectively. In the EVI, all three bands contribute to the total uncertainty during the period of maximum greenness. The decrease of the absolute uncertainty in the EVI before and after the greenness maximum (see Fig. 8, middle row left) can be explained by the removal of the canopy background targeted in the EVI formula [see (4)], which increasingly dominates the spectral properties as the plants mature. In mathematical terms, the numerator becomes smaller and the denominator bigger. This limits the overall impact of uncertainties since the resulting EVI values are small in any case when the canopy is not at its greenness peak.

GLAI is a physiological parameter derived from the inversion of a physically based RTM. It is, therefore, based on the entire spectral information and, hence, the uncertainties from all S2 bands. Furthermore, the inversion of the radiative transfer equation is ill-posed [64]. Using the median of the 100 simulated pixel spectra with the lowest spectral RMSE is an approach to solve the inverse problem. We assume to eliminate random sources of uncertainty to some extent. Still, the RMSE is a similarity measure that gives equal weight to all spectral bands. Thus, bands dominated by random uncertainty effects (low radiance levels) are weighted the same as bands where random and systematic effects occur to similar degrees (red edge and NIR bands). We hypothesize that the increase in reflectance in the red-edge and NIR bands as the growing season progresses conditions a greater weighting of the systematic uncertainties of these, synchronously increasing the absolute uncertainty in GLAI.

Differences found between the crops revealed that potato exhibited the highest median uncertainty values, while the grasslands showed lowest median uncertainties. This can be

explained by the number of observations that include plant and soil background, which in case of potatoes is high because the plants are grown in dams that are separated by trenches with usually little green canopy cover. Simply put, row crops such as potatoes violate the assumptions of the turbid medium PROSAIL model. As explained by the example of uncertainty dynamics in wheat (see above), the relative uncertainty is higher at higher biomass levels. This is also evident in the plot of spatiotemporal variability of uncertainty in potato (see Fig. 22). Grassland (see Figs. 15 and 16) is always green, so relative uncertainties are constantly low.

3) *LSP Uncertainty*: Interscene correlation determines LSP uncertainty. If uncertainties are uncorrelated among S2 scenes, LSP uncertainties are higher if they are fully correlated. In other words, if uncertainties are fully correlated, the LSP retrieval approach based on seasonal amplitude thresholds is robust against propagated radiometric uncertainty, and the impact on LSP metrics is small.

Full interscene correlation ($\alpha = 1$) biases the time series. The series is shifted in either positive or negative y -axis direction by a constant factor [up- and downward arrows in Fig. 3(b)]. Although the absolute amount of the shift scales for each data point differently because uncertainties in the spectral indices and GLAI are not constant over time (see Section V-B), the shape of the curve is largely preserved. Furthermore, the impact on the LSP metrics is dampened by using a relative threshold instead of absolute numbers. In contrast, when assuming zero error correlation ($\alpha = 0$) among the S2 scenes, each data point in the time series can be shifted in a different direction along the y -axis. In addition, the shift is not by a constant factor for the entire time series. Consequently, different values for the seasonal amplitude may be obtained, and the timing of the LSP metrics can differ between MC scenario runs.

A closer look at Figs. 9–11 shows differences between the crop types and between EVI, NDVI, and GLAI. Depending on the shape of the growth curve, even small changes can have larger effects on the LSP calculation: Soybeans, for example, have a well-defined growth curve whose ascending and descending branches have a steep gradient. In contrast, the ascending branch of winter cereals is less well defined in the spectral indices, since the actual growth period starts in autumn of the previous year and the canopy appears green early in spring. Thus, if the growth curve is well accented, uncertainty has less influence because small changes in the time series values have less influence on LSP estimation.

This points to a problem with the TIMESAT approach: The seasonal amplitude is meaningful if the green-up and brown-down branches are symmetric and a single clear maximum exists. However, the approach reaches its limit in the presence of strong asymmetries and, thus, can produce implausible results for SOS and EOS. This could be an explanation for the secondary peak evident in the SOS and EOS uncertainty distributions (see Figs. 9 and 10): If time series are asymmetric or have multiple peaks, uncertainty can cause shifts in SOS and EOS by several tens of days, which is reflected in the uncertainty distribution. Since the secondary peak occurs

in all time series and all crops, we assume that incorrect labels (main crop reported by the farmer was wrong) of the parcels as well as spectral mixed pixel effects could have an influence.

4) *Educated Guess About Interscene Error Correlation*: The question that follows from Section VI-A3 is which of the two interscene error correlation assumptions is closer to reality. While we cannot give a definitive answer at this point, considerations of the multitemporal behavior of the uncertainty contributors in the S2-RUT (see box in Fig. 4, top left) allow a first guess based on [38] and [43].

For example, some sources of uncertainty are not correlated between scenes. These include the instrument noise, as well as the analog-to-digital quantization and the L1C image quantization into a 12-bit floating point (radiance measurement) and 16-bit integer system (stored reflectance factors), respectively. Furthermore, we assume that crosstalk and the stability of the dark signal are uncorrelated between the scenes as residuals of signal correction.

This contrasts with effects that are correlated between scenes. It is known that the random part of the out-of-field (OOF) straylight has a pattern on the scan line of the MSI detector array, which should be independent of the scene. Likewise, the systematic part of the OOF straylight is partially correlated between scenes, since the OOF only changes partially over time. For the diffuser absolute knowledge, there is a strong correlation over the entire time span since the preflight Bidirectional Reflectance Distribution Function (BRDF) model is common to all scenes. The interpolation of the BRDF at different solar angles is considered not to significantly modify the error correlation between scenes. In contrast, for the diffuse cosine effect associated with microvibrations and thermal cycling, the interscene correlation is broken with recalibration, which occurs approximately every 18 days. The gamma correction, which includes corrections for nonlinearity and nonuniformity, exhibits high correlation between scenes that are radiometrically similar because the uncertainty budget of this contributor depends on the radiance level. Consequently, the effect is less correlated for scenes that are radiometrically less similar. Therefore, the interscene correlation of gamma correction depends on the development of vegetation over time.

In summary, we assume that uncertainties between scenes tend to be correlated rather than uncorrelated. The actual uncertainties in the LSP metrics might, therefore, be more likely in the range of values from the MC run of the zero interscene correlation. Still, the degree of interscene correlation is most likely a function of time. This means interscene correlation might not be constant over the growing season and be coupled with the phenological development of vegetation. Thus, our assumption needs to be further investigated.

B. Significance of Radiometric Uncertainty in LSP Studies

As shown in Section V-C, the interscene error correlation has an impact on the uncertainty in the LSP metrics. In case of full error correlation among the S2 scenes, the uncertainties

are mostly small. In case of uncorrelated errors, the uncertainties are clearly higher. Especially for fast growing crops like maize, uncertainties in SOS of the order of ≥ 10 days imply large physiological differences between the determined SOS dates. This is important when LSP metrics are used to detect phenological shifts, for example, to study the effects of climate change on plant growth [65] or to quantify differences in the onset of phenological stages for wider areas [66] or among varieties of the same crop. If the uncertainties are of a similar order of magnitude or even significantly larger than the expected phenological shifts, in the worst case, no reliable statement can be made. For example, using MODIS time series for China for the period 2001–2014 as an example, Luo and Yu [67] were able to show that the variability in SOS and EOS expressed as a single standard deviation at the pixel level ranges from 0 to greater than 30 days. This is in the order of magnitude of the uncertainty for the case without error correlation (see Figs. 9 and 10, top row). Furthermore, the identified uncertainties in LOS are not negligible, especially when LOS is used as a proxy for vegetation productivity. For instance, Park et al. [68] reported an increase in LOS of 2.6 days per decade for continental-scale LSP retrieval in Northern America from Global Inventory Modeling and Mapping Studies NDVI. This increase in LOS is slightly larger than the uncertainties in the fully correlated case (see Fig. 11, bottom row) but clearly smaller than uncertainty estimates from the uncorrelated scenario runs (see Fig. 11, top row).

