Masterarbeit

Zur Erlangung des akademischen Grades Master of Science

Upscaling of Field Measurements for the Validation of the Sentinel-2 Level-2A Product

eingereicht von:	Britta Themann
Gutachter/innen:	Prof. Dr. Patrick Hostert Dr. David Frantz Dr. Magdalena Main-Knorn (DLR)

Eingereicht am Geographischen Institut der Humboldt-Universität zu Berlin am: 30.04.2019

Contents

Journal style adopted from Remote Sensing of the Environment

List of Figures

List of Tables

Master Thesis

Upscaling of Field Measurements for the Validation of the Sentinel-2 Level-2A Product

Britta Themann [1]

[1] Geography Department, Humboldt University Berlin, Rudower Chaussee 16, 12489 Berlin

Received: 30 April.2019

Abstract: In times of rapid global environmental change, the free distribution of the Sentinel-2 surface reflectance product has greatly improved the availability of high-resolution satellite imagery capable of capturing such changes. Having previously undergone geometric, radiometric and atmospheric correction, it is essential to determine the data quality before their further use in scientific studies. Preferably, validation of the Sentinel-2 Level-2A surface reflectance (L2A SR) product is performed using higher resolution reference data. In this study, field measurements and HySpex high-resolution airborne imagery for an area near Lake Stechlin, Germany, serve as reference to assess the Sentinel-2 L2A SR data quality. After spectral resampling of both field measurements and airborne imagery to Sentinel-2 spectral resolution, the field measurements are upscaled to match the spatial resolution of the airborne imagery using simple averaging. To account for the geolocation error, this is performed repeatedly with different geographic offsets until correlation with the HySpex pixel values is maximized. The resulting field measurement pixels are used to determine the quality of the airborne imagery, which subsequently undergoes spatial resampling, i.e. upscaling, to Sentinel-2 spatial resolution. Following this, surface reflectance from pixel samples is used to validate the Sentinel-2 L2A SR product with the upscaled HySpex data serving as reference. Accuracy, Precision and Uncertainty (APU) estimates of surface reflectance are reported together with uncertainties arising at different steps in the analysis. Ultimately, this is an assessment of Sen2Cor, the atmospheric correction algorithm used to produce the Sentinel-2 L2A SR product. The upscaling of field measurements using simple averaging produced realistic estimates but could be further improved by applying a more suitable field sampling scheme that allows for sophisticated upscaling techniques. HySpex imagery shows a high correlation with the upscaled field measurements comparable to results in PFLUG (2019). Validation of the Sentinel-2 L2A SR product revealed an overestimation of surface reflectance and APU statistics mostly outside their specifications. In comparison to similar studies (DOXANI et al. 2018; ESA 2019; PFLUG 2019), the Sentinel-2 imagery over Lake Stechlin contains larger uncertainties, which may originate from differences between Sentinel-2 and HySpex imagery in georeferencing, solar and viewing angles, IFOVs and processors as well as the retrieved AOT value at 550 nm as processing input.

Keywords: atmospheric correction, earth observation, Level-2A, Sentinel-2, surface reflectance, Sen2Cor, upscaling, validation

1. Introduction

In the light of global environmental change, there is an ever-increasing demand for high-resolution satellite imagery for the monitoring of environmental changes. As part of the European Copernicus Program, the Sentinel-2 mission provides such imagery free of charge to any user world-wide (ESA 2019). Before the ready-to-use surface reflectance product can be obtained via download from the Copernicus Open Access Hub (https://scihub.copernicus.eu/dhus/#/home), the raw data from the sensor undergoes a refinement process, which includes radiometric, geometric and atmospheric correction. Uncertainties arising from this refinement process need to be quantified to ensure the reliability of any conclusions drawn from a remote sensing analysis based on those products. To this end, the performance of atmospheric correction process of the Sentinel-2 Level-2A surface reflectance (L2A SR) product is evaluated through a validation process comparing satellite imagery with reference data from the earth's surface.

The quality of a remote sensing product is determined by comparing the respective image with reference data of the same entity acquired at the ground. A popular way to obtain ground measurements for the validation of optical remote sensing imagery is through networks of automated measurement sites. The Radiometric Calibration Network, short RadCalNet, run by a working group of the Committee on Earth Observation Satellites (CEOS) is intended for radiometric calibration (CEOS 2019), but is also exploited for L2A SR product validation. However, its 4 test sites in US, France, China and Namibia do not provide any surface reflectance measurements over dark targets, such as water or forest (CZAPLA-MYERS et al. 2016; JING et al. 2019). For the systematic validation of the Sentinel-2 L2A SR product, the world-wide AErosol Robotic NETwork project, short AERONET (HOLBEN et al. 1998), provides aerosol and water vapor measurements as reference data from which, in turn, reference data for surface reflectance can be derived (GASCON et al. 2017). Although the derivation of surface reflectance from AERONET sites is a convenient way to generate an abundance of reference pixels, it relies on radiative transfer models, whose own uncertainty of up to 1% are then included into the reference data (KOTCHENOVA et al. 2006; KOTCHENOVA & VERMOTE 2007). The provision of additional reference data through field campaigns not relying on radiative transfer models is therefore desired. However, field campaigns for the collection of reference data can be costly and time-consuming endeavors limiting the number of measurements taken and the area covered. There is also a mismatch between the area represented by a field measurement (~ 10 cm) and the size of the respective pixel (~ 1 m) in the remotely sensed imagery. To make a valid comparison between a pixel from remote sensing imagery and the respective location on the earth's surface, the field measurement needs to accurately cover the pixel's footprint. Since instruments used in the field are usually not capable of covering an area this large, multiple measurements within the pixel footprint need to be aggregated to produce one value representing the pixel footprint. Knowledge of the exact position of each pixel footprint, i.e. the geolocation error, is required, since spatial aggregation produces different results, depending on how measurements are aggregated. This phenomenon was first described by GEHLKE & BIEHL (1934) and is termed the modifiable area unit problem, short MAUP. Therefore, when aggregating field data to match the spatial resolution of a remotely sensed product, the selected method also needs to account for a possible geolocation error in the data.

2. Objective

This study aims at validating the Sentinel-2 L2A SR product for an area near Lake Stechlin, Germany, by upscaling field measurements to high-resolution HySpex imagery, which then serves as reference data to the Sentinel-2 L2A SR product. The airborne HySpex imagery thereby mediates between spatial resolutions of field and Sentinel-2 data.

In the course of a field campaign near Lake Stechlin, surface reflectance measurements were collected over water and grass in synchrony with fly-overs by the airborne HySpex sensor and the spaceborne Sentinel-2 satellite. First, both field measurements and the HySpex imagery are spectrally resampled to match the Sentinel-2B spectral bands. Following a thorough literature review on upscaling methods, i.e. the change of a value's supported area, the field measurements are upscaled to the spatial resolution of the HySpex data while accounting for its geolocation error. For both measurement sites, over water and grass, error estimates are calculated to assess the quality of the airborne HySpex imagery against the upscaled reference pixels from the field measurements. After their validation, spatial resampling is performed to upscale the HySpex data to Sentinel-2 spatial resolution. Following this, surface reflectance values for different landcover classes are sampled from the Sentinel-2 imagery and compared to their respective reference values in the upscaled HySpex data. Finally, the validation of the Sentinel-2 L2A SR product is conducted by calculating Accuracy, Precision and Uncertainty (APU) estimates for all sample pixels and pixels from each landcover class and reported together with uncertainties arising at different steps in the analysis. It ultimately produces uncertainty figures for Sen2Cor, the atmospheric correction algorithm used to produce the Sentinel-2 L2A product. This campaign data-based method supplements the systematic Sentinel-2 L2A SR product validation utilizing radiative transfer models and data from the global AERONET network.

3. Background

To understand the necessary steps for validating the Sentinel-2 Level-2A surface reflectance product, the basics of the Sentinel-2 mission and the product's atmospheric correction and validation processes are explained. Furthermore, the scale problem, spatial effects and sampling issues are discussed. Lastly, an overview of upscaling methods used in the literature is provided.

3.1. The Sentinel-2 Mission

The Sentinel-2 mission is part of the Copernicus program, an earth observation initiative headed by the European Commission (EC) in partnership with the European Space Agency (ESA). It provides services for land management, the marine environment, atmosphere, emergency response, security and climate change (ESA 2019). As part of the Sentinel family, the Sentinel-2 mission consists of two identical satellites tasked to provide high-resolution optical imagery for land services. Both share the same orbit at a mean altitude of 786 km but are phased at 180° to each other to accomplish a revisit time at the equator of only 5 days while covering latitudes of 56° S to 84° N. While Sentinel-2A was launched on June 23rd, 2015, its twin Sentinel-2B was launched on March 7th, 2017. They each carry a passive optical MultiSpectral Imager (MSI) with a swath width of 290 km and 13 spectral bands covering wavelengths between 442.2 and 2185.7 nm; specifically, 4 spectral bands in the visible and near infrared domain at 10 m, 6 in the red-edge and shortwave-infrared domain at 20 m and 3 spectral bands for atmospheric correction at 60 m spatial resolution (ESA 2015; ESA 2019). An overview of the technical details regarding MSI bands are presented in Table 1. Since 2018, systematically produced orthorectified and atmospherically corrected Level-2A imagery has been made available to users through the Copernicus Open Access Hub (MSIL2A). A summary of information on the quality of Sentinel-2 products is provided in the respective Data Quality Report (ESA 2018; ESA 2019).

Table 1. Technical details of the MultiSpectral Imager onboard Sentinel-2B (ESA 2019, adjusted).

Band	Band Designation	Spatial Resolution [m]	Central Wavelength [nm]	Bandwidth [nm]
1	Aerosols	60	442.2	45
2	Blue	10	492.1	98
3	Green	10	559.0	46
4	Red	10	664.9	39
5	VNIR[1] vegetation red edge	20	703.8	20
6	VNIR vegetation red edge	20	739.1	18
7	VNIR vegetation red edge	20	779.7	28
8	NIR[2]	10	832.9	133
8A	VNIR vegetation red edge	20	864.0	32
9	Water vapor absorption	60	943.2	27
10	Cirrus	60	1376.9	76
11	SWIR[3]	20	1610.4	141
12	SWIR	20	2185.7	238

[1] VNIR = visible and near-infrared, 2 NIR = near-infrared, 3 SWIR = shortwave-infrared.

3.2. Atmospheric Correction of Sentinel-2 TOA Data

Atmospheric correction is an important process transforming top-of-atmosphere (TOA) at-sensor radiance into bottom-of-atmosphere (BOA) reflectance suitable for user applications or comparison with data from field spectrometers. The differences in TOA and BOA data stem from light interacting with the earth's atmosphere and its constituents. When radiation from the sun bounces off a surface feature on earth and reaches a spaceborne sensor, it will have passed through the entire thickness of the atmosphere twice (JONES & VAUGHAN 2010). While passing through, atmospheric effects, such as atmospheric scattering and absorption, alter the intensity and spectral composition of the sensed radiation (ACHARYA 2017; ANDREWS 2017; ISLAM et al. 2018). The effects of atmospheric scattering and absorption are explained in more detail in of the Appendix A. The extent to which the sensor signal is distorted depends on the atmospheric path length, the signal's magnitude, the current atmospheric conditions and the wavelengths under investigation. For instance, the signal detected at an airborne sensor will be less distorted than that of a spaceborne sensor due to a significantly shorter atmospheric path length (LILLESAND et al. 2015). When measuring surface reflectance at ground level, a reference panel is used to remove atmospheric effects in the downwelling radiation from the signal (see Appendix C, (13) and (14)). The negligible remaining upward atmospheric path length from ground to sensor allows for virtually no atmospheric distortion in the measured surface reflectance, which is why it is used as reference for validation purposes and sometimes referred to as *ground truth*. However, considering that the field data can still contain measurement errors, the use of the term *ground truth* is misleading and therefore depreciated. Instead, the term *reference data* will be used.

In addition to the interaction with the atmosphere, radiation is split into a reflected, absorbed and transmitted fraction by the surface feature it touches (JONES & VAUGHAN 2010; LILLESAND et al. 2015). Consequently, the signal reaching a sensor is made up by the energy reflected and scattered by the atmosphere itself and the surface feature's reflectance, reduced by atmospheric scattering and absorption. This raw TOA radiance detected at a spaceborne sensor cannot be directly compared to reference field measurements which capture BOA reflectance. The satellite imagery therefore needs to undergo an atmospheric correction process which aims at compensating for the above-mentioned atmospheric effects and produces a close approximation of the BOA reflectance, which can then be validated using reference data from field measurements and is suitable for user applications (LIANG et al. 2012). An overview of atmospheric correction algorithms, as well as an in-detail description of radiation terminology is given in Appendices B and C.

Sentinel-2 data users can decide themselves whether they want to work with the Sentinel-2 Level-1C (L1C) product, representing TOA reflectance, and perform the atmospheric correction themselves using their processing software of choice or weather they directly download atmospherically corrected L2A product in BOA reflectance from the Copernicus Open Access Hub. The L1C data is produced by converting radiance to TOA reflectance, generating quality and cloud masks and computing viewing and solar angles, as well as meteorological data for each tile and including them in the metadata. The TOA reflectance is further reprojected to a cartographic reference frame and undergoes radiometric and geometric correction. To systematically produce the L2A product, each granule of the L1C product undergoes processing with the Sen2Cor processor. Sen2Cor is available as a third-party plugin of the Sentinel-2 toolbox. The three standard operations applied to the L1C imagery are cloud detection and scene classification, retrieval of Aerosol Optical Thickness (AOT) and Water Vapor (WV), and TOA to BOA conversion. Sen2Cor thereby relies on radiative transfer Look-Up Tables (LUTs) and a Digital Elevation Model (DEM) as input (GASCON et al. 2017; ESA 2015). Figure 1 illustrates this processing chain.