In addition, comparing radiometric uncertainty to other sources of uncertainties in LSP retrieval that have already been investigated is important. For example, the authors of [69] and [70] found that changes in spatial aggregation can cause differences in LSP metrics greater than 60 days, using mangrove forests and agricultural cropland, respectively. This exceeds the uncertainties identified from radiometry. In this regard, Helman [27] highlights the effect of spectral mixing due to coarse spatial resolution: Changes in species composition could cause a similar change in the spectral properties of a pixel as an actual change in LSP [71]. Another source of uncertainty is introduced by the choice of the time series model: Lara and Gandini [72] were able to show, using crop and grassland phenology from MODIS data, that uncertainties in SOS and LOS can range from 20 to 50 days based on the choice and parameterization of the time series model.

C. Limitations

As shown in Fig. 4, radiometry is one out of many sources of uncertainty that make up the total uncertainty budget of LSP metrics. Since radiometric uncertainty affects all steps of EO processing chains, such as AC or scene classification, quantifying and propagating radiometric uncertainties is an important first step. However, this also means that in further steps, sources of uncertainty have to be quantified, which are currently ignored ($\mu(0)$ in Fig. 4).

In terms of computational requirements, the proposed MC framework is rather slow and resource intensive. Preferably, uncertainties would be propagated analytically using

the chain rule of uncertainty propagation. However, this requires knowledge about error covariance matrices and sensitivity coefficients. Currently, this information is hardly available in the EO domain, thus hampering the advancement of EO as a measurement science, as claimed by Mittaz et al. [36].

Furthermore, the link to agronomically relevant estimates of phenological development in crops such as the BBCH scale—a decimal scale for crop development comparable to the scale developed by Zadoks et al. [73]—is currently not given. Therefore, future research should focus on the impact of radiometric uncertainty on more advanced phenological metrics such as the start of heading in wheat [74]. With the present work, we provide a baseline to address these points thoroughly. Therefore, we published the calculated uncertainties in the L2B products (EVI, NDVI, and GLAI) as a dataset for this study.

VII. CONCLUSION

In this article, we proposed a framework for propagating radiometric uncertainties in S2 L1C TOA reflectance factors along an EO data processing chain into LSP metrics. The framework follows GUM specifications and demonstrates how uncertainty output from the S2-RUT can be propagated into higher level EO products. Not only did we trace radiometric uncertainties but also showed how radiometric uncertainty is embedded in the overall uncertainty budget of LSP metrics.

Our results for various agricultural crops reveal that NDVI and EVI are more robust to radiometric uncertainty than GLAI from RTM inversion. Still, this is not yet a statement on the suitability of spectral indices or physically based crop traits for phenology assessment. This applies to all crops suggesting that no crop-specific uncertainty exists. The S2 SCL product was mostly invariant to radiometric uncertainty, except for spectrally mixed pixels. These occur, for example, at cloud shadow edges or with translucent clouds. Furthermore, our results show that propagated radiometric uncertainty influences the timing of LSP metrics in the order of a few days. In extreme cases, uncertainties can take up to a few weeks. Uncertainties in LSP metrics should be taken into account when interpreting LSP data or when comparing remotely sensed phenology estimates to ground observations. However, more in-depth research on interscene error correlation is needed to further assess the magnitude of uncertainty in LSP metrics. In addition, sources of uncertainty not addressed in this study—e.g., from the AC—should be quantified.

Finally, we would like to emphasize that the methods and findings of our research are fully reproducible as code and L2B data products are freely available. Thus, future research can follow up our approaches and elaborate on open questions—also with regard to other disciplines and geographical regions.

APPENDIX

TABLE IV

MEAN PERCENTAGE OF PIXELS PER SELECTED SCL CLASS ± STANDARD DEVIATION DERIVED FROM 300 MC REALIZATIONS FOR ALL 33 S2 SCENES