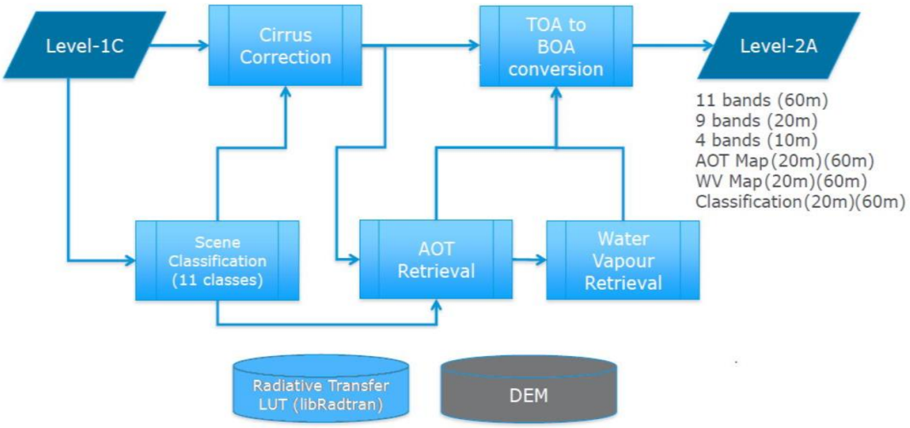


Figure 1. Processing chain of Sentinel-2 Level-2A data (GASCON et al. 2017).

When using Sen2Cor, the user can select options like cirrus, terrain, adjacency and empirical BRDF corrections. However, there is no option for sun glint correction over water surfaces (GASCON et al. 2017). The LUTs have been generated by libRadtran, a freely available collection of C and FORTRAN functions and programs for calculation of atmospheric solar and thermal radiation (ESA 2015; EMDE et al. 2016). By default, 90 m spatial resolution topographic data from the Shuttle Radar Topography Mission (SRTM) (FARR et al. 2007) are used as input DEM. Although not embedded, they are automatically downloaded unless another DEM is provided (GASCON et al. 2017). The scene classification includes 11 classes[1] and is primarily used to distinguish cloudy, clear and water pixels rather than representing a land cover classification map. Together with Quality Indicators (QI) for cloud and snow probabilities (0-100%) at 20 and 60 m spatial resolution, it is used internally for atmospheric correction. For retrieval of WV column in centimeters, the ratio between Band 8A and Band 9 is exploited using a Sentinel-2-adapted Atmospheric Pre-corrected Differential Absorption (APDA) algorithm (GASCON et al. 2017; SCHLÄPFER et al. 1998). Retrieval of AOT at 550 nm is performed using the Dark Dense Vegetation (DDV) pixel method by KAUFMAN & SENDRA (2007). When lacking DDV pixels, a constant AOT value specified by the start visibility set in the configuration file with a default of 40 km is applied. The BOA surface reflectance imagery is provided by Sen2Cor with a scaling factor of 10 000 in different spatial resolutions depending on the band. The VIS and NIR bands 2, 3, 4 and 8 are delivered in 10 m spatial resolution. The red edge bands 5, 6, 7 and 8A, as well as the SWIR bands 11 and 12 together with the resampled VIS bands (2, 3, 4) are provided in 20 m and bands 1 and 9, as well as all other resampled bands except 10 and 8 in 60 m spatial resolution (GASCON et al. 2017).

3.3. Validation of Sentinel-2 Level-2A Product

As part of the Sentinel Core Ground Segment (CGS), the Sentinel-2 Mission Performance Center (MPC) is tasked with the performance assessment and the continued evolution activities concerning all Sentinel-2 data (ESA 2015). The validation results for geometric and radiometric properties of the MSI Level-1C and Level-2A data are released each month in the respective Data Quality Report (ESA 2019). The calibration and validation of the Sentinel-2 Level-2A processor Sen2Cor is conducted by Telespazio France (TPZ-F) and the German Aerospace Centre (DLR), united in the Expert Support Laboratories (ESL) team ESL-L2A (PFLUG et al. 2016). The validation can be divided into three main tasks: validation of cloud screening and scene classification (SCL), AOT and WV products and the BOA reflectance product (GASCON et al. 2017).

Reference data are provided by or derived from the global AERONET project, a federation of ground-based remote sensing aerosol networks initiated by NASA over 25 years ago, which today

---

[1] saturated or defective, dark area pixels, cloud shadows, vegetation, not vegetated, water, unclassified, cloud medium probability, cloud high probability, thin cirrus, snow

has expanded to a wide range of collaborators from national agencies to institutes, universities and other partners. A map of AERONET sites is available on NASA's AERONET website (https://aeronet.gsfc.nasa.gov/). Its database of aerosol optical, microphysical and radiative properties for aerosol characterization is readily accessible to the public and delivers data for the validation of satellite retrievals (HOLBEN et al. 1998; NASA 2019). Validation of the Sentinel-2 L2A products by ESA (2019) was performed using a selection of 25 AERONET test sites for AOT and WV products (ESA 2019) and the dataset from the Atmospheric Correction Inter-comparison Exercise (ACIX) for the BOA SR product (ESA 2019; DOXANI et al. 2018).

For the SCL validation, a representative subset of pixels from the classified scene are selected by means of stratified random sampling ensuring statistical consistency and the inclusion of spatially limited classes. These samples are then visually inspected with the aid of an RGB composite (bands 4, 3 ,2), Color Infra-Red composites emphasizing vegetation (bands 8A, 4, 3) or snow (bands 12, 11, 8A), the spectral profile, as well as the snow and cloud confidence Quality Images, labelled with the corresponding class number and added to a reference database. Eventually, a confusion matrix is calculated by comparing the class according to the SCL with the class assigned in the reference database (GASCON et al. 2017). The reported validation measures include the Overall Accuracy and Errors of Omission and Commission calculated for all individual classes and for consolidated 'valid' (vegetation, non-vegetated, water) and 'cloud' classes (cloud medium probability, cloud high probability, this cirrus) (ESA 2019).

The validation of AOT and WV is conducted by directly comparing the Sen2Cor output with the AERONET sunphotometer measurements. The Sen2Cor product values are averaged over an area of 9 km by 9 km around the AERONET station. For WV validation, all soil and vegetation pixels are used, while AOT uses additional water pixels. The results are illustrated in a graph correlating the AERONET reference (x) with Sen2Cor product averages (y) for both cloudy (> 5 %) and cloudless (< 5 %) conditions for the same area (ESA 2019; GASCON et al. 2017). The Level-2A Data Quality Report further includes histograms of AOT and WV difference to their respective AERONET reference values including a fitted precision and uncertainty curve alongside a table listing additional statistics (ESA 2019). The presence of DDV pixel has proven to be particularly important for an accurate AOT retrieval (KAUFMAN & SENDRA 2007). As AOT and WV are key parameters in the atmospheric correction process, their retrieval accuracy can strongly affect the quality of the BOA SR product (GASCON et al. 2017).

The remaining task is the validation of the BOA SR product. To this end, DOXANI et al. (2018) created reference data by applying an atmospheric correction to L1C subsets of 9 km by 9 km around the AERONET station using the 6S radiative transfer code (see Appendix B) with AOT, aerosol model and column water vapor derived from AERONET sunphotometer measurements as inputs. The results from comparing the Sen2Cor BOA SR product with the reference surface reflectance data are expressed in terms of Accuracy (A), Precision (P) and Uncertainty (U) values (DOXANI et al. 2018). The accuracy value describes the mean difference to the reference value, the precision value represents the variation around it and the uncertainty value is the quadratic sum of the former two (ESA 2019). The respective formulas are given below ((1)-(4)), where $\rho_{i,\lambda}^{SR}$ is the surface reflectance of observation i at wavelength λ and n_λ number of observation at wavelength λ:

$$\Delta\rho_{i,\lambda}^{SR} = (\rho_{i,\lambda}^{SR_{PROCESSOR}} - \rho_{i,\lambda}^{SR_{AERONET}}) \tag{1}$$

$$A = \frac{1}{n_\lambda}(\sum_{i=1}^{n_\lambda} \Delta\rho_{i,\lambda}^{SR}) \tag{2}$$

$$P = \sqrt{\frac{1}{(n_\lambda - 1)}\sum_{i=1}^{n_\lambda}(\Delta\rho_{i,\lambda}^{SR} - A)^2} \tag{3}$$

$$U = \sqrt{\frac{1}{n_\lambda}\sum_{i=1}^{n_\lambda}(\Delta\rho_{i,\lambda}^{SR})^2} \tag{4}$$

APU values are calculated for 0.01-wide SR bins and plotted in APU histograms per band (DOXANI et al. 2018). The Sentinel-2 L2A Data Quality Report further provides absolute and relative APU band averages over all pixels for the same dataset but with a different version of Sen2Cor. The respective graph of absolute APU values also includes band averages of reference SR and specifications. The relative APU graph presents accuracy, precision and uncertainty relative to the average surface reflectance reference per band (ESA 2019).

When reference data is calculated using radiative transfer models, uncertainties inherent in the model will propagate into the resulting data. Simulations have shown uncertainties connected to radiative transfer codes of up to 1%, which is considerable (KOTCHENOVA et al. 2006; KOTCHENOVA & VERMOTE 2007). Therefore, validation studies not relying on surface reflectance derived from radiative transfer models, but on actual surface reflectance measurements from the field is needed to supplement the evaluation effort. Providing precisely such a supplementary, radiative transfer model-independent validation is one objective of this paper.

3.4. Implications of Scale

Scale is a central issue not only in the validation and calibration of models and datasets in different fields of study (ATKINSON 2013; WU & LI 2009), but in any type of spatial analysis (O'SULLIVAN & UNWIN 2010). The smaller the scale, the higher the level of detail present in the derived information (ATKINSON 2013). In remote sensing, scale can refer to the spectral scale, i.e. spectral resolution (hyperspectral versus multispectral data), or spatial scale, i.e. spatial resolution (high- and medium-resolution data). Although the scale concept is applicable to both spatial and spectral units, it is most often used in the spatial context. O'SULLIVAN & UNWIN (2010) give an illustrative example of a city, which on a continental scale is represented by a point, but when zooming in to regional scale becomes an area object and to local scale a complex collection of point, line, area and network objects. They further warn that the scale we work with affects the representation and therefore the analysis undertaken (O'SULLIVAN & UNWIN 2010).

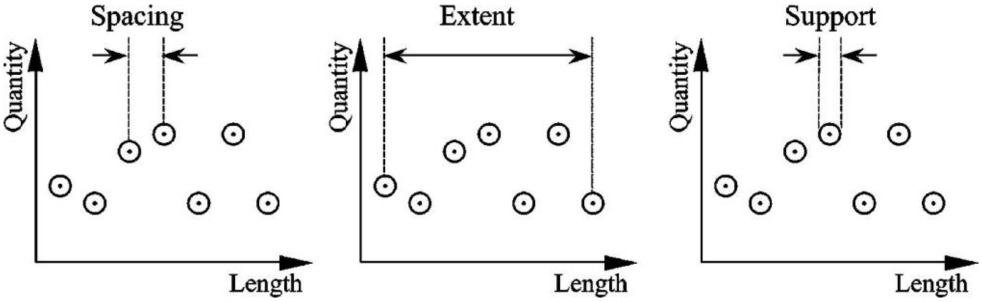


Figure 2. The scale triplet (after BLÖSCHL & SIVAPALAN (1995) in WESTERN & BLÖSCHL (1999)).

As illustrated in Figure 2, spatial scale consists of three main characteristics: spacing, extent, and support. Their combined effect acts like a filter and is relative to the scale of the natural variability (BLÖSCHL & SIVAPALAN 1995). The spacing refers to the distance between the measurement points or between computational points in a model. While the variability remains unaffected with an increase in spacing, the apparent spatial size of features increases and the detail decreases. The extent describes the total coverage of the measurements or model. An increase in extent leads to the inclusion of larger features into the data, as well as an increase in variability and average size of the feature (WESTERN & BLÖSCHL 1999). The support is the space (or time) over which each observation averages the underlying variations. Therefore, an increase in support will decrease the variability in the data due to averaging (WESTERN & BLÖSCHL 1999). Sometimes even shape and orientation of the support are important. For instance, in remotely sensed imagery, the support is characterized by its point spread function approximated by a Gaussian weighting function (ATKINSON & TATE 2000; JONES & VAUGHAN 2010). As it assigns more weight to the center of the support, i.e. pixel, than to its edges, it places emphasis on the underlying spatial variation at the pixel center (ATKINSON & TATE 2000). To best capture the real variability of the observed feature, spacing and support would have to be minimal and the extent very large (WESTERN & BLÖSCHL 1999). A common issue to arise when dealing with regression analysis or comparisons of spatial data is the mismatch in scales of the support. In remote sensing, this problem presents itself when raster imagery is to be compared with point measurements from the ground, as it is done for the validation of remote sensing products. To make a valid comparison, the data with the smaller support, here the ground measurements, need to be scaled up to the same scale of support as the large-support data, i.e. the remotely sensed imagery (ATKINSON & TATE 2000).

For a long time, the change of the support has been the subject of much research (GEHLKE & BIEHL 1934; OPENSHAW 1984; FOTHERINGHAM & WONG 1991; DARK & BRAM 2007; WU & LI 2009; GOODCHILD 2011; SALMIVAARA et al. 2015). Changing the support by aggregation of data originally compiled at a more detailed level has implications for data statistics. Pattern and relationships may vary significantly depending on the scale selected for a study (O'SULLIVAN & UNWIN 2010). First discovered in the 1930s by GEHLKE & BIEHL (1934), this effect was later referred to as the Modifiable Areal Unit Problem (MAUP) (OPENSHAW & TAYLOR 1979). The MAUP effect can be attributed to two things: the aggregation effect, which refers to the combination of observations, which will produce an outcome closer to the mean of the overall data; and the zoning effect, which represents the differences in results under different aggregation schemes (O'SULLIVAN & UNWIN 2010). Regression relationships between two variables are strengthened by aggregation of small support areas into larger ones (GOODCHILD 2011; OPENSHAW & TAYLOR 1979). However, there is no increased significance, as the improvement in correlation is counterbalanced by the loss of degrees of freedom (GOODCHILD 2011). In terms of zoning effect, OPENSHAW and TAYLOR were able to demonstrate that correlations anywhere between extremely negative and extremely positive can be produced by repeatedly reaggregating the same county data for Iowa in different ways (OPENSHAW & TAYLOR 1979). Most famously, this effect is exploited by US politicians when redrawing electoral district boundaries to maximize their votes (O'SULLIVAN & UNWIN 2010), also known as gerrymandering. Since the MAUP effect can result in substantial differences in conclusions and recommendations, it has major implications for decision-making for policy and planning (SALMIVAARA et al. 2015; O'SULLIVAN & UNWIN 2010).

Although many attempts have been made to tackle the MAUP problem (DARK & BRAM 2007; OPENSHAW 1984), no generic solution has been found that is available across different types of analysis and disciplines. Neither has a comprehensive solution been found capable of easily and accurately quantifying the MAUP effect (OPENSHAW 1984). However, acknowledging the problem is a first step and for correlation analysis, searching for stable results across different aggregation scales may make results more reliable (WONG 2009). The MAUP is a systematic phenomenon that concerns practically any study using spatially-referenced data. Nevertheless, it does not get the attention it deserves (OPENSHAW & TAYLOR 1979; SALMIVAARA et al. 2015; GOODCHILD 2011). OPENSHAW (1983) suggests that a lack of understanding the MAUP effect has led many to neglect this issue. He therefore advises to treat the MAUP as an exploratory and descriptive tool, as has been done with spatial autocorrelation (OPENSHAW 1983).

3.5. Spatial Effects and Sampling

Common spatial effects occurring across scales are nonuniformity of space, distance, adjacency, interaction, and neighborhood effects (O'SULLIVAN & UNWIN 2010). Nonuniformity of space simply refers to spatial heterogeneity, or in other words the fact that space is not uniform (O'SULLIVAN & UNWIN 2010). Spatial heterogeneity may not be constant across scales. Variables can be homogeneous at one scale, but heterogenous at another. Likewise, heterogeneity may be increased with an increase in spatial extent of the study area. It is therefore important to understand the nature of the underlying spatial variation in order to change the scale of data (ATKINSON & TATE 2000).

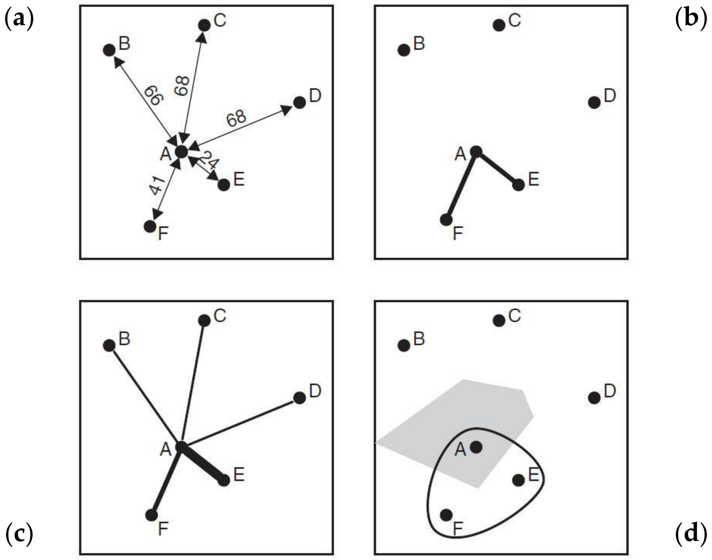


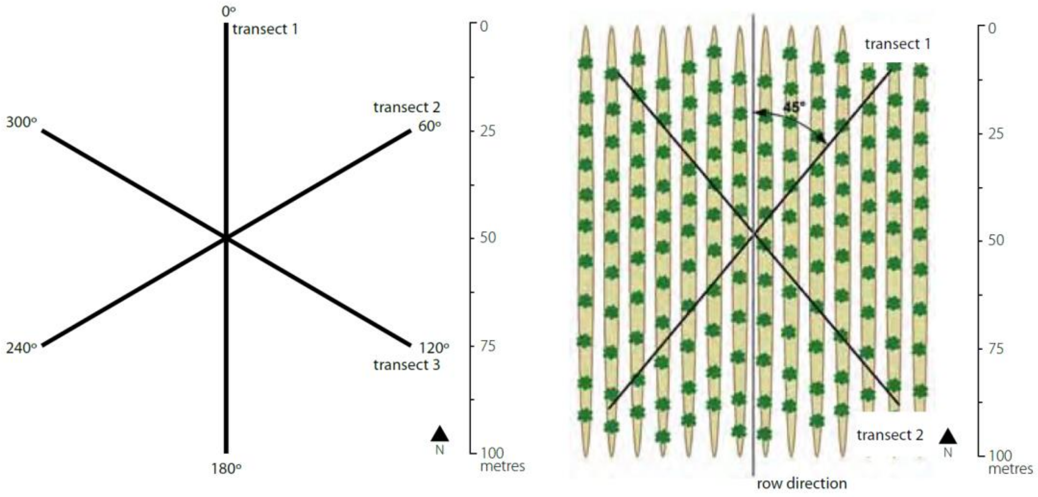
Figure 3. A schematic representation of (a) distance, (b) adjacency, (c) interaction, and (d) neighborhood concepts. (O'SULLIVAN & UNWIN 2010).

Figure 3 illustrates the concepts of distance (a), adjacency (b), interaction (c) and neighborhood (d) (O'SULLIVAN & UNWIN 2010). For small study areas, the distance between points is usually measured as Euclidean distance, while for larger study areas the curvature of the Earth has to be considered. The binary representation of distance is adjacency, which describes whether two points or other spatial entities are adjacent to each other or not. This concept is important for interpolation and measuring autocorrelation (O'SULLIVAN & UNWIN 2010). The notion of nearer things being more closely related than distant things describes the concept of interaction and is also referred to as the First Law of Geography (TOBLER 1970). Usually, some form of inverse distant weighting is applied to determine interaction between spatial entities resulting in closer entities to have stronger interactions (O'SULLIVAN & UNWIN 2010). In terms of geostatistics, interaction across space manifests itself as spatial autocorrelation (CLIFF & ORD 1973). The increased likelihood of locations near each other being more similar than locations far from another introduces redundancy into the data as each additional data point included will provide less new information (O'SULLIVAN & UNWIN 2010). Typical measures of spatial autocorrelation are the Moran's I ((22)-(24) in Appendix E) (MORAN 1948) and Geary's C (CLIFF & ORD 1973) (O'Sullivan&UnwinÉ. Results can indicate a positive autocorrelation, negative autocorrelation or zero autocorrelation. A positive autocorrelation refers to nearby observations being similar to one another, while negative autocorrelations refers to them being different from one another. Cases where values are random across space have zero autocorrelation (CLIFF & ORD 1973; O'SULLIVAN & UNWIN 2010). The function describing the nature of the spatial interaction, also referred to as spatial dependence (ATKINSON & TATE 2000), is called the *semivariogram*. It illustrates the semivariance between observations with distance (lag). Semivariance is defined as the half the expected squared difference between two observations of a certain distance (ATKINSON & TATE 2000). Statistical interpolation methods, such as Block Kriging (Appendix E), use semivariograms to make predictions for a different spatial support than that of the observations (BILL 2010). The last spatial effect presented in Figure 3 (d) is the neighborhood effect. The concept of neighborhood can mean different things. It can define all spatial entities adjacent to a certain entity, all spatial entities within a certain distance of that entity or even all surrounding entities sharing the similar traits that make this neighborhood distinctly different from other neighborhoods (O'SULLIVAN & UNWIN 2010).

In practice, the above-mentioned spatial concepts have implications for the sampling of field (or remotely sensed) data. Not only should the sample size be large enough to appropriately reflect the spatial heterogeneity, but the spacing between observations also needs to be sufficiently sparse to reduce redundancy in the data. This keeps autocorrelation low and, because of the increased distance between adjacent observations, reduces their level of interaction. Moreover, an observation should be internally uniform, i.e. only contain the surface of interest and not be contaminated by other features. The feature of interest is usually a neighborhood distinctly different from other neighborhoods, e.g. water versus grassland.

By sampling data in a certain way, their real representation is filtered leaving only the selected samples to draw conclusions from (Atkinson & Tate 2000; Blöschl & Sivapalan 1995). Key elements of a sampling framework are the sampling scheme, sampling density, sample size and spatial coverage. The sampling scheme refers to spatial pattern in which observations are sampled, such as random, stratified random, systematic or equilateral triangular schemes. The number of observations per unit area is the sampling density and the total number of observations sampled is called the sample size. These three elements together determine the spatial coverage, the distribution in terms of distance and direction of the sample observations. Remotely sensed data, for instance provides complete, uninterrupted, spatial coverage (Atkinson & Tate 2000).

Finding a sampling framework suitable for changing the support across scales, for example from point measurements to pixel footprint, is particularly difficult. While Johannsen & Daughtry (2009) summarize considerations for planning and acquiring reference data for agricultural application, Muir et al. (2011) developed a sampling framework for vegetative and non-vegetative fractional landcover in Australia. It is optimized to change the support of field data to accurately simulate pixel sizes from Landsat and MODIS imagery facilitating their validation. Building on sampling approaches by Brady et al. (1995), Scarth (2006) and Schmidt et al. (2010) , Muir et al. (2011) introduce two modified discrete point transect sampling methods; one for natural and pastoral surfaces (Figure 4 (a)) and one for vegetation sown in rows (Figure 4 (b)).



(**a**) (**b**)

Figure 4. Sampling transects optimized for change of support: (a) Star-shaped transects for natural and pastoral surfaces; (b) Cross-shaped transects for vegetation sown in rows (Muir et al. 2011).

To minimize within-site variation when selecting a measurement site, they recommend representative and homogeneous sites with minimal topographic variation. As their intention is to link field measurements to medium-resolution Landsat and MODIS imagery, with spatial resolutions of 30 m and 500 m respectively, a site should be at least 100 times 100 m in dimensions, equivalent to an area of about three times three Landsat pixels. Ideally, the site should be located where the homogenous area around the site extends to a total area of 500 times 500 m to allow for a change of support to MODIS pixel size. Visual inspection of imagery may aid the selection process. The number of sampling sites depends on both the objective and the landscapes occurring in the scene, however a minimum of five sites should be acquired (MUIR et al. 2011). Surface features or classes characterized by a high variability may require more sampling sites (JOHANNSEN & DAUGHTRY 2009). To make the best use of observations, they should be distributed over multiple sites rather than all being part of one site (MCCOY 2005). The acquisition time of the measurements should be set as close as possible to that of the target imagery. Seasons, land management phases and phenology should also be taken into account. Per site, a sampling size of 300 observations for natural vegetation and pastures (SCARTH 2006) and 200 for vegetation in rows is recommended (SCHMIDT et al. 2010). The sampling scheme suggested by MUIR et al. (2011) for natural vegetation and pastures, shown in Figure 4, consists of three 100 m transects laid in a star-shape at 60° to one another starting with a North-to-South transect. Observations are made every meter of each transect, providing a sample size of 300 observations (3*100). For vegetation in rows, such as cropland, only two 100 m transects are laid at angle of 45° to the sowing lines and sampled every meter along the transect. In total, this procedure produces a sample size of 200 observations (2*100). MUIR et al. (2011) also provide a checklist for the equipment needed to apply the sampling scheme in the field.

3.6. Upscaling of Field Measurements

For comparing information from multiple datasets of different resolutions, as is required for validating a remote sensing product, spectral and spatial scales must match. While spectral matching can be achieved by spectrally resampling data using the spectral response function of the target data sensor, the matching of spatial scales is more complex. The decrease or increase in spatial resolution to create data with matching spatial scales is also referred to up- or downscaling (ATKINSON & TATE 2000; ATKINSON 2013). The principles of up- and downscaling are illustrated in Figure 5 (pdf = point distribution function).

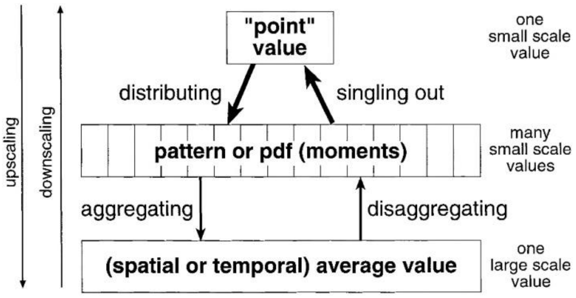


Figure 5. Graphical representation of the processes of upscaling and downscaling (from BLÖSCHL & SIVAPALAN (1995), reproduced with permission by ATKINSON & TATE (2000)).

The validation of remotely sensed products is a particularly common objective in the upscaling efforts (BACCINI et al. 2007; CROW et al. 2005; CROW et al. 2012; LOEW & SCHLENZ 2011; MIRALLES et al. 2010; SHI et al. 2015; WANG et al. 2015). Thereby higher resolution reference data is upscaled to the spatial resolution of the remote sensing product. As surfaces may appear homogeneous on one scale, for example in remote sensing imagery, they may reveal substantial heterogeneity on a different scale, e.g. among field measurements (ATKINSON & TATE 2000). Scale-dependent heterogeneity has long been identified as one of the most fundamental constraints in multiscale comparisons and spatial data integration (QUATTROCHI & LAM 1991). Consequently, the core motivation of upscaling data is the increase in support to approximate its representation at a coarser scale while considering its spatial heterogeneity (KING 1991). However, a thorough investigation of the underlying spatial variation in the data needs to be conducted before rescaling is attempted (ATKINSON & TATE 2000). (Semi)Variograms are essential to assess spatial variation for data rescaling. They give important information on spatial dependence, statistically represented as spatial autocorrelation (ATKINSON & TATE 2000; BILL 2010). Even when this information is available, the spatial dependence of a selected measurement site may prove to be unique impeding generalization (DAVIS et al. 1991).

In order to upscale higher resolution reference data to a certain pixel size, the position of the target pixel borders need to be known. In other words, we need to know which reference samples relate to which target pixel. Then, some sort of aggregation can be applied to all reference observations covered by the target pixel. If samples do not completely fill the target pixel, the empty pixel part needs to extrapolated and if samples lie within more than one pixel, disaggregation of the respective sample becomes necessary (LEITÃO et al. 2018). The potential geolocation error inherent in many remote sensing products complicates the upscaling process significantly. If it is unclear how reference data and remote sensing product pixels align, the position of pixel borders for data aggregation is also unknown. One solution to this is using regression analysis to find the best fitting sample-to-pixel allocation (ZANDLER et al. 2015).