sensing date	no data	saturated or defective	dark area pixels	cloud shadows	vegetation	non vegetated	water	unclassified	cloud medium probability	cloud high probability	thin cirrus	snow
2019-02-14	(0.0±1.5)e+2	0.0±nan	13.0±4.0	6.0±4.0	52.1±0.4	6.8±1.5	0.4±1.5	15.5±2.0	1.0±4.0	0.0±5.0	0.0±nan	5.9±0.4
2019-02-16	(0±4)e+1	0.0±nan	9±5	3±8	56.6±0.4	12.2±1.4	0.3±1.8	13.6±1.7	1.0±4.0	0.0±5	0.0±nan	3.86±0.34
2019-02-21	0.0±nan	0.0±nan	9±5	2±12	52.9±0.5	21.0±1.1	0.2±1.7	13.4±1.8	0.5±3.3	0.0±6	0.0±nan	1.0±1.0
2019-02-24	0.0±nan	0.0±nan	9.2±2.3	0.0±20	60.1±0.4	22.4±0.9	0.2±1.5	7.1±1.9	0.4±3.2	0.0±4.0	0.0±nan	0.2±1.9
2019-02-26	(0±6)e+1	0.0±nan	7.0±2.6	1.0±12	60.5±0.4	23.6±0.8	0.2±2.0	7.6±1.7	0.4±3.0	0.0±6	0.0±nan	0.1±2.7
2019-03-16	0.0±nan	0.0±nan	5.3±3.0	0.9±1.4	71.8±0.2	14.6±1.1	0.1±1.0	6.4±2.2	0.6±2.4	0.2±2.1	0.0±nan	0.0±nan
2019-03-21	0.0±nan	0.0±nan	3.0±2.8	0.0±27	72.79±0.22	19.2±0.8	0.1±0.7	4.4±1.6	0.4±1.7	0.0±6	0.0±nan	0.0±nan
2019-03-23	0.0±nan	0.0±nan	2.1±2.5	0.0±29	72.4±0.2	20.6±0.8	0.1±1.7	4.2±1.4	0.4±2.1	0.0±5	0.0±nan	0.0±nan
2019-03-28	0.0±nan	0.0±nan	1.6±3.0	0.6±2.3	73.79±0.20	18.9±0.8	0.1±1.4	3.9±1.5	0.6±2.0	0.2±3.2	0.2±3.5	0.0±nan
2019-03-31	0.0±nan	0.0±nan	2.1±3.0	0.0±24.0	74.6±0.2	18.6±0.8	0.1±1.0	4.2±1.6	0.4±2.2	0.0±5.0	0.0±nan	0.0±nan
2019-04-20	0.0±nan	0.0±nan	0.0±5.0	0.0±25.0	83.3±0.1	13.0±0.6	0.1±1.0	2.7±1.6	0.4±3.3	0.0±4.0	0.0±nan	0.0±nan
2019-05-30	0.0±25	0.0±nan	1.0±5	3.5±0.9	65.1±0.2	7.7±1.0	0.1±1.1	5.7±1.2	4.3±1.4	12.4±0.4	0.2±1.8	0.0±1.2
2019-06-01	0.0±nan	0.0±nan	0.1±3.1	0.0±19	84.7±0.1	12.8±0.5	0.1±1.0	1.6±1.8	0.0±4.0	0.0±4.0	0.0±nan	0.0±nan
2019-06-04	0.0±nan	0.0±nan	0.2±1.4	0.0±nan	85.7±0.09	11.1±0.7	0.0±nan	2.6±1.6	0.0±6	0.0±4.0	0.0±26.0	0.0±nan
2019-06-19	0.0±nan	0.0±nan	0.1±2.1	(0.0±1.6)e+2	86.9±0.1	9.6±0.7	0.1±2.4	2.1±1.9	1.0±4.0	0.2±3.0	0.0±13	(0±4)e+1
2019-06-24	0.0±nan	0.0±nan	0.2±2.0	(0.0±1.7)e+3	81.5±0.6	6.7±1.3	0.0±6	1.6±3.4	1.0±4.0	0.2±2.9	9±6	0.0±nan
2019-06-26	0.0±nan	0.0±nan	0.2±0.9	0.0±15.0	89.6±0.1	7.8±1.0	(0.0±2.0)e+2	1.6±1.9	1.0±4.0	0.2±3.2	0.0±nan	0.0±nan
2019-06-29	(0.0±1.9)e+2	0.0±nan	0.2±3.1	(0±6)e+1	89.3±0.1	8.4±0.8	0.0±5	1.5±1.9	0.4±2.5	0.2±3.1	(0.5±0.0)e+2	(0.0±4.0)e+1
2019-07-04	0.0±nan	0.0±nan	0.2±2.9	(0±6)e+2	90.3±0.1	7.2±0.9	0.0±14.0	1.8±1.6	0.4±3.5	0.0±4.0	0.0±nan	0.0±nan
2019-07-16	0.0±nan	0.0±nan	0.0±5	2.0±0.8	76.8±0.1	11.7±0.7	0.1±0.7	3.0±1.6	2.3±2.0	3.8±1.2	0.0±nan	(0.0±1.2)e+3
2019-07-24	(0.0±1.5)e+2	0.0±nan	0.0±6	(0.0±2.4)e+2	82.1±0.1	15.4±0.5	0.1±1.0	1.8±1.6	0.0±4.0	0.0±4.0	0.0±nan	0.0±nan
2019-07-26	0.0±nan	0.0±nan	0.1±2.9	0.0±nan	81.3±0.1	16.1±0.5	0.0±4.0	1.8±1.8	0.0±4.0	0.2±3.5	0.0±nan	0.0±nan
2019-08-08	(0±6)e+2	0.0±nan	1.0±4.0	5.0±0.6	65.7±0.2	10.9±0.7	0.1±0.9	5.1±2.2	4.2±0.9	7.9±0.7	0.0±17.0	0.0±6
2019-08-18	(0.0±1.6)e+2	0.0±nan	0.0±5	0.0±nan	83.8±0.1	13.4±0.5	0.1±0.6	2.0±1.6	0.3±3.0	0.2±3.5	0.0±nan	(0±4)e+2
2019-08-25	0.0±nan	0.0±nan	0.0±5.0	(0±5)e+1	85.0±0.1	11.9±0.5	0.1±1.5	2.2±1.5	0.0±4.0	0.0±4.0	0.0±nan	0.0±nan
2019-08-28	0.0±nan	0.0±nan	0.0±6	(0±4)e+1	84.77±0.08	11.7±0.6	0.0±4.0	2.8±1.2	0.0±4.0	0.0±4.0	0.0±nan	0.0±nan
2019-08-30	0.0±nan	0.0±nan	0.0±5	(0±8)e+2	85.1±0.1	11.8±0.5	0.1±1.3	2.4±1.4	0.3±3.5	0.0±4.0	0.0±nan	0.0±nan
2019-09-04	0.0±nan	0.0±nan	0.0±6	0.0±12	85.7±0.1	11.0±0.6	0.1±1.2	2.5±1.3	0.4±3.5	0.0±5.0	0.0±nan	(0.7±0.0)e+3
2019-09-12	(0.0±2.0)e+2	0.0±nan	2.0±2.6	0.0±11	86.1±0.11	8.4±1.0	0.1±1.4	2.8±1.9	0.0±4.0	0.0±4.0	0.0±nan	0.0±nan
2019-09-17	0.0±nan	0.0±nan	0.0±7	0.0±13	78.2±1.2	11.0±5.0	0.0±nan	10±4.0	1.0±5.0	0.0±8	0.0±nan	0.0±nan
2019-09-19	(0.0±1.1)e+2	0.0±nan	0.0±5	(0.0±1.1)e+2	85.0±0.1	11.1±0.6	0.1±1.5	2.9±1.1	0.4±3.0	0.0±5	0.0±nan	0.0±nan
2019-09-29	0.0±nan	0.0±nan	1.0±5	0.0±15	86.7±0.1	8.8±0.7	0.1±0.7	3.2±1.1	0.3±3.5	0.0±4.0	0.0±nan	0.0±nan
2019-10-12	0.0±nan	0.0±nan	5.1±3.5	0.0±5	83.44±0.09	4.7±1.2	0.0±7	5.7±2.3	0.0±5	0.0±4.0	1.0±7	0.0±nan
2019-10-14	(0±6)e+1	0.0±nan	1.0±4.0	0.0±15	84.1±0.1	9.2±0.7	0.1±1.0	4.7±1.1	0.0±4.0	0.0±4.0	0.0±nan	0.0±nan

The overall number of pixels per scene is 360 000 (20 × 20 m spatial resolution).

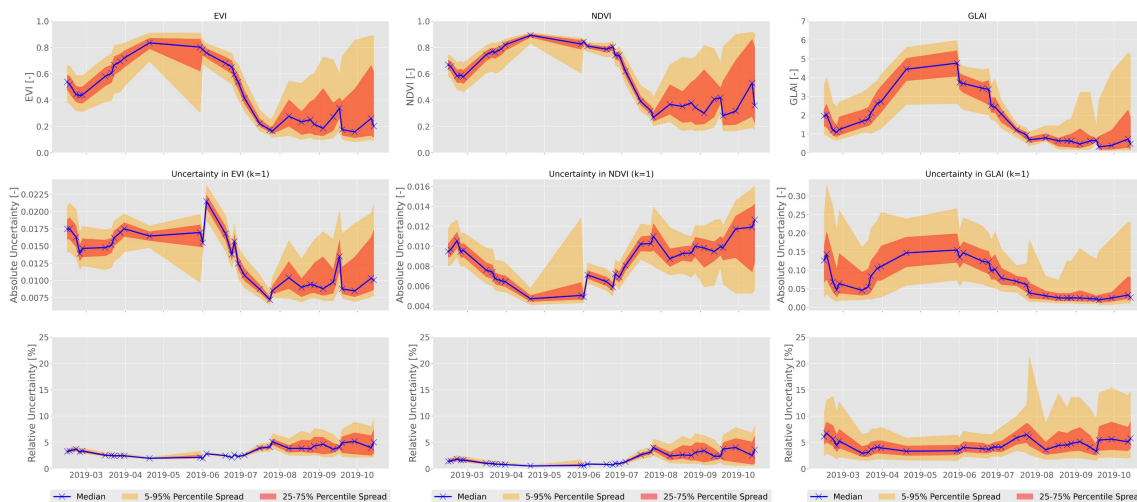


Fig. 12. Spatiotemporal variability of EVI (left column), NDVI (mid column), and GLAI (right column) for all pixels annotated as rapeseed (70.4 ha). For each S2 scene, the median value (blue line), central 50% (red), and 90% spread (orange) across all pixels is shown not classified as cloud, shadow, or snow based on the SCL data of the original S2 outputs after Sen2Cor. The top row shows the spatial variability in the actual EVI, NDVI, and GLAI derived from the S2 scenes. In the mid row, the absolute uncertainties derived from the MC simulations are depicted, whereas the bottom row shows the uncertainties in relative terms.