Through the increasing integration of multiscale remote sensing data, developing appropriate upscaling techniques has become an important issue in a range of disciplines. While many studies investigate the upscaling of variables, such as methane emissions in the Arctic (DAVIDSON et al. 2017), soil moisture (WESTERN et al. 2002) globally (CROW et al. 2012), in Germany (LOEW & SCHLENZ 2011), in the US (CROW et al. 2005; CROW et al. 2012; JOSHI et al. 2011; LANNOY et al. 2007; MIRALLES et al. 2010) and in China (WANG et al. 2015; ZHANG et al. 2018), carbon stocks in Argentina (GONZÁLEZ-ROGLICH & SWENSON 2016) and Brazil (LEITÃO et al. 2018), biomass in the US (BACCINI et al. 2007) and vegetation related indices and properties in the US (ANDERSON et al. 2004) and China (SHI et al. 2015), no body of literature appears to exist on upscaling spectra, from which those variables are generally derived.

Two papers reviewing different upscaling techniques are WESTERN et al. (2002) and CROW et al. (2012). While WESTERN et al. (2002) discuss behavioral techniques relying on statistical analysis, e.g. Ordinary Kriging, and process-based techniques relying on process understanding within a deterministic reductionist framework, CROW et al. (2012) evaluates enhanced upscaling techniques involving time stability concepts, field campaign data and land surface modeling, as well as Block Kriging. Not only the use of multi-temporal data is quite common for the upscaling process (ANDERSON et al. 2004; LANNOY et al. 2007), but also the use of high-resolution imagery to bridge the spatial resolution gap between ground measurement and target resolution (ANDERSON et al. 2004; LEITÃO et al. 2018; GONZÁLEZ-ROGLICH & SWENSON 2016). Frequently, more than one upscaling step is applied or multi-scale pixel estimations are attempted (ANDERSON et al. 2004; GONZALEZ-ROGLICH & SWENSON 2016; SHI et al. 2015; ZHANG et al. 2018). The nature of the field data to be upscaled differs among the variety of studies. There is upscaling from point to pixel (CROW et al. 2005; CROW et al. 2012; LANNOY et al. 2007; SHI et al. 2015; WANG et al. 2015; ZHANG et al. 2018), as pursued in this paper, point to plot (ANDERSON et al. 2004) and often also from plot to pixel benefitting from a somewhat averaged measure of the variable across the area of each plot (BACCINI et al. 2007; DAVIDSON et al. 2017; LEITÃO et al. 2018; GONZALEZ-ROGLICH & SWENSON 2016). An evaluation of the upscaling error connected to the upscaling of soil moisture data is provided by LOEW & SCHLENZ (2011) and MIRALLES et al. (2010). LOEW & SCHLENZ (2011) used a temporally adaptive technique and MIRALLES et al. (2010) a triple collocation approach. Both approaches make use of multi-temporal data and land surface models, not available for this study.

4. Data

Field measurements and airborne imagery used in this study were collected near Lake Stechlin on May 4th, 2018 during a joint campaign between the German Aerospace Center (DLR) and the Leibnitz Institute of Freshwater Ecology and Inland Fisheries (IGB). Lake Stechlin is located Northwest of Berlin in the North of the German Province Brandenburg at 53°9'6'' N and 13°1'34'' E. The campaign date and time (mid-morning) were synchronized with the fly-over date and time of both the Sentinel-2B and Landsat 8 satellite to obtain optimal reference data for the respective satellite products.

4.1. Field Mearurements near Lake Stechlin

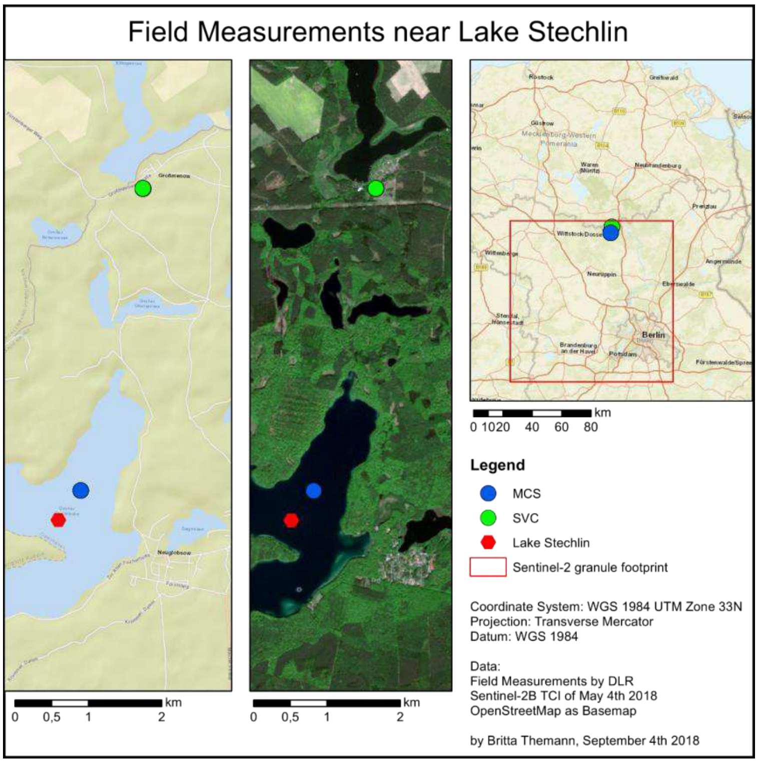


Figure 6. Field Measurements near Lake Stechlin.

The field measurements acquired near Lake Stechlin are shown in Figure 6 and serve as reference data for the validation of airborne and spaceborne imagery. Among other campaign measurements excluded from this study, reference data for the water surface and body was collected using the Zeiss Multi Channel Spectrometer (MCS) 55 UV-NIR, an ASD spectrometer, a profiler sonde for various water properties and constituents, and by conduction laboratory analysis of water samples. For this study, only the MCS data were used.

The measurements were taken at the deepest part, roughly in the center of Lake Stechlin at around 53°9'19"N 13°1'52"E. The sensor was mounted onto a crane-like structure and fixed on the side of a boat to capture the water surface over which it hovered. Between 9:28 and 11:00 am, 12 measurements were taken at the same coordinates over the relatively calm water surface. At those coordinates, the MCS measurements and the campaign boat share the same Sentinel-2 (60 m resolution) pixel footprint. To avoid including the reflectance from the boat into the upscaling process, the MCS measurement coordinates have been artificially placed, each at slightly different coordinates, past the closest Sentinel-2 pixel border into the neighboring pixel. This way, the water measurements will be upscaled to a pure water pixel, rather than a mixed pixel containing boat and water. As water constantly moves, assigning each measurement to different coordinates is not expected to have a large impact on the validation results. The MCS measurements at their newly assigned coordinates are shown in Figure 7.

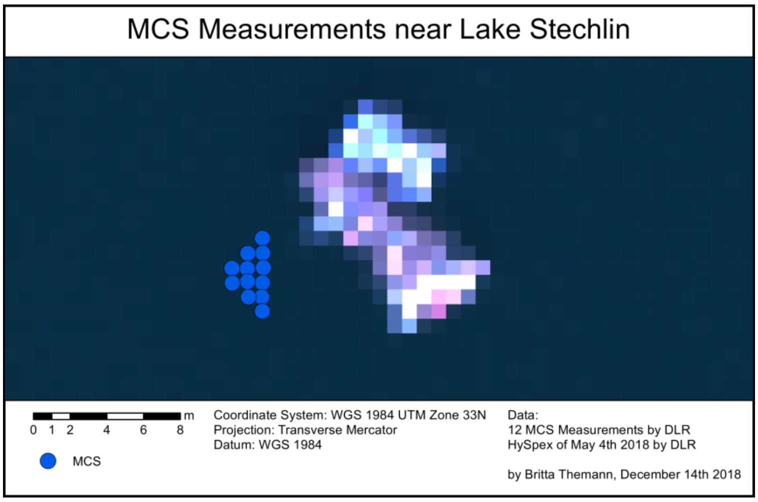


Figure 7. MCS Measurements near Lake Stechlin.

Reference data for land surfaces were provided by a Spectra Vista Corporation (SVC) spectrometer, type HR-1024i. It was used to collect 22 measurements over grass on a meadow around 5 km North of Lake Stechlin at around 53°11'33" N 13°2'29" E, with at least 30 m distance to the bordering forest North and West of it. As illustrated in Figure 8, the first SVC measurement was taken at 9:48 am and the last at 10:40 am. The meadow is characterized by mixed patches of healthy and dried grass, as well as bushes and trees scattered over the meadow Northeast, East and South of the SVC measurement site.

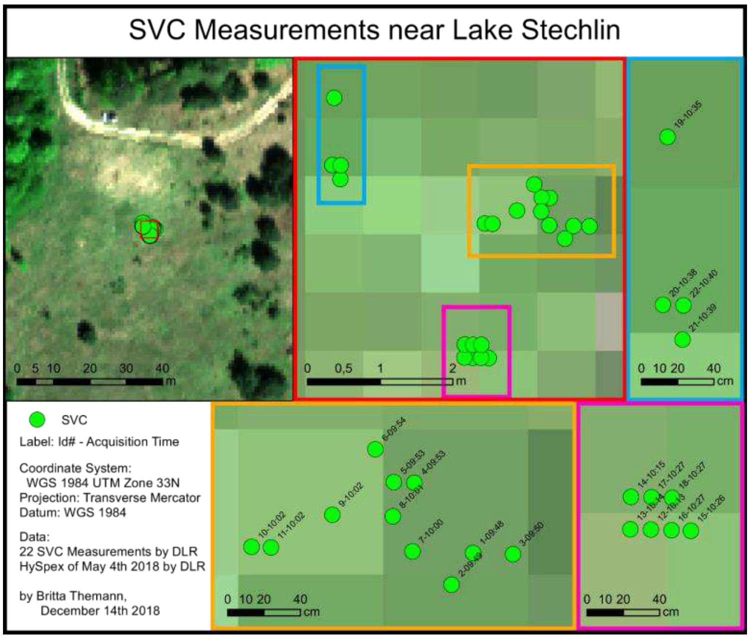


Figure 8. SVC Measurements near Lake Stechlin.

4.2. *HySpex Data*

The hyperspectral imagery, seen in Figure 9 (a), was captured by the HySpex system onboard the DLR aircraft Do228. It consists of two pushbroom imaging spectrometers: VNIR-1600 and SWIR-320m-e; the former covering the visible and near infrared spectral domain (416.00 to 988.40 nm) with 160 spectral bands, the latter covering the short infrared spectral domain (967.79 to 2496.53 nm) with 256 spectral bands. The scanners are facing nadir from an altitude of between 1000 to 2000 m and record all across-track pixels within a swath of between 1.6 and 3.2 km simultaneously at spatial resolutions of around 1 to 2 m. The resulting flight line imagery is the product of successively combining the across-track frames while the aircraft steadily moves along the track (DLR 2015).

During the campaign near Lake Stechlin, 10 flight lines were acquired, 2 from 2600 m altitude, 4 from 1000 m altitude with an integration time optimized for water surfaces and 4 from 1000 m altitude with an optimized integration time for land surfaces. For this study, flight line 3 acquired at 9:51 am on May 4th, 2018 at ~1000 m altitude with an optimized integration time for water surfaces and a spatial resolution of 0.8 m was selected. It is approximately 11 km long and 700 m wide. Extending from 53°7'24" N 13°1'21" E (Southwest corner) and 53°7'20" N 13°1'58" E (Southeast corner) to 53°13'17" N 13°2'48" E (Northwest corner) and 53°13'13" N 13°3'25" E (Northeast corner), it covers both field measurement sites and is not impacted by sun glint near the water measurements. Each flight line produces two images, one from the VNIR sensor and one from the SWIR. From both sensors, orthorectified Level-1C at-sensor radiance data and Level-2A surface reflectance data were provided separately (DLR 2015). The L2A SR product has previously undergone terrain correction using a DEM and atmospheric correction. The latter was performed using the ATCOR4 model based on atmospheric LUTs generated by the radiative transfer model MODTRAN4 and a retrieved average AOT value at 550 nm of 0.119 (DLR 2015). This study was limited to HySpex Level-2A VNIR data, which serve as auxiliary data for the upscaling of field measurements to Sentinel-2 pixel size.

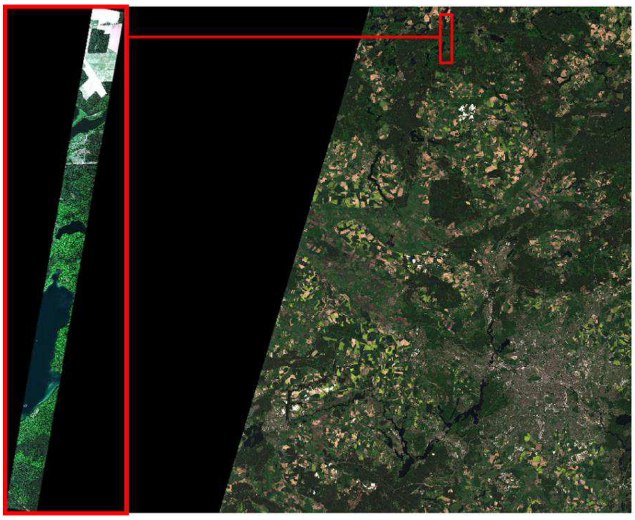

(**a**) (**b**)

Figure 9. Airborne HySpex data and spaceborne Sentinel-2 data: (a) HySpex flight line 3, acquired at 9:51 am on May 4th, 2018 at an altitude of 1000 m and a spatial resolution of 0.8 m, covering the field measurement sites near Lake Stechlin; (b) Sentinel-2B imagery acquired at 10:10 am on May 4th, 2018 with spatial resolutions of 10, 20 and 60 m, covering the field measurement sites near Lake Stechlin (Relative Orbit number R022, tile number field T33UUU).

4.3. Sentinel-2B Data

The Sentinel-2 image tile to be validated, shown in; Figure 9 (b), was acquired by the MultiSpectral Instrument (MSI) onboard the Sentinel-2B satellite at 10:10 am on May 4th, 2018 (Relative Orbit number R022, tile number field T33UUU). From South to North and West to East, it covers an area from Beelitz to Lychen and from Brandenburg an der Havel to the eastern end of Berlin (from 52°13'58" N 12°21'29" E and 52°15'18" N 13°40'42" E to 53°13'44" N 12°47'25" E and 53°14'31"N 13°38'53" E). Each band is delivered as a 100 km by 100 km ortho-image tile (or granule) in UTM/WGS84 projection (ESPG 32633) (ESA 2015). The geolocation error at the date of acquisition was slightly over 12 m for Level-1C or higher level data (ESA 2018). This study included both TOA Level-1C and BOA Level-2A reflectance. The latter were produced by applying the atmospheric correction algorithm Sen2Cor within the Sentinel-2 Toolbox.