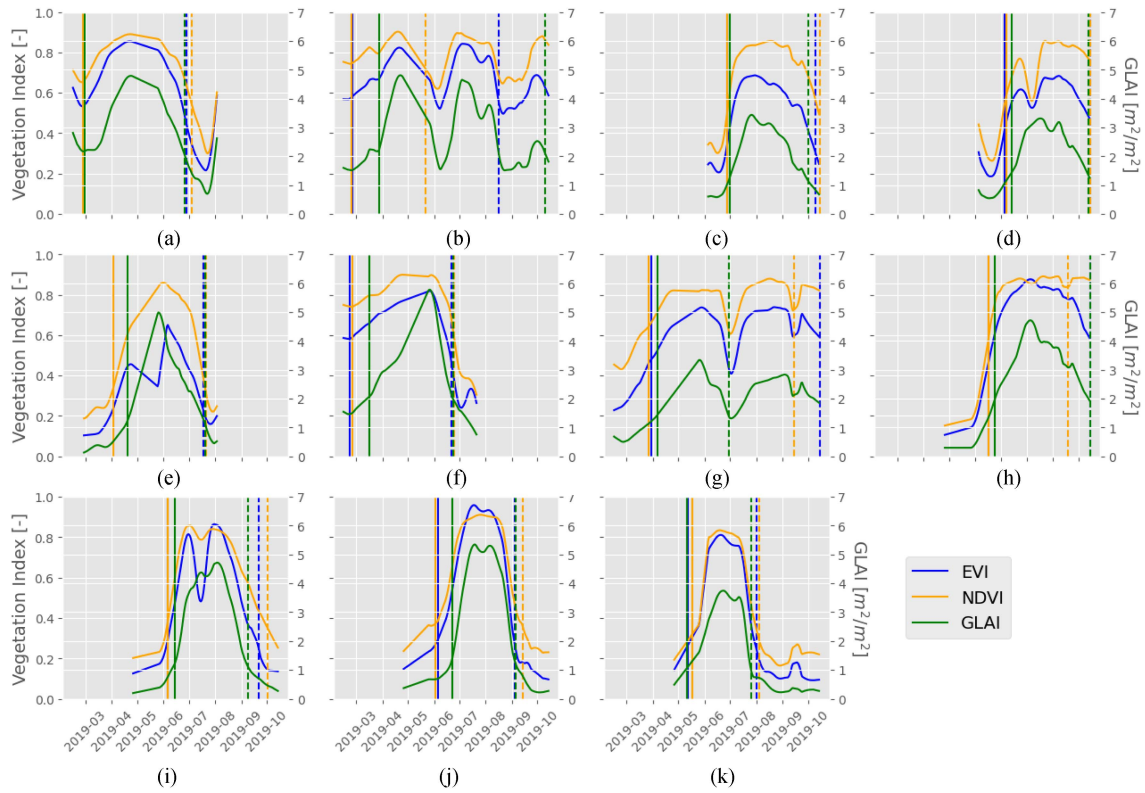


Fig. 13. Sample time series of different crop types from randomly selected pixels showing EVI (blue), NDVI (orange), and GLAI (green). The curves have been outlier-filtered and interpolated to daily values using Savitzky–Golay. (a) Rapeseed. (b) Permanent grassland. (c) Grain maize. (d) Silage maize. (e) Winter wheat. (f) Winter barley. (g) Extensively used grassland. (h) Sugar beet. (i) Sunflower. (j) Soybean. (k) Potato.

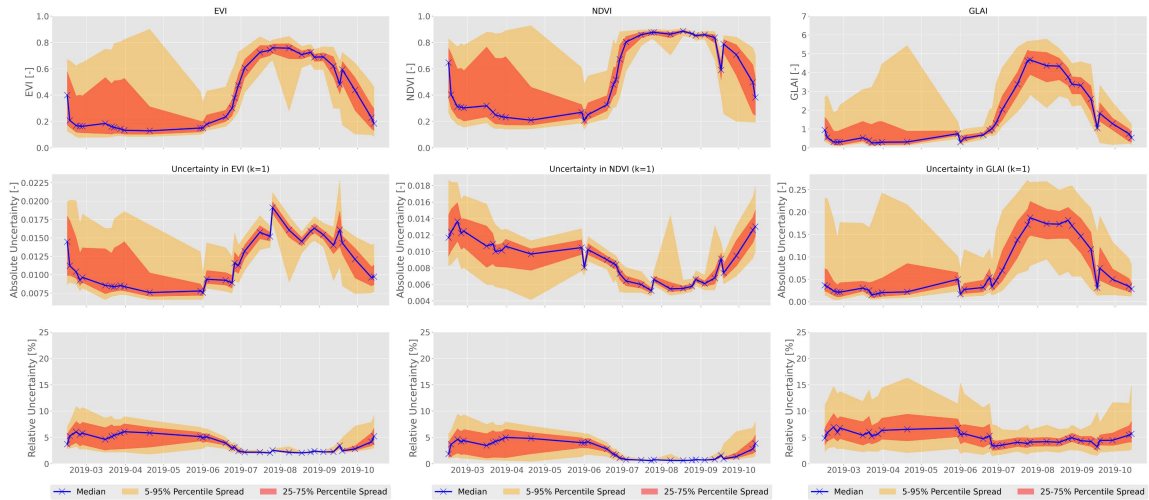


Fig. 14. Spatiotemporal variability of EVI (left column), NDVI (mid column), and GLAI (right column) for all pixels annotated as grain maize (41.3 ha). For each S2 scene, the median value (blue line), central 50% (red), and 90% spread (orange) across all pixels is shown not classified as cloud, shadow, or snow based on the SCL data of the original S2 outputs after Sen2Cor. The top row shows the spatial variability in the actual EVI, NDVI, and GLAI derived from the S2 scenes. In the mid row, the absolute uncertainties derived from the MC simulations are depicted, whereas the bottom row shows the uncertainties in relative terms.

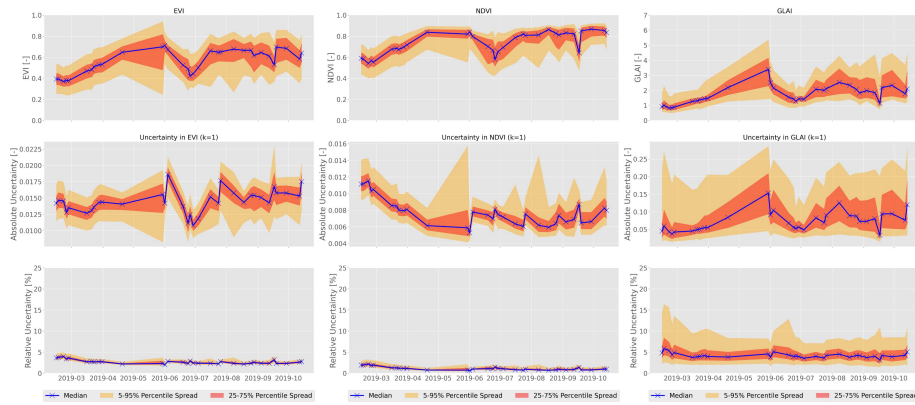


Fig. 15. Spatiotemporal variability of EVI (left column), NDVI (mid column), and GLAI (right column) for all pixels annotated as extensively used grassland (62.9 ha). For each S2 scene, the median value (blue line), central 50% (red), and 90% spread (orange) across all pixels is shown not classified as cloud, shadow, or snow based on the SCL data of the original S2 outputs after Sen2Cor. The top row shows the spatial variability in the actual EVI, NDVI, and GLAI derived from the S2 scenes. In the mid row, the absolute uncertainties derived from the MC simulations are depicted, whereas the bottom row shows the uncertainties in relative terms.

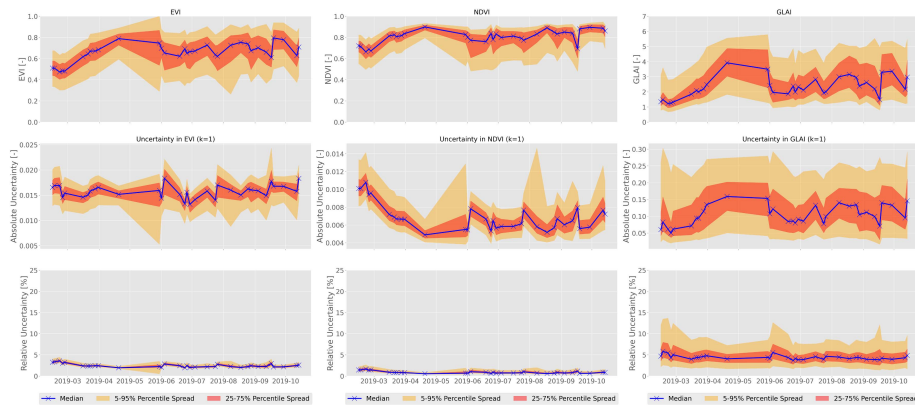


Fig. 16. Spatiotemporal variability of EVI (left column), NDVI (mid column), and GLAI (right column) for all pixels annotated as permanent grassland (168.9 ha). For each S2 scene, the median value (blue line), central 50% (red), and 90% spread (orange) across all pixels is shown not classified as cloud, shadow, or snow based on the SCL data of the original S2 outputs after Sen2Cor. The top row shows the spatial variability in the actual EVI, NDVI, and GLAI derived from the S2 scenes. In the mid row, the absolute uncertainties derived from the MC simulations are depicted, whereas the bottom row shows the uncertainties in relative terms.

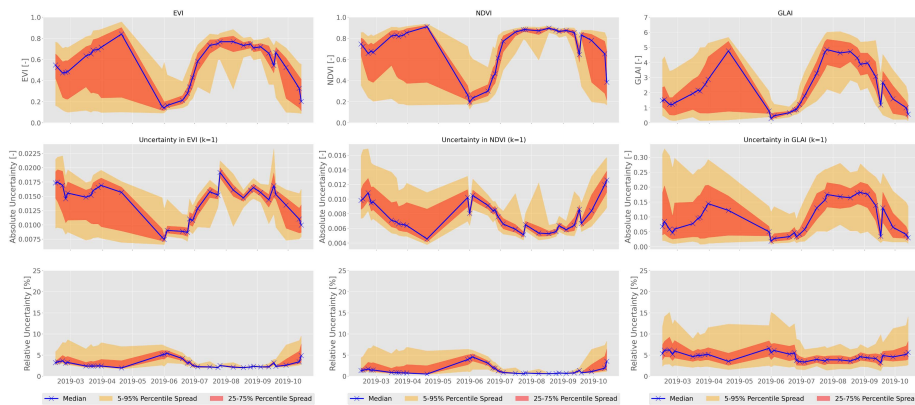


Fig. 17. Spatiotemporal variability of EVI (left column), NDVI (mid column), and GLAI (right column) for all pixels annotated as silage maize (175.6 ha). For each S2 scene, the median value (blue line), central 50% (red), and 90% spread (orange) across all pixels is shown not classified as cloud, shadow, or snow based on the SCL data of the original S2 outputs after Sen2Cor. The top row shows the spatial variability in the actual EVI, NDVI, and GLAI derived from the S2 scenes. In the mid row, the absolute uncertainties derived from the MC simulations are depicted, whereas the bottom row shows the uncertainties in relative terms.

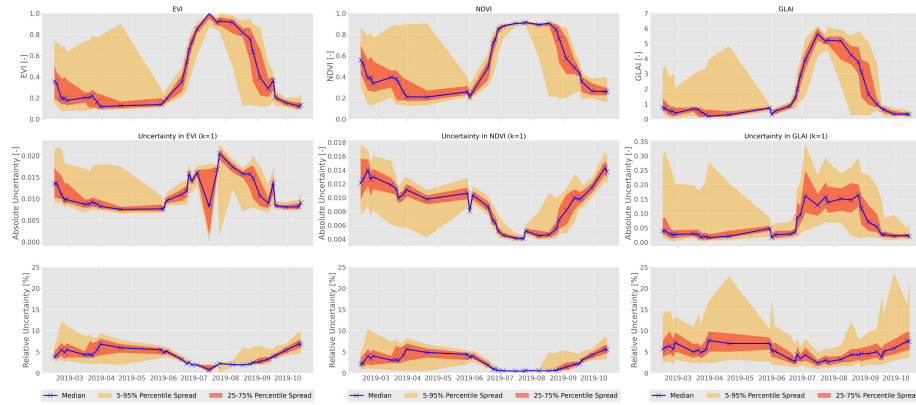


Fig. 18. Spatiotemporal variability of EVI (left column), NDVI (mid column), and GLAI (right column) for all pixels annotated as soybean (7.8 ha). For each S2 scene, the median value (blue line), central 50% (red), and 90% spread (orange) across all pixels is shown not classified as cloud, shadow, or snow based on the SCL data of the original S2 outputs after Sen2Cor. The top row shows the spatial variability in the actual EVI, NDVI, and GLAI derived from the S2 scenes. In the mid row, the absolute uncertainties derived from the MC simulations are depicted, whereas the bottom row shows the uncertainties in relative terms.

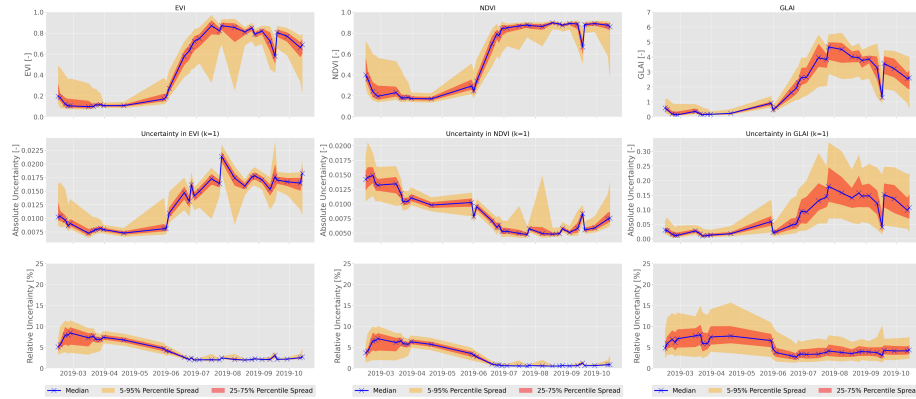


Fig. 19. Spatiotemporal variability of EVI (left column), NDVI (mid column), and GLAI (right column) for all pixels annotated as sugar beet (65.4 ha). For each S2 scene, the median value (blue line), central 50% (red), and 90% spread (orange) across all pixels is shown not classified as cloud, shadow, or snow based on the SCL data of the original S2 outputs after Sen2Cor. The top row shows the spatial variability in the actual EVI, NDVI, and GLAI derived from the S2 scenes. In the mid row, the absolute uncertainties derived from the MC simulations are depicted, whereas the bottom row shows the uncertainties in relative terms.

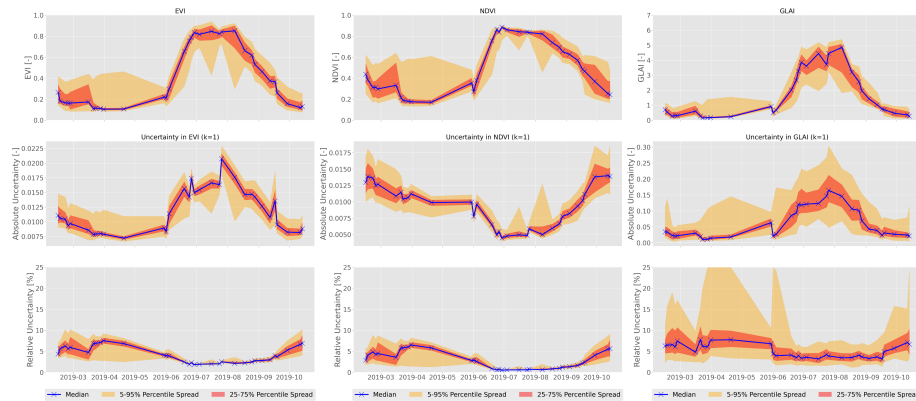


Fig. 20. Spatiotemporal variability of EVI (left column), NDVI (mid column), and GLAI (right column) for all pixels annotated as sunflower (7.9 ha). For each S2 scene, the median value (blue line), central 50% (red), and 90% spread (orange) across all pixels is shown not classified as cloud, shadow, or snow based on the SCL data of the original S2 outputs after Sen2Cor. The top row shows the spatial variability in the actual EVI, NDVI, and GLAI derived from the S2 scenes. In the mid row, the absolute uncertainties derived from the MC simulations are depicted, whereas the bottom row shows the uncertainties in relative terms.