5. Methodology

This chapter describes in more detail the steps undertaken to validate the Sentinel-2 Level-2A surface reflectance product. All processing was conducted using the open source programming language R with minor tasks solved using the open source geographic information system QGIS. First, the preparation of field and HySpex data is described and the resulting data are introduced. Following this, methods for the upscaling of field measurements, the validation of HySpex data and their spatial resampling, as well as HySpex and Sentinel-2 pixel sampling are explained before elaborating on the validation of the Sentinel-2 Level-2A surface reflectance product.

5.1. Data Preparation

The MCS spectra collected over water were provided as radiance reflectance. To account for irradiance coming from the entire hemisphere rather than a single stream of light at an infinitesimally small angle, it was converted to irradiance reflectance by multiplying with pi (see Appendix C) (NICODEMUS et al. 1977; SCHAEPMAN-STRUB et al. 2006). Figure 10 illustrates the resulting MCS irradiance reflectance spectra over water. For all further analysis, the term surface reflectance designating surface irradiance reflectance was adopted for its brevity.

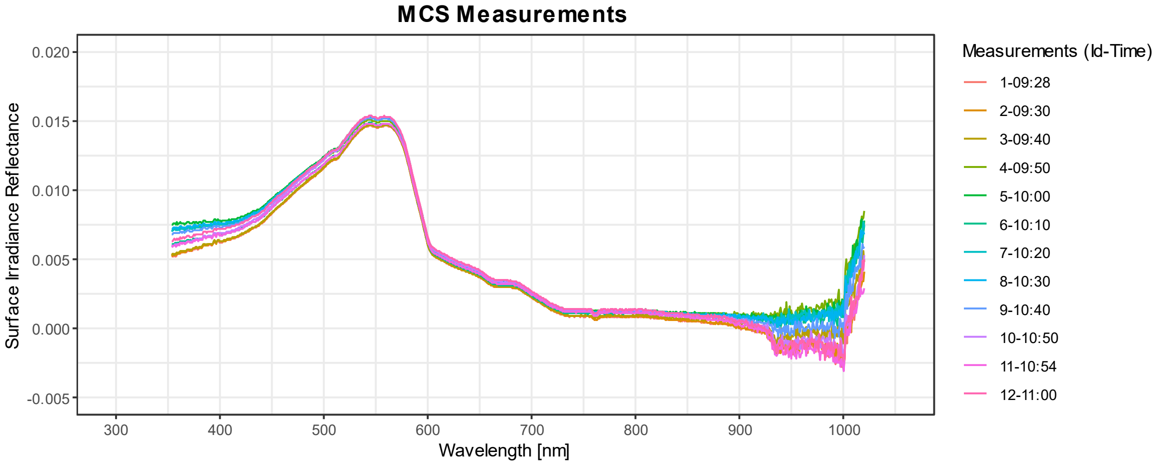


Figure 10. MCS surface irradiance reflectance spectra over water.

The provided SVC reflectance data over grass displays visible data jumps where changes between temperature-dependent sensors have occurred. As only the spectral domain between 411 and 907 nm, equivalent to Sentinel-2B Band 1 to 8A, were used for further analysis, these jumps and the data uncertainty they cause (see Appendix D) do not affect any of the study's results. Furthermore, the raw SVC measurements were obtained as a ratio of radiance over grass and radiance emitted from a white Spectralon reference panel (13). Although the reference panel is supposed to serve as a perfect Lambertian reflector, it rarely reflects all the incoming radiation. Consequently, the panel's reflectance factor, i.e. its absolute reflectance, was determined in a laboratory and multiplied with the provided reflectance to calibrate the SVC measurements (14). The mean calibration uncertainty resulting from picking one of two different sets of reference panel measurements from the laboratory is 0.00451. Figure 11 (a) shows the SVC measurements' mean spectra before (dark green) and after calibration (light green). A slight decrease in surface reflectance can be observed across the entire spectrum. The resulting SVC measurement spectra are illustrated in Figure 11 (b).

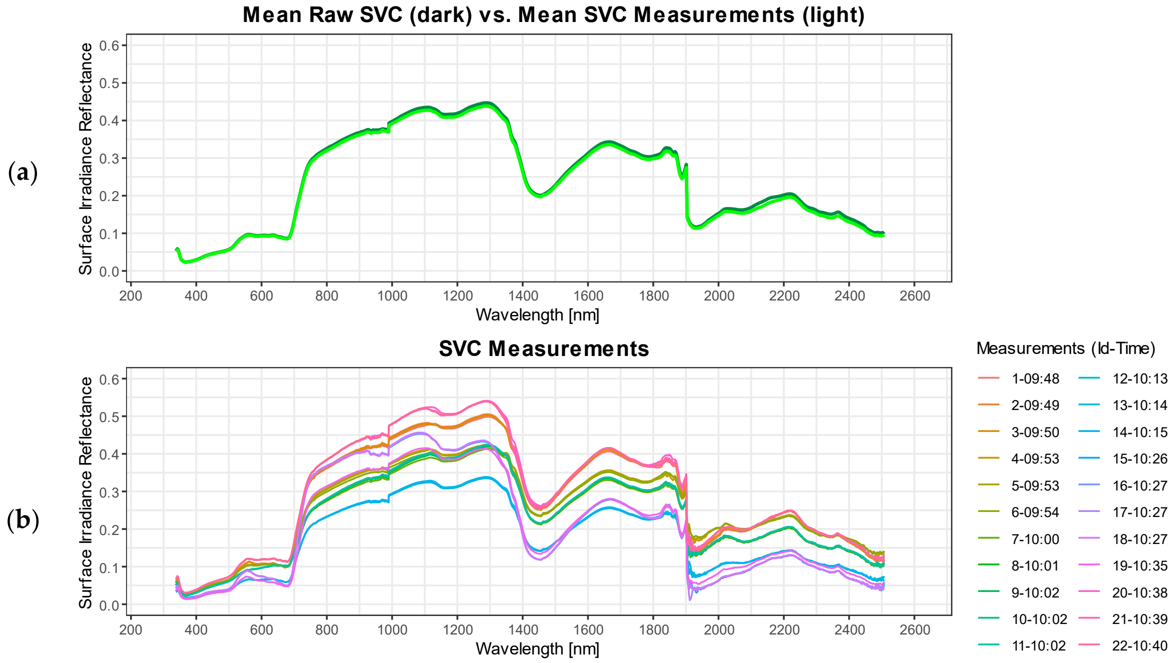


Figure 11. Raw and calibrated SVC spectra: (a) Decrease in mean surface irradiance reflectance after calibration of the raw SVC spectra; (b) SVC surface irradiance reflectance spectra over grass.

To compare surface reflectance values from different sensors, they need to share the same spectral resolution. Hence, MCS, SVC and HySpex reflectance data were spectrally resampled to match Sentinel-2B spectral bands using the Sentinel-2B spectral response function (ESA 2017). Figure 12 shows the spectrally resampled field measurements. Throughout the study, only Sentinel-2B bands 1 to 8A will be evaluated.

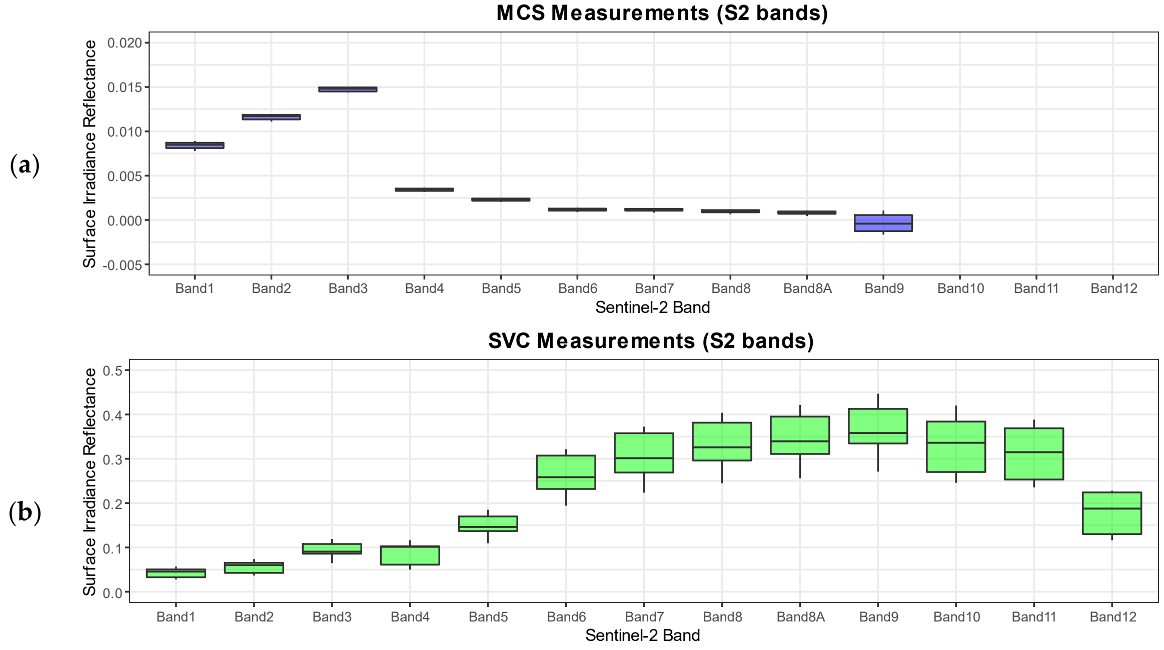


Figure 12. MCS (a) and SVC data (b) spectrally resampled to Sentinel-2B spectral resolution. The boxes represent the 25 % and 75 % quantiles, the horizontal line within the median and the whiskers extend towards minimum and maximum values excluding outliers.

5.2. Regression-based Upscaling using Simple Averaging

Regarding the validation of the Sentinel-2 L2A product, the mismatch in spatial scales between field measurements, high-resolution HySpex and medium-resolution Sentinel-2 remote sensing data has to be overcome by upscaling reference data from the ground and airborne data to Sentinel-2 spatial resolution. In this study, the term *upscaling* refers to the change of spatial scales from point measurements on the ground to HySpex pixel size and from HySpex to Sentinel-2 pixel size. Strictly speaking, the spectrometer measurements in the field are not point measurements with an infinitesimally small spatial support but have a support of ~10 cm. To simplify the upscaling in this study, field measurements are interpreted as point measurements. This issue is further examined in the Discussion. The alignment of spatial scales between HySpex and Sentinel-2 pixels was performed by spatially resampling the HySpex raster data using simple averaging.

One in the literature recurring upscaling technique is Block Kriging (ATKINSON & TATE 2000; CROW et al. 2012; WANG et al. 2015; ZHANG et al. 2018). This statistically linear unbiased weighted interpolation technique is widely used to upscale data over larger supports than those of the original observations. It has two major advantages; it uses spatial dependency between observations represented in a variogram for the calculation of weights and thereby minimizes the estimation variance, and it automatically accounts for the size of the support (ATKINSON & TATE 2000; BILL 2010). Other upscaling techniques have been dismissed due to the requirement of multitemporal data or other auxiliary datasets not available in this context (CROW et al. 2012; WESTERN et al. 2002). Although it appears that Block Kriging has never been applied to surface reflectance, its good performances for variables derived from surface reflectance indicate its suitability for upscaling point measurement spectra from the field to HySpex pixel size. Therefore, a Block Kriging attempt, documented in Appendix E, was made on the SVC field measurements over grass. The establishment of variograms per spectral band, however, revealed that the data fails to fulfil the statistical prerequisites of stationarity and spatial autocorrelation needed for Block Kriging. Hence, neither linear, exponential, spherical nor Gaussian models represented the spatial dependency well enough to make predictions for unknown points. Consequently, this study reverted to a more primitive upscaling technique: simple averaging within the target support.

As the coordinates used for the MCS measurements in this analysis have been relocated due to proximity to the boat, they now each occupy their own HySpex pixel footprint away from the boat. Therefore, averaging over multiple measurement points is unnecessary. Instead, each HySpex pixel footprint is simply assigned the value of the MCS measurement it contains. While upscaling of the MCS measurements over water is a straightforward process, the upscaling of SVC measurements over grass has to account for a geolocation error of up to 3 HySpex pixels. Although this spatial uncertainty may be equally present near the MCS measurement site, the altering of their original coordinates has made such consideration redundant. The consequence of the geolocation error is an uncertainty of the exact pixel positions. However, when attempting to average all field measurements lying within a particular pixel, such information is essential. The work-around for this issue is a regression-based approach in which measurements are repeatedly averaged for different potential pixel positions, i.e. offsets, until correlation between HySpex values and the upscaling result is maximized. As the maximum geolocation error is estimated to be around 3 HySpex pixels (~2,5 m), correlation was tested for an area stretching ~2,5 m in each direction from the SVC measurement coordinates. Within that area, the evaluation of correlation was limited to every 0.267 m (1/3 HySpex pixel) to keep a balance between result accuracy and computing time. At each spatial offset, the measurements were averaged over the pixel footprint covering them and a weighted mean RMSE across all upscaled pixels calculated ((5)-(7)). The weight per upscaled pixel was provided by the ratio between the number of measurements falling within that pixel footprint and the total number of measurements, here 22. Correlation between HySpex and the upscaled SVC measurements is maximized where the spatial offset produces the smallest weighted mean RMSE.

$$RMSE = \sqrt{\frac{\sum(HySpex_i - Upscaled_i)^2}{n_i}} \qquad (per\ band/upscaled\ pixel/offset) \qquad (5)$$

$$RMSE_{weighted} = RMSE \cdot \frac{n_{M(pixel)}}{n_{M(total)}} \qquad (per\ upscaled\ pixel/offset) \qquad (6)$$

$$RMSE_{weighted\ mean} = \frac{\sum RMSE_{weighted}}{n_{pixel}} \qquad (per\ offset) \qquad (7)$$

where

$i = band\ number$
$n_i = number\ of\ bands$
$n_{M(pixel)} = number\ of\ measurements\ per\ pixel$
$n_{M(total)} = number\ of\ measurements\ in\ total$
$n_{pixel} = number\ of\ pixels$

5.3. Validation of the Sentinel-2 Level-2A Surface Reflectance Product

The validation of the Sentinel-2 L2A SR product consists of two parts. First, the HySpex L2A SR data is validated by comparing upscaled (averaged) field measurements with the respective HySpex pixels. Subsequently, the HySpex data, of which the reliability has been determined, is adopted as the new reference data to validate the Sentinel-2 L2A SR product.