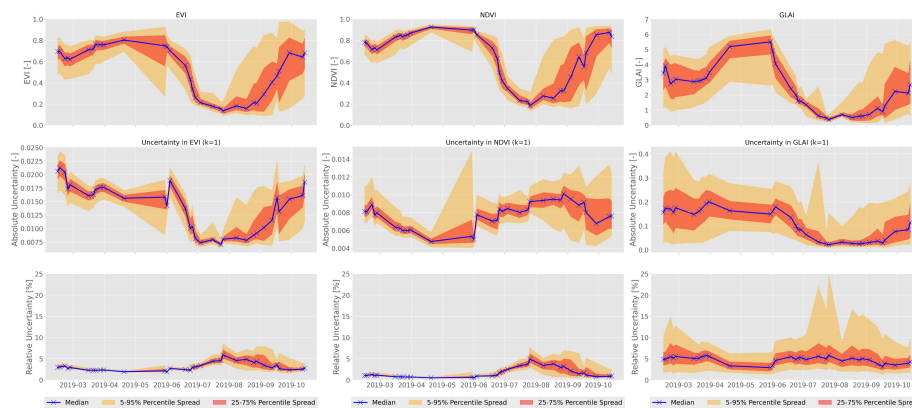


Fig. 21. Spatiotemporal variability of EVI (left column), NDVI (mid column) and GLAI (right column) for all pixels annotated as winter barley (63.4 ha). For each S2 scene, the median value (blue line), central 50% (red), and 90% spread (orange) across all pixels is shown not classified as cloud, shadow, or snow based on the SCL data of the original S2 outputs after Sen2Cor. The top row shows the spatial variability in the actual EVI, NDVI, and GLAI derived from the S2 scenes. In the mid row, the absolute uncertainties derived from the MC simulations are depicted, whereas the bottom row shows the uncertainties in relative terms.

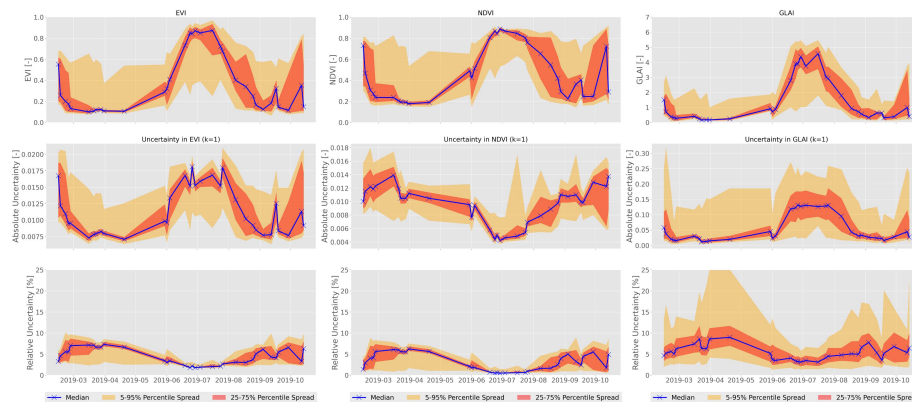


Fig. 22. Spatiotemporal variability of EVI (left column), NDVI (mid column), and GLAI (right column) for all pixels annotated as potato (4.7 ha). For each S2 scene, the median value (blue line), central 50% (red), and 90% spread (orange) across all pixels is shown not classified as cloud, shadow, or snow based on the SCL data of the original S2 outputs after Sen2Cor. The top row shows the spatial variability in the actual EVI, NDVI, and GLAI derived from the S2 scenes. In the mid row, the absolute uncertainties derived from the MC simulations are depicted, whereas the bottom row shows the uncertainties in relative terms.

Code Availability: Code used for the data processing and analysis is available online.⁷

Data Availability: We provide standard uncertainties for NDVI, EVI, and GLAI (L2B products) for an entire growing season as an output of the radiometric uncertainty propagation chain. The data are available online.⁸

REFERENCES

- [1] A. Gonsamo and J. M. Chen, "Circumpolar vegetation dynamics product for global change study," *Remote Sens. Environ.*, vol. 182, pp. 13–26, Sep. 2016.
- [2] J. A. Caparros-Santiago, V. Rodriguez-Galiano, and J. Dash, "Land surface phenology as indicator of global terrestrial ecosystem dynamics: A systematic review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 171, pp. 330–347, Jan. 2021.
- [3] W. Wu, Y. Sun, K. Xiao, and Q. Xin, "Development of a global annual land surface phenology dataset for 1982–2018 from the AVHRR data by implementing multiple phenology retrieving methods," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 103, Dec. 2021, Art. no. 102487.
- [4] R. B. Myneni, F. G. Hall, P. J. Sellers, and A. L. Marshak, "The interpretation of spectral vegetation indexes," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 2, pp. 481–486, Mar. 1995.
- [5] W. Verhoef and H. Bach, "Remote sensing data assimilation using coupled radiative transfer models," *Phys. Chem. Earth, Parts A/B/C*, vol. 28, nos. 1–3, pp. 3–13, Jan. 2003.
- [6] L. Homolová, Z. Malenovský, J. G. P. W. Clevers, G. García-Santos, and M. E. Schaepman, "Review of optical-based remote sensing for plant trait mapping," *Ecol. Complexity*, vol. 15, pp. 1–16, Sep. 2013.
- [7] E. K. Melaas, M. A. Friedl, and Z. Zhu, "Detecting interannual variation in deciduous broadleaf forest phenology using Landsat TM/ETM data," *Remote Sens. Environ.*, vol. 132, pp. 176–185, May 2013.
- [8] D. K. Bolton, J. M. Gray, E. K. Melaas, M. Moon, L. Eklundh, and M. A. Friedl, "Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery," *Remote Sens. Environ.*, vol. 240, Apr. 2020, Art. no. 111685.
- [9] H. Croft et al., "The global distribution of leaf chlorophyll content," *Remote Sens. Environ.*, vol. 236, Jan. 2020, Art. no. 111479.
- [10] F. Tian et al., "Calibrating vegetation phenology from Sentinel-2 using eddy covariance, PhenoCam, and PEP725 networks across Europe," *Remote Sens. Environ.*, vol. 260, Jul. 2021, Art. no. 112456.
- [11] K. M. de Beurs and G. M. Henebry, "Land surface phenology, climatic variation, and institutional change: Analyzing agricultural land cover change in Kazakhstan," *Remote Sens. Environ.*, vol. 89, no. 4, pp. 497–509, Feb. 2004.
- [12] L. Zeng, B. D. Wardlow, D. Xiang, S. Hu, and D. Li, "A review of vegetation phenological metrics extraction using time-series, multispectral satellite data," *Remote Sens. Environ.*, vol. 237, Feb. 2020, Art. no. 111511.

⁷[Online]. Available: <https://doi.org/10.5281/zenodo.6669854>

⁸[Online]. Available: <https://doi.org/10.3929/ethz-b-000574818>

- [13] D. Lloyd, "A phenological classification of terrestrial vegetation cover using shortwave vegetation index imagery," *Int. J. Remote Sens.*, vol. 11, no. 12, pp. 2269–2279, Dec. 1990.
- [14] B. C. Reed, J. F. Brown, D. VanderZee, T. R. Loveland, J. W. Merchant, and D. O. Ohlen, "Measuring phenological variability from satellite imagery," *J. Vegetation Sci.*, vol. 5, no. 5, pp. 703–714, Oct. 1994.
- [15] T. Sakamoto, M. Yokozawa, H. Toritani, M. Shibayama, N. Ishitsuka, and H. Ohno, "A crop phenology detection method using time-series MODIS data," *Remote Sens. Environ.*, vol. 96, nos. 3/4, pp. 366–374, Jun. 2005.
- [16] M. Meroni, M. M. Verstraete, F. Rembold, F. Urbano, and F. Kayitakire, "A phenology-based method to derive biomass production anomalies for food security monitoring in the Horn of Africa," *Int. J. Remote Sens.*, vol. 35, no. 7, pp. 2472–2492, Apr. 2014.
- [17] F. Gao et al., "Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery," *Remote Sens. Environ.*, vol. 188, pp. 9–25, Jan. 2017.
- [18] C. Diao, "Remote sensing phenological monitoring framework to characterize corn and soybean physiological growing stages," *Remote Sens. Environ.*, vol. 248, Oct. 2020, Art. no. 111960.
- [19] M. Salinero-Delgado et al., "Monitoring cropland phenology on Google Earth engine using Gaussian process regression," *Remote Sens.*, vol. 14, no. 1, Jan. 2022, Art. no. 146.
- [20] K. White, J. Pontius, and P. Schaberg, "Remote sensing of spring phenology in northeastern forests: A comparison of methods, field metrics and sources of uncertainty," *Remote Sens. Environ.*, vol. 148, pp. 97–107, May 2014.
- [21] K. Kowalski, C. Senf, P. Hostert, and D. Pflugmacher, "Characterizing spring phenology of temperate broadleaf forests using Landsat and Sentinel-2 time series," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 92, Oct. 2020, Art. no. 102172.
- [22] R. Shen, X. Chen, L. Chen, B. He, and W. Yuan, "Regional evaluation of satellite-based methods for identifying leaf unfolding date," *ISPRS J. Photogrammetry Remote Sens.*, vol. 175, pp. 88–98, May 2021.
- [23] H. Wagenseil and C. Samimi, "Assessing spatio-temporal variations in plant phenology using Fourier analysis on NDVI time series: Results from a dry Savannah environment in Namibia," *Int. J. Remote Sens.*, vol. 27, no. 16, pp. 3455–3471, Aug. 2006.
- [24] L. Gu et al., "Characterizing the seasonal dynamics of plant community photosynthesis across a range of vegetation types," in *Phenology of Ecosystem Processes: Applications in Global Change Research*, A. Noormets, Ed. New York, NY, USA: Springer, 2009, pp. 35–58.
- [25] A. Vrieling et al., "Vegetation phenology from Sentinel-2 and field cameras for a Dutch barrier island," *Remote Sens. Environ.*, vol. 215, pp. 517–529, Sep. 2018.
- [26] M. Moon, A. D. Richardson, and M. A. Friedl, "Multiscale assessment of land surface phenology from harmonized Landsat 8 and Sentinel-2, PlanetScope, and PhenoCam imagery," *Remote Sens. Environ.*, vol. 266, Dec. 2021, Art. no. 112716.
- [27] D. Helman, "Land surface phenology: What do we really 'see' from space," *Sci. Total Environ.*, vol. 618, pp. 665–673, Mar. 2018.
- [28] E. Amin, J. Verrelst, J. P. Rivera-Caicedo, L. Pipia, A. Ruiz-Verdú, and J. Moreno, "Prototyping Sentinel-2 green LAI and brown LAI products for cropland monitoring," *Remote Sens. Environ.*, vol. 255, Mar. 2021, Art. no. 112168.
- [29] R. Pazúr, N. Huber, D. Weber, C. Ginzler, and B. Price, "A national extent map of cropland and grassland for Switzerland based on Sentinel-2 data," *Earth System Sci. Data*, vol. 14, no. 1, pp. 295–305, Jan. 2022.
- [30] W. J. Frampton, J. Dash, G. Watmough, and E. J. Milton, "Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 82, pp. 83–92, Aug. 2013.
- [31] G. Misra, F. Cawkwell, and A. Wingler, "Status of phenological research using Sentinel-2 data: A review," *Remote Sens.*, vol. 12, no. 17, Aug. 2020, Art. no. 2760.
- [32] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, "Monitoring vegetation systems in the Great Plains with ERTS," *NASA Special Publ.*, vol. 351, no. 1974, Art. no. 309.
- [33] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, "Overview of the radiometric and biophysical performance of the MODIS vegetation indices," *Remote Sens. Environ.*, vol. 83, no. 1, pp. 195–213, Nov. 2002.
- [34] M. A. White and R. R. Nemani, "Real-time monitoring and short-term forecasting of land surface phenology," *Remote Sens. Environ.*, vol. 104, no. 1, pp. 43–49, Sep. 2006.
- [35] S. E. Hunt et al., "A quality assurance framework for satellite Earth observation missions," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 608–611.
- [36] J. Mittaz, C. J. Merchant, and E. R. Woolliams, "Applying principles of metrology to historical Earth observations from satellites," *Metrologia*, vol. 56, no. 3, May 2019, Art. no. 032002.
- [37] J. Gorroño et al., "A radiometric uncertainty tool for the sentinel 2 mission," *Remote Sens.*, vol. 9, no. 2, Feb. 2017, Art. no. 178.
- [38] J. Gorroño et al., "Providing uncertainty estimates of the Sentinel-2 top-of-atmosphere measurements for radiometric validation activities," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 650–666, Jan. 2018.
- [39] C. J. Merchant et al., "Sea surface temperature datasets for climate applications from phase 1 of the European space agency climate change initiative (SST CCI)," *Geosci. Data J.*, vol. 1, no. 2, pp. 179–191, 2014.
- [40] F. Mílin, G. Sclep, T. Jackson, and S. Sathyendranath, "Uncertainty estimates of remote sensing reflectance derived from comparison of ocean color satellite data sets," *Remote Sens. Environ.*, vol. 177, pp. 107–124, May 2016.
- [41] L. I. W. McKinna, I. Cetinić, A. P. Chase, and P. J. Werdell, "Approach for propagating radiometric data uncertainties through NASA ocean color algorithms," *Front. Earth Sci.*, vol. 7, Jul. 2019, Art. no. 176.
- [42] *Supplement 1 to the 'Guide to the Expression of Uncertainty in Measurement'—Propagation of Distributions Using a Monte Carlo Method*, JCGM 101:2008, Int. Org. Standardization, Geneva, Switzerland, 2008.
- [43] J. Gorroño, A. C. Banks, N. P. Fox, and C. Underwood, "Radiometric inter-sensor cross-calibration uncertainty using a traceable high accuracy reference hyperspectral imager," *ISPRS J. Photogrammetry Remote Sens.*, vol. 130, pp. 393–417, Aug. 2017.
- [44] C. L. Wyatt, *Radiometric Calibration: Theory and Methods*. New York, NY, USA: Academic, 1978.
- [45] L. Ma et al., "Uncertainty analysis for RadCalNet instrumented test sites using the Baotou sites BTCN and BSCN as examples," *Remote Sens.*, vol. 12, no. 11, Jan. 2020, Art. no. 1696.
- [46] L. V. Graf, G. Perich, and H. Aasen, "EOdal: An open-source Python package for large-scale agroecological research using Earth Observation and gridded environmental data," *Comput. Electron. Agriculture*, vol. 203, Dec. 2022, Art. no. 107487.
- [47] U. Muller-Wilm, J. Louis, R. Richter, F. Gascon, and M. Niezette, "Sentinel-2 level 2A prototype processor: Architecture, algorithms and first results," in *Proc. ESA Living Planet Symp.*, 2013, pp. 9–13.
- [48] M. Main-Knorn, B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm, and F. Gascon, "Sen2Cor for Sentinel-2," in *Proc. SPIE*, vol. 10427, Oct. 2017, Art. no. 1042704.
- [49] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, Dec. 2017.
- [50] I. Kamenova and P. Dimitrov, "Evaluation of Sentinel-2 vegetation indices for prediction of LAI, fAPAR and fCover of winter wheat in Bulgaria," *Eur. J. Remote Sens.*, vol. 54, pp. 89–108, Feb. 2021.
- [51] S. Jacquemoud et al., "PROSPECT SAIL models: A review of use for vegetation characterization," *Remote Sens. Environ.*, vol. 113, pp. S56–S66, Sep. 2009.
- [52] J.-B. Féret, A. Gitelson, S. Noble, and S. Jacquemoud, "PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle," *Remote Sens. Environ.*, vol. 193, pp. 204–215, May 2017.
- [53] M. Danner, K. Berger, M. Wocher, W. Mauser, and T. Hank, "Efficient RTM-based training of machine learning regression algorithms to quantify biophysical & biochemical traits of agricultural crops," *ISPRS J. Photogrammetry Remote Sens.*, vol. 173, pp. 278–296, Mar. 2021.
- [54] L. Eklundh and P. Jönsson, "TIMESAT: A software package for time-series processing and assessment of vegetation dynamics," in *Remote Sensing Time Series: Revealing Land Surface Dynamics*, ser. Remote Sensing and Digital Image Processing, C. Kuenzer, S. Dech, and W. Wagner, Eds. Cham, Switzerland: Springer, 2015, pp. 141–158.
- [55] L. Trotter and T. Robinson, "Frontiers/phenology v1.0," 2021. [Online]. Available: <https://zenodo.org/record/4738425>
- [56] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964.
- [57] J. Chen, P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh, "A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter," *Remote Sens. Environ.*, vol. 91, nos. 3/4, pp. 332–344, Jun. 2004.