The HySpex data were validated by comparing them to the averaged MCS measurements and the averaged SVC measurements at the best correlating geolocation error offset. To do so, the regression between HySpex and the upscaled field surface reflectance values, their differences per pixel and band as well as the weighted mean RMSE across all pixels were evaluated.

Before comparing Sentinel-2 surface reflectance to the HySpex reference data, the HySpex imagery was spatially resampled to 10 m resolution for bands 2, 3, 4 and 8, to 20 m for bands 5, 6, 7 and 8A and to 60 m resolution for Band 1, so they match the spatial resolution of Sentinel-2 bands. The applied resampling method was simple averaging. Following this, polygons covering pure pixels of different landcover classes (water, forest, cropland, grassland, urban area) were drawn visually with the aid of Sentinel-2 True Color Image (TCI) at 60 m resolution. All HySpex and Sentinel-2 pixels (10, 20 and 60 m) lying within polygons were sampled for the validation of the Sentinel-2 L2A SR product.

To quantify the quality of the Sentinel-2 L2A SR product for the sampled area near Lake Stechlin, the absolute and relative APU measures commonly used in validation literature are adopted (DOXANI et al. 2018; ESA 2019). This facilitates comparison between a validation approach using a radiative transfer model with AERONET-derived AOT values for atmospheric correction and a field campaign-based approach without the use of a radiative transfer model. For each band, mean APU statistics are calculated and vidualized. Furthermore, APU graphs include the mean surface reflectance of the reference data, here HySpex imagery, and the mean specifications (specs) defined as < 0.05 * mean SR(reference) + 0.005 (ESA 2019). In addition to this, relative APU measures are calculated by dividing each absolute mean APU value by the respective mean surface reflectance of the reference (HySpex) data. The resulting relative APU values are plotted in percent for each band. APU histograms (DOXANI et al. 2018) for individual bands illustrate APU values per mean HySpex SR bin. The bin width is to 0.01 and 0.001 for water.

To support the APU graph, a histogram of SR difference between Sentinel-2 (Sen2Cor) and HySpex data including a fitted precision and uncertainty curve was also provided. Lastly, a regression plot illustrates the correlation between the mean surface reflectance per band for the two datasets. Since researchers using Sentinel-2 SR imagery are typically interested in a specific landcover, all five plot types (absolute APU, relative APU, APU histograms per band, histogram and regression) were produced for each landcover class and all of them combined.

6. Results

In the following, the spectrally resampled data going into the study are described and the results of the sampling process and the validation of both HySpex and Sentinel-2 data reported before they are discussed in a later chapter.

6.2. Data Analysis

To get familiar with the data going into the study, the surface reflectance distributions of the spectrally resampled data over water and grass have been compared at the measurement coordinates and for a larger area immediately surrounding the respective measurement sites. They were also evaluated with regard to typical spectral signature of the respective surface using the USGS Spectral Characteristics Viewer (USGS 2019) which is based on the USGS Spectral Library (KOKALY et al. 2017). Figure 13 shows the SR distribution of MCS, SVC, HySpex and Sentinel-2 over water (a) and grass (b). The box hinges represent the first (25% quantile) and third quartile (75% quantile), the line in the box the median and the whiskers the distance to the extrema, excluding outliers. Outliers are defined as any value farther than 1.5 times the inter-quartile range away from the respective hinge.

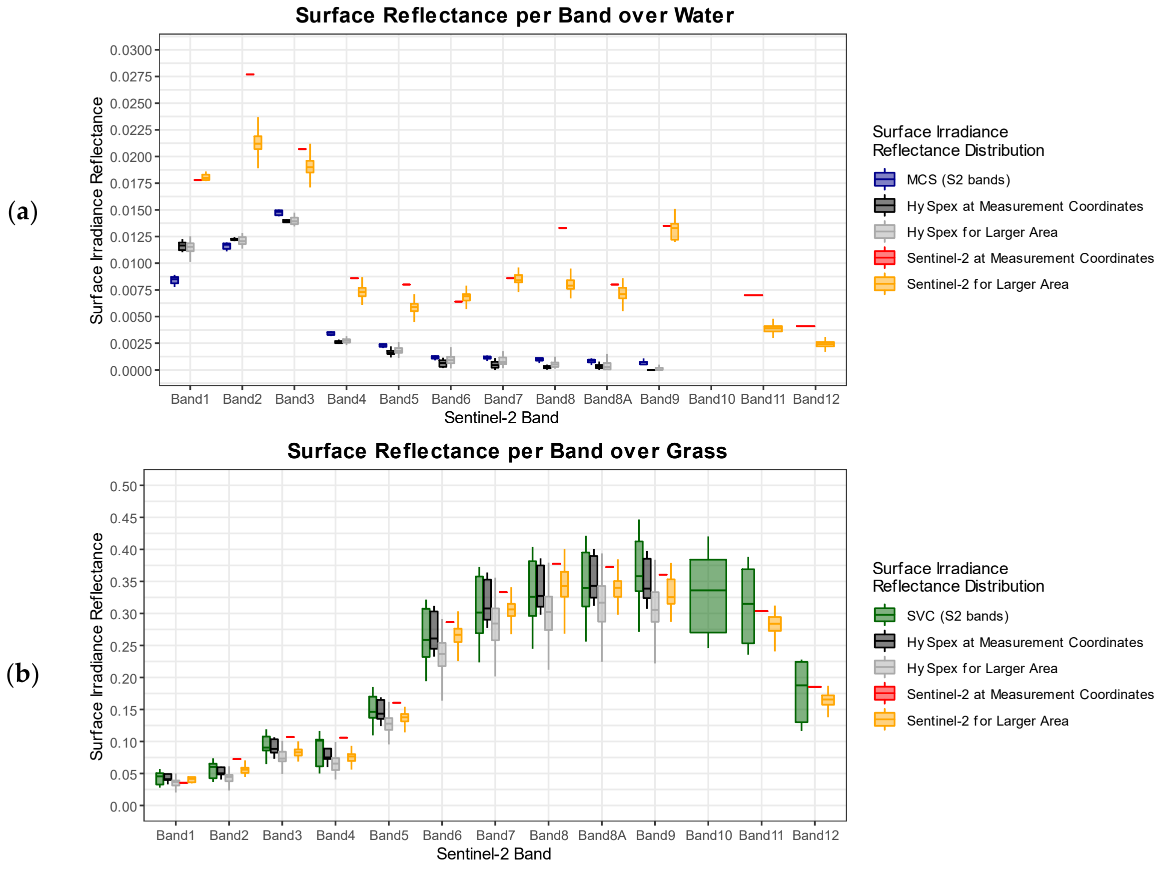


Figure 13. Surface reflectance distributions at measurement coordinates and over the larger area: (a) SR distribution over water; (b) SR distribution over grass.

The MCS surface reflectance throughout the bands closely resembles the spectral signature of clear water (KOKALY et al. 2017; LILLESAND et al. 2015) with values below 0.015 due to strong absorption in the near-infrared (NIR) and transmission in the visible wavelengths (VIS). The peak being positioned in Band 3 (green) rather than Band 2 (blue) may indicate some concentration of chlorophyll in the water of Lake Stechlin. Comparison between HySpex spectra extracted at the measurement coordinates and from a point grid over the measurement sites' immediate sourroundings reveal that the MCS measurement site is representative of the larger environment and does not contain an unusual or atypical reflectance distribution. The equivalent comparison with Sentinel-2 pixel values demonstrate larger differences than those found in the HySpex data. Particularly, in Band 2 (blue), Band 5 (NIR/red edge) and Band 8 (NIR), Sentinel-2 reflectance at the measurement site shows increased values. The Sentinel-2 reflectance also generally follows the typical clear water spectrum (KOKALY et al. 2017). MCS reflectance in Band 1 (aerosol) and 2 (blue) are lower in comparison to the respective HySpex reflectance, which is not the case in other bands. Underestimation of retrieved AOT going into the HySpex atmospheric correction process may have lead to insuffient correction of the effects of aersol scattering, particularly at shorter wavelengths (Band 1). Dark surfaces, such as water, are more stongly affected by aerosol concentration in the atmosphere than lighter surfaces. As their reflectance is dominated by absorption with only very little radiation directly reflected toward the sensor, any reflectance added by aerosol scattering appears as a pronouced reflectance increase in the signal (see also dark object subtraction in Appendix B). Due to the longer atmospheric path length travelled by the signal in order to reach the HySpex sensor compared to the MCS sensor, its imagery is much more susceptible to the influence of aerosols. Mie scattering causes the light touching the aerosol particle to be scattered back towards the sensor before it can reach the water surface and be absorbed resulting in increased values. Remaining effects of aerosol scattering after atmospheric correction, or rather the lack thereof, may therefore be responsible for the lower MCS surface reflectance values in Band 1 and 2. In all other bands, the MCS measurements were very similar to, but slightly higher than the HySpex data. This may be due an undercorrection for water vapor in the atmosphere below the HySpex sensor, absorbing radiation and attenuating the received signal. Sentinel-2 reflectance over water is much larger in all bands than MCS or HySpex reflectance.

The surface reflectance distributions over grass display much higher values than over water as more light is reflected and less light absorbed by the surface (Figure 13 (b)). Reflectance values span from 0.025 to 0.45. The ranges of reflectance values per band also exceed those over water demonstrating a particularly pronouced spatial heterogeneity at different SVC measurement coordinates and the grid points covering the entire meadow. Furthermore, with increased reflectance towards longer wavelengths, the range of represented values also increases. Overall, all measured and extracted reflectance values follow the typical spectral signature of grass (KOKALY et al. 2017; LILLESAND et al. 2015) with a minor peak in Band 3 (green), a steep increase between Band 4 (red) and Band 7 (vegetation red edge) and a decrease in reflectance in the shortwave-infrared, i.e. Band 10, 11 and 12. The rapid reflectance increase at the red edge is noticably less steep than commonly observed, which can be attributed to the dryness of the grass at the acquisition time (KOKALY et al. 2017; LILLESAND et al. 2015). Comparison between HySpex data from the measurement coordinates and the larger area suggest that the SVC measurement sites are on the more reflective end of the meadow's range of reflectances consistently showing higher medians in all bands. The same is true for the Sentinel-2 reflectance, but with a slightly lower reflectance in Band 1. Similar to reflectances over water, the reflectance medians of the Sentinel-2 data are larger than those of the HySpex data. The SVC measurements cover such a large range of reflectances per band that the respective pixel values from HySpex and Sentinel-2 fit well into that range. There are strong similarities between the measured SVC reflectances and HySpex pixel values extracted at the measurement coordinates. A decrease in the interquartile-range going from field to airborne to space-borne sensed reflectance can be attributed to the decrease in spatial resolution from point measurement to 0.8 m to 10, 20 or 60 m resolution respectively. With increasing support, the aggregation effect, previously discussed with regard to the MAUP, reduces contrast between surface reflectance observations by averaging over a larger area.

6.2. Validation of the HySpex SR product using Upscaled Field Measurements

As each MCS measurement coordinate was artificially placed into a different HySpex pixel footprint, no averaging was performed. Nevertheless, by assigning each MCS measurement to the respective pixel footprint, the upscaling process increased the measurement's support justifying the use of the term *upscaled* with respect to the MCS measurements. A map of the upscaled MCS measurements is displayed in Figure 14 (a). The comparison of surface reflectance between the 12 MCS measurements over water and the respective HySpex pixels reveals a mean RMSE across the spectrum (Band 1-8A) and all pixels of about 0.00129. This amounts to a mean NRMSE, the RMSE normalized over the field measurements' reflectance range, of about 9.287% (Table 2). Looking at the correlation of HySpex and upscaled MCS reflectance values in Figure 14 (b), a slight underestimation of reflectance in all HySpex bands, except Band 1 (aerosol) and 2 (blue) can be observed. A mean coefficient of correlation (R) of 0.97686 and a mean coefficient of determination (R^2) of 0.95428 is achieved across all pixels (Table 2). This is even more obvious, when looking at surface reflectance differences (HySpex - upscaled) per band in Figure 14 (c) with absolute differences mostly below 0.0015, but as large as 0.004 for the aerosol band (Band 1). As previously mentioned, this circumstance can be attributed to insufficient correction for aerosol scattering in the HySpex imagery by the ATCOR processor. Surface reflectance differences between pixels in Figure 14 (d) do not display any striking anomalies.

Table 2. Surface reflectance error estimates for HySpex L2A imagery.

	Water	**Grass (excluding outlier pixels)**			
Estimate[1]	**Mean**	**Mean**		**Weighted mean**	
RMSE	0.00129	0.02739	(0.1462)	0.00241	(0.002)
NRMSE	0.09287	0.08126	(0.5136)	0.00754	(0.00703)
R	0.97686	0.99877	(0.99793)	-	(-)
R^2	0.95428	0.99754	(0.99586)	-	(-)

[1] (weighted) mean across Band 1-8A and across all pixels

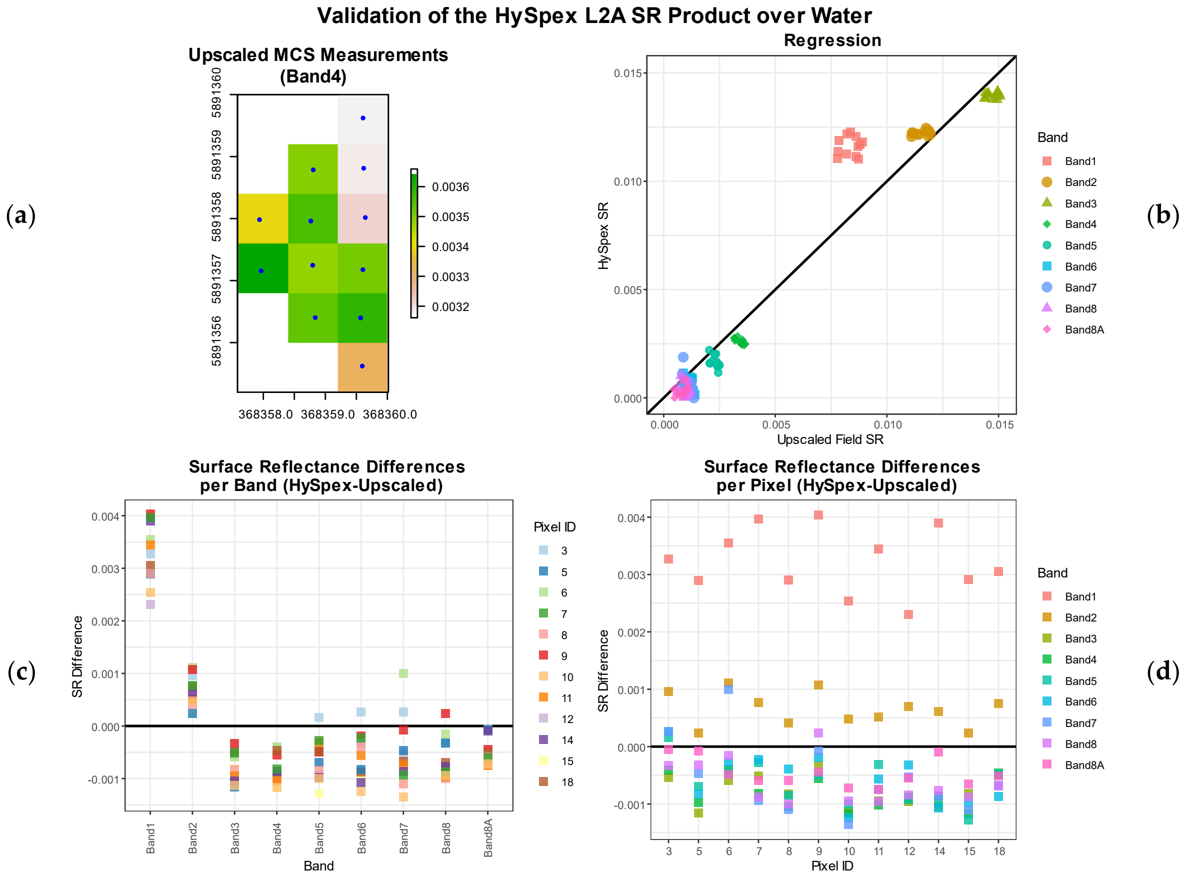


Figure 14. Validation of the HySpex L2A SR product over water: (a) upscaled MCS measurements, (b) regression and (c) SR difference per band and (d) per pixel.

To upscale the SVC measurements, the HySpex geolocation error was taken into account by finding the maximally correlated geographical offset between the upscaling result and the HySpex data at the measurement site. For computational reasons, the SVC measurements were moved rather than the HySpex pixel, which contain the geolocation error. Results are illustrated in Figure 15. The best geographical offset was found at 0 m latitude (S) and 0.533 m longitude (E), which corresponds to a HySpex geolocation error of 0.533 m West with a minimum weighted mean RMSE of about 0.00241 and a weighted NRMSE of 0.754 % across all pixels (Table 2). The real geolocation error may be slightly different in any direction as correlation between upscaling results and HySpex data was only calculated every 0.267 m.

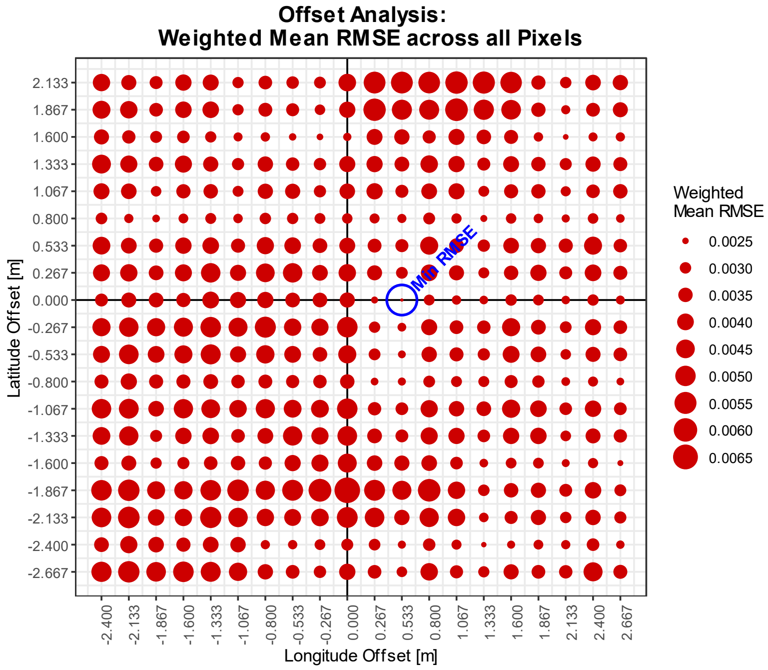


Figure 15. Identification of the geolocation error. Based on the mean RMSE across all pixels weighted by the number of measurements per pixel, the correlation is maximized for a geographical offset of the field measurements of 0 m latitude (S) and 0.533 m longitude (E).

At the best correlated geographical offset, the upscaled SVC measurements and the respective HySpex pixel demonstrate a high correlation with a mean R of 0.99877, a mean R^2 of 0.99754 (Table 2) and most reflectance differences below 0.025. Exceptions to this are three HySpex pixels values, which are up to 0.09 lower than the upscaled SVC reference data. The regression plot in Figure 16 (b) also shows the spectrum divided into lower values in Band 1 to 5 and higher values in bands 6 to 8A. This divide marks the steep reflectance increase in the red edge bands (Band 5 to 8A) typical for vegetated surfaces. The surface reflectance differences plotted per band (c) and pixel (d) in Figure 16 clearly identify pixels 47,58 and 73 as containing the outliers observed in the regression plot. These HySpex pixels display much lower reflectances and, with longer wavelengths, differ increasingly from the equivalent upscaled pixel, which have been averaged over only 1 (58, 73) or 2 (47) measurements. A rerun of the upscaling process excluding the outlier pixels resulted in a slightly different geolocation error in the HySpex data of 0.533 m North and 00.533 m West and a weighted mean RMSE and NRMSE across all pixels of around 0.002 and 0.703 % respectively.

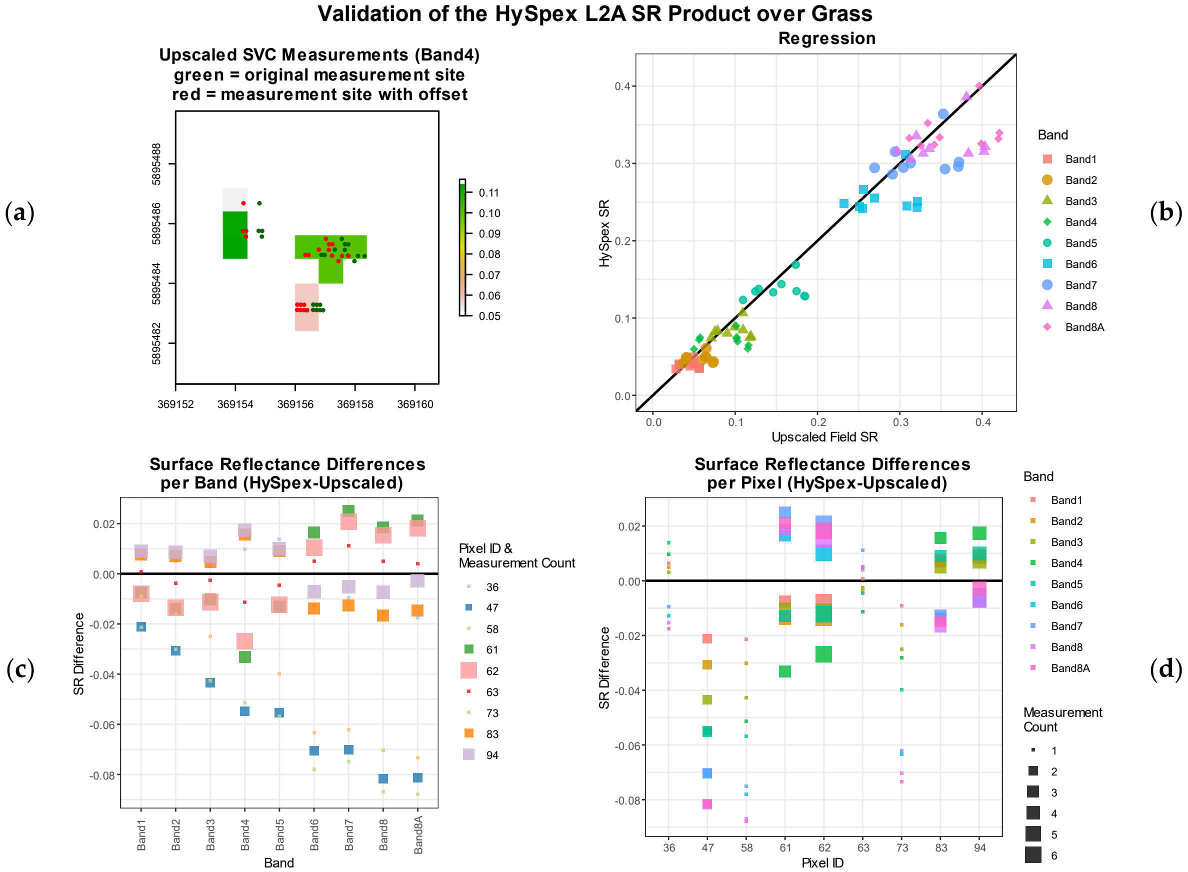
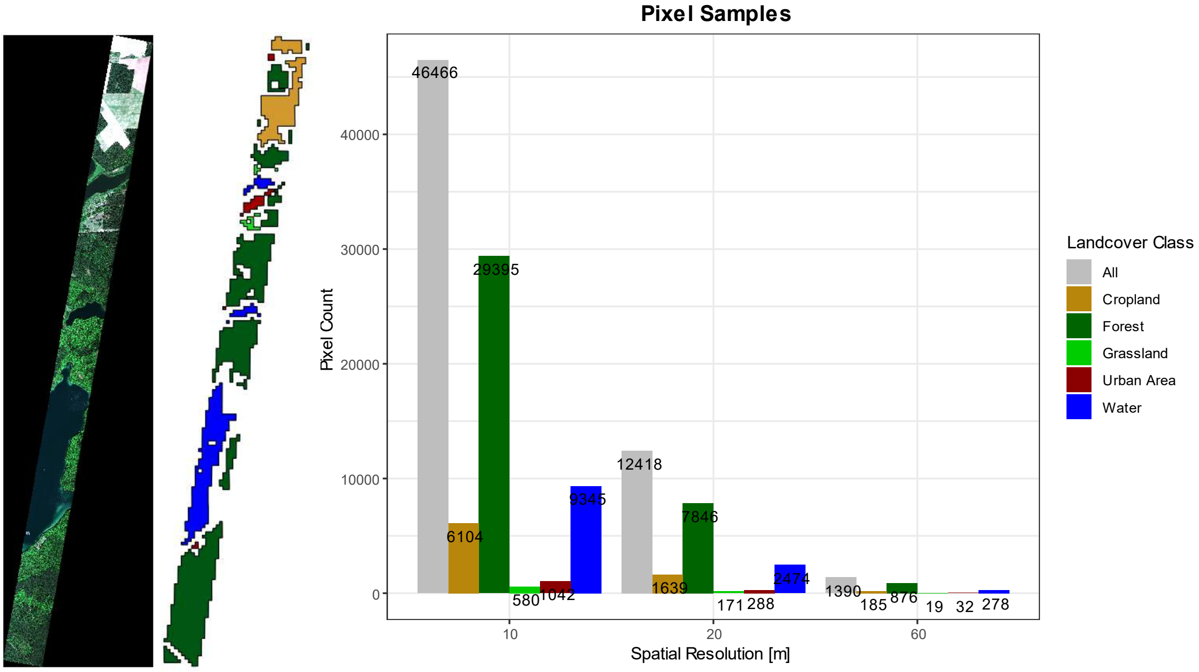


Figure 16. Validation of the HySpex L2A SR Product over Grass: (a) upscaled SVC measurements, (b) regression and (c) SR difference per band and (d) per pixel.

6.3. Pixel Sampling

Pixel samples from the spatially resampled HySpex imagery (Figure 17 (a)), were obtained for five different landcover classes and used as reference data to validate the Sentinel-2 L2A SR product. The polygons from which the surface reflectance values were extracted are illustrated in Figure 17 (b). To avoid ambiguous signals from mixed pixels, they exclusively cover pixels containing only one landcover type: water, forest, cropland, grassland or urban area. Pixels including sun glint or objects in the water were also omitted. Due to differences in spatial resolutions and number of pixels of each landcover type within the study area, the sample size varies greatly between bands and landcover classes (Figure 17 (c)). More than half of all pixels are forest and the other half is almost entirely made up by cropland and water. While these classes evidently possess sufficient representation throughout the study area, grassland and urban area are grossly underrepresented at lower spatial resolutions. Particularly in Band 1 at 60 m resolution, the lack of samples in these two classes, with only 19 and 32 pixels respectively, does not allow for valid conclusions concerning the quality of the Sentinel-2 imagery. In total, 46466 pixels were sampled at 10 m (Band 2, 3, 4 and 8), 12418 at 20 m (Band 5, 6, 7 and 8A) and 1390 at 60 m (Band 1) spatial resolution. The most accurate quality figures will therefore be attained for Band 2, 3, 4 and 8.



(**a**) (**b**) (**c**)

Figure 17. Overview of the sampled pixels per resolution and landcover class: (a) HySpex TCI; (b) validation polygons; (c) sample size per resolution and landcover class.

6.4. Validation of the Sentinel-2 L2A SR Product using Upscaled HySpex Data

Figure 18 shows surface reflectance distributions at the respective measurement site and for the entire respective landcover classes water (a) and grassland (b). At the measurement sites, the upscaled HySpex data (pink) and the Sentinel-2 data (red) only consist of a single pixel, showing as a small horizontal bar. Due to the size of the dataset, the display of outliers has been suppressed.

The surface reflectance distributions for the MCS measurements (blue) in Figure 18 (a) and their upscaled equivalent (turquoise) are identical across all bands since only one measurement point was placed into each pixel when they were artificially relocated. Averaging over a total of one point within a pixel simply allocates that value to the pixel. The large HySpex value ranges over water (grey) match the MCS data (blue) much better than the previous comparison with its immediate surroundings suggested. Even in Band 1 (aerosols), the water reflectance values from HySpex (grey) extend as low as the MCS reflectances (blue). HySpex reflectances at the measurement site (black) remain on the upper end of the value range over water (grey) for bands 1 and 2, in the center for Band 3 and on the lower end of the value range or lower for bands 4 to 8A. Their upscaled equivalent (pink) shows a slightly higher reflectance, most noticeable in Band 1, but nevertheless remains much smaller than the Sentinel-2 reflectance at the measurement site (red), by reflectance differences of around 0.005 (Band 1) to 0.015 (Band 2).

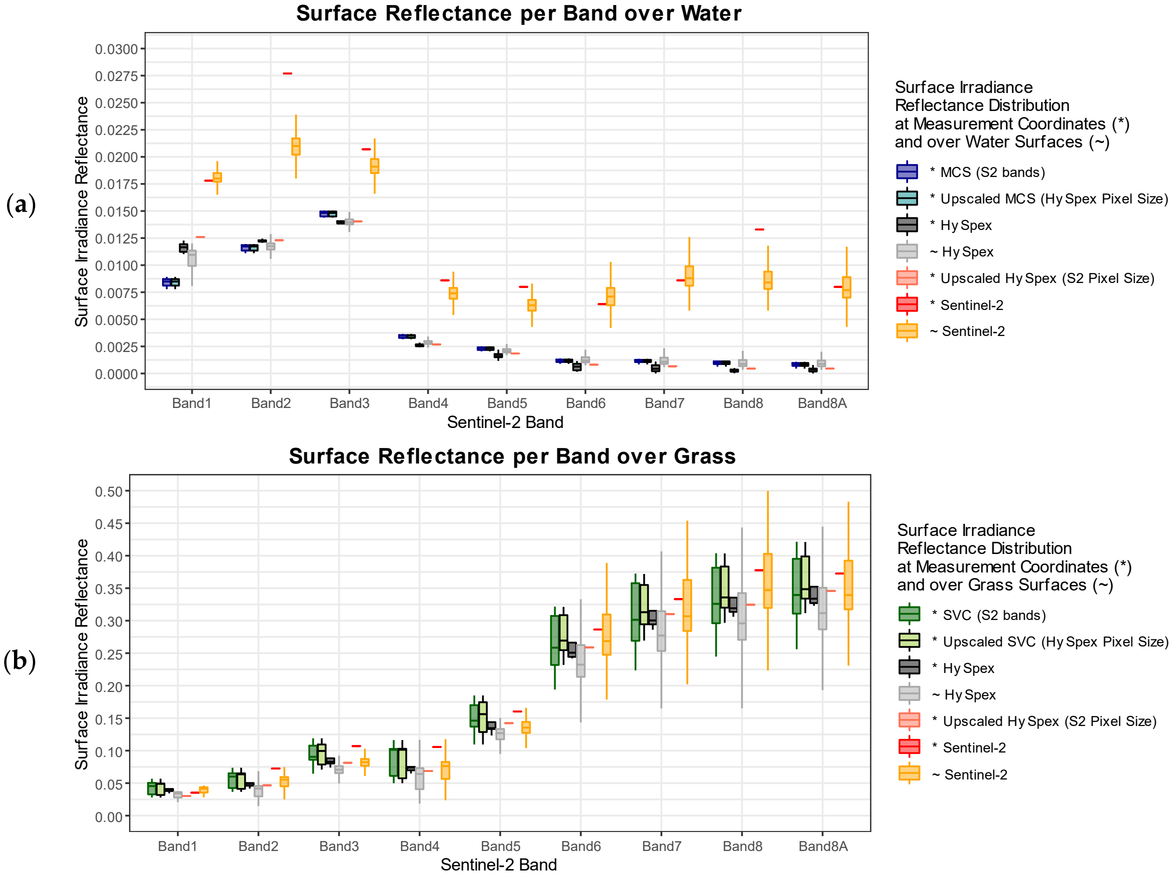


Figure 18. Comparison of surface reflectance distributions of original and upscaled data: (a) over water; (b) over grass (b); both at the measurement coordinates (*) and for all pixels within the respective validation polygons (~).

The measurement coordinates used for comparison of surface reflectance over grass (Figure 18 (b)) have been previously corrected for the geolocation error. Consequently, the extracted HySpex reflectances at the measurement site (black) also change compared to Figure 13 (a). Their ranges (black) are now considerably smaller than at the original coordinates, but still fit well within the SVC reflectance value range (dark green). This suggest a more homogeneous grass surface at the updated coordinates. The upscaling of SVC measurements (light green) has lifted their median values for most bands or left them unchanged (Band 4 and 6). Comparison of HySpex (black/red) and Sentinel-2 (red/yellow) reflectance extracted from the measurement site and over the entire landcover class grassland confirms previous findings, in which the measurement site was found more reflective than most grassland surfaces. The upscaled HySpex pixel value (pink) is slightly lower than the original HySpex data (black) in the visible wavelength domain (bands 1 to 4) and slightly higher in near-infrared (bands 5 to 8A). Yet, the reflectance in the Sentinel-2 pixel at the measurement site (red) is about 0.01 (Band 1) to 0.1 (Band 8) higher than the upscaled HySpex reflectance (black).

The changes in median values when upscaling field or HySpex data can be attributed to the zoning effect (see MAUP); depending on where pixel borders are located, the averaged value distribution differs. This can be particularly problematic in the upscaling of SVC measurements based on an estimated HySpex geolocation error. The correlation between HySpex pixel values and upscaled SVC measurements used for its estimation was only tested every 0.267 meters. This leaves an uncertainty of < 0.267 m in the geolocation error of the HySpex data. Had any other distance been selected to perform the step-wise approximation towards maximized correlation, the upscaling results may differ considerably.

An RMSE of 0.0287, an NRMSE of 7.987 %, an R value of 0.96925 and an R^2 value of 0.93964 has been calculated for Sentinel-2 L2A SR data using all pixel samples. Visualization of Sentinel-2 L2A SR validation results (Table 3) for all landcover classes combined are presented in Figure 19. It includes absolute (a) and relative (c) APU band averages, APU statistics for HySpex SR bins per band (e), as well as SR difference (b) and regression (d) plots. The lowest accuracy value, i.e. the best accuracy, is achieved by Band 4 with a value of 0.00932, closely followed by Band 1 with an accuracy value of 0.00970. The highest accuracy value of 0.04245 was attained by Band 8, which also has by far the worst precision and greatest uncertainty with values of 0.04514 and 0.06196 respectively. The best precision and uncertainty, i.e. lowest values, were reached by Band 1 with 0.00307 and 0.01018 respectively. However, Band 1 does not, in fact, perform the best when looking at the relative APU values. Band 1 attains a relative accuracy value of ~ 53.8 %, which is the second worst after Band 2 with a value of ~ 57 %. The lowest relative accuracy, precision and uncertainty values occurred in Band 8A with ~15.5 %, ~ 14.9 % and ~ 21.5 % respectively. The mismatch in the ranking of bands according to their absolute and relative APU estimates can be explained by strong differences in mean surface reflectance across the spectrum, as seen in the surface reflectance curve in Figure 19. The mean HySpex surface reflectance resembles the attenuated spectral signature of vegetation. This is not surprising considering the dominance of forest throughout the study area. Following the relatively abrupt increase in reflectance between the visible domain and near-infrared, absolute APU values are relatively small with values below 0.02 for bands 1 to 5 and are much larger for bands 6 to 8A with values up to 0.06. Conversely, the relative APU values tend to decrease towards longer wavelengths. Specifically, accuracy and uncertainty values drop from around 60 % down to around 15 % and 20 % respectively, while the relative precision values start out considerably lower at 17 % in Band 1 and 26 % in Band 2 and then generally follow the accuracy curve to around 15 % in Band 8A. All graphs in Figure 19 demonstrate an overestimation of the surface reflectance in the Sentinel-2 data; both mean APU graphs (a and c) have positive accuracy values, the histogram of surface reflectance (b) is skewed towards positive values and the points in the regression plot (d) are above the regression line. With exception of the precision for Band 1 and 2, all APU band values are outside the specifications. While low reflectance bands (Band 1 to 5) miss the specification margin only by around 0.01 or less, high reflectance bands (Band 6 to 8A) are off by up to 0.04. Between accuracy, precision and uncertainty, the uncertainty values are the largest, followed by the accuracy values in both absolute and relative terms. On average, precision values are lowest indicating repeatability as the best validation statistic. A good repeatability represented by low precision values mean that Sen2Cor reliably produces very similar differences in surface reflectance to the reference data for different pixels regardless of how large these differences are, i.e. how large the accuracy values are.

Table 3. Absolute and relative APU values for the validation of Sentinel-2 imagery.

Band	Spatial resolution [m]	Sample size	A	P	U	Rel. A [%]	Rel. P [%]	Rel. U [%]
Band1	60	1390	0.0097	0.00307	0.01018	53.815	17.032	56.478
Band2	10	46466	0.01264	0.00577	0.01389	56.951	25.997	62.583
Band3	10	46466	0.01092	0.0103	0.01501	25.242	23.809	34.696
Band4	10	46466	0.00932	0.00896	0.01293	29.665	28.518	41.155
Band5	20	12418	0.01201	0.01388	0.01835	16.334	18.878	24.957
Band6	20	12418	0.0332	0.02622	0.04231	22.074	17.433	28.131
Band7	20	12418	0.03045	0.02782	0.04124	16.762	15.314	22.701
Band8	10	46466	0.04245	0.04514	0.06196	22.081	23.480	32.229
Band8A	20	12418	0.03105	0.02988	0.04309	15.467	14.884	21.464

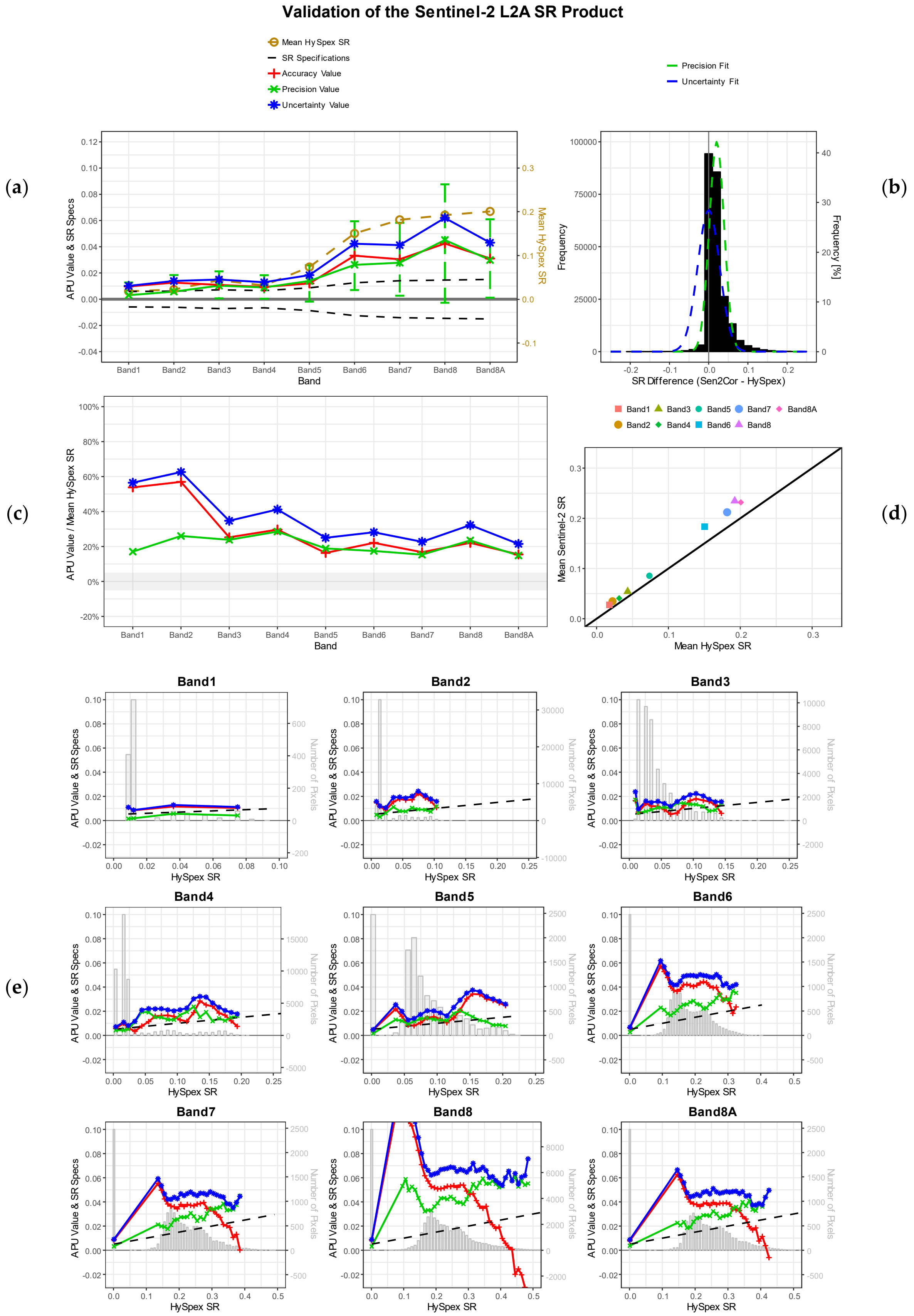


Figure 19. Visualization of the Sentinel-2 L2A SR validation using all sample pixels: (a) Absolut mean APU plot; (b) SR difference histogram; (c) relative mean APU plot; (d) mean SR regression plot; (e) APU histograms per band.

The APU histograms in Figure 19 (e) illustrate the absolute APU values, specifications and number of pixels per surface reflectance bin in the upscaled HySpex data. The respective graphs for each landcover class can be found in Appendix F. The bin width is set at 0.01 and 0.001 for water pixels. The displayed surface reflectance is between 0 and 0.1 for Band 1, between 0 and 0.25 for the low reflectance bands (bands 2 to 5), and between 0 and 0.5 for high reflectance bands (bands 6 to 8A). Due to the low surface reflectance over water, a SR domain between 0 and 0.025 is displayed for all bands of the landcover class water. APU values represented by less than 50 pixels are removed from the plot, unless all bins of the respective band plot are backed by less than 50 pixels.

In Band 1 (aerosol), 2 (blue) and