- [58] J. Zhou, L. Jia, M. Menenti, and B. Gorte, "On the performance of remote sensing time series reconstruction methods—A spatial comparison," *Remote Sens. Environ.*, vol. 187, pp. 367–384, Dec. 2016.
- [59] A. de Castro, J. Six, R. Plant, and J. Peña, "Mapping crop calendar events and phenology-related metrics at the parcel level by object-based image analysis (OBIA) of MODIS-NDVI time-series: A case study in central California," *Remote Sens.*, vol. 10, no. 11, Nov. 2018, Art. no. 1745.
- [60] P. Jönsson and L. Eklundh, "Seasonality extraction from time-series of satellite sensor data," in *Frontiers of Remote Sensing Information Processing*. Singapore: World Scientific, Jul. 2003, pp. 487–500.
- [61] J. Louis et al., "Sentinel-2 global surface reflectance Level-2a product generated with Sen2Cor," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 8522–8525.
- [62] C.-C. Liu et al., "Clouds classification from Sentinel-2 imagery with deep residual learning and semantic image segmentation," *Remote Sens.*, vol. 11, no. 2, Jan. 2019, Art. no. 119.
- [63] K. Prabhakara, W. D. Hively, and G. W. McCarty, "Evaluating the relationship between biomass, percent groundcover and remote sensing indices across six winter cover crop fields in Maryland, USA," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 39, pp. 88–102, Jul. 2015.
- [64] R. Zurita-Milla, V. Laurent, and J. van Gijsel, "Visualizing the ill-posedness of the inversion of a canopy radiative transfer model: A case study for Sentinel-2," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 43, pp. 7–18, Dec. 2015.
- [65] I. Garonna, R. de Jong, and M. E. Schaepman, "Variability and evolution of global land surface phenology over the past three decades (1982–2012)," *Glob. Change Biol.*, vol. 22, no. 4, pp. 1456–1468, 2016.
- [66] T. C. Nietupski, R. E. Kennedy, H. Temesgen, and B. K. Kerns, "Spatiotemporal image fusion in Google Earth Engine for annual estimates of land surface phenology in a heterogenous landscape," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 99, Jul. 2021, Art. no. 102323.
- [67] Z. Luo and S. Yu, "Spatiotemporal variability of land surface phenology in China from 2001–2014," *Remote Sens.*, vol. 9, no. 1, 2017, Art. no. 65.
- [68] T. Park et al., "Changes in growing season duration and productivity of northern vegetation inferred from long-term remote sensing data," *Environ. Res. Lett.*, vol. 11, no. 8, Aug. 2016, Art. no. 084001.
- [69] N. Younes, K. E. Joyce, and S. W. Maier, "All models of satellite-derived phenology are wrong, but some are useful: A case study from northern Australia," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 97, May 2021, Art. no. 102285.
- [70] L. G. J. van Bussel, F. Ewert, and P. A. Leffelaar, "Effects of data aggregation on simulations of crop phenology," *Agriculture, Ecosyst. Environ.*, vol. 142, no. 1, pp. 75–84, Jul. 2011.
- [71] X. Chen, D. Wang, J. Chen, C. Wang, and M. Shen, "The mixed pixel effect in land surface phenology: A simulation study," *Remote Sens. Environ.*, vol. 211, pp. 338–344, Jun. 2018.
- [72] B. Lara and M. Gandini, "Assessing the performance of smoothing functions to estimate land surface phenology on temperate grassland," *Int. J. Remote Sens.*, vol. 37, no. 8, pp. 1801–1813, Apr. 2016, doi: [10.1080/2150704X.2016.1168945](https://doi.org/10.1080/2150704X.2016.1168945).
- [73] J. C. Zadoks, T. T. Chang, and C. F. Konzak, "A decimal code for the growth stages of cereals," *Weed Res.*, vol. 14, no. 6, pp. 415–421, 1974.
- [74] K. Harfenmeister, S. Itzerott, C. Weltzien, and D. Spengler, "Detecting phenological development of winter wheat and winter barley using time series of Sentinel-1 and Sentinel-2," *Remote Sens.*, vol. 13, no. 24, Dec. 2021, Art. no. 5036.

Lukas Valentin Graf received the B.Sc. degree in geography from the Ludwig Maximilian University of Munich, Munich, Germany, in 2018, and the M.Sc. degree in geographic information science from the University of Salzburg, Salzburg, Austria, in 2020. He is currently working toward the Ph.D. degree with the Crop Science Group, ETH Zürich, Zürich, Switzerland.

His research interests include retrieval of crop phenology using satellite remote sensing and crop modeling.

Javier Gorroño received the M.Sc. degree in telecommunications engineering from the Universitat Politècnica de València, València, Spain, in 2010, and the Ph.D. degree in electrical engineering from the University of Surrey, Guildford, U.K., in 2018.

He is working on several Earth observation satellite missions, such as Copernicus Sentinel 2 and TRUTHS mission. He is currently an ESA Living Planet Fellow with the Universitat Politècnica de València, working on the development of uncertainty estimates associated with the surface reflectance products of the Sentinel 2 mission. His research interests include the quality assessment of satellite data products and the detection/quantification of methane point-source emissions from space.

Andreas Hueni received the B.Sc. degree in computer science from the University of Applied Science Brugg-Windisch, Windisch, Switzerland, in 1997, the PG diploma in geographic information systems and the M.Phil.(Sc.) degree in earth science from Massey University, Palmerston North, New Zealand, in 2005 and 2006, respectively, and the Ph.D. degree in geography from the University of Zürich, Zürich, Switzerland, in 2011.

He has worked on the APEX project since 2007. He is currently an Academic Associate with the Remote Sensing Laboratories, University of Zürich, responsible for APEX sensor and data calibration and project manager of the Airborne Research Facility for the Earth System. His research interests include the calibration of spectrometers, uncertainty analysis, and the design of combined database and software systems such as the spectral database SPECCHIO.

Achim Walter received the Diploma in physics and biology and the Ph.D. degree in plant sciences from the University of Heidelberg, Heidelberg, Germany, in 1995 and 1997, respectively.

After a postdoctoral stay with the Biosphere 2 Center, Columbia University, New York, NY, USA, in 2001, he was a Group Leader with Forschungszentrum Jülich, Jülich, Germany, from 2003 to 2010. Since 2010, he has been a Professor with ETH Zürich, Zürich, Switzerland, where he was also the Director of Studies from 2011 to 2017, responsible for the further development of teaching in agricultural sciences. He is a Member of the Scientific Advisory Board of FiBL, Switzerland; Agroscope, Switzerland; IPK Gatersleben, Germany; and the Phenorob Cluster of Excellence, Germany. His research group develops phenotyping methods and focuses on the possibilities of digitalization and artificial intelligence in the field of sustainable agriculture.

Helge Aasen received the Ph.D. degree in geographic information science from the University of Cologne, Cologne, Germany, in 2016.

He is currently the Leader of the Earth Observation of Agroecosystems team, Agroscope, Zürich, Switzerland, where he focuses on the usage of satellite and field phenotyping remote sensing methods to investigate plant-environment interactions. His aim is to produce actionable information that can guide decision making for various stakeholders in agriculture and beyond. His main research interests include ecophysiological earth observation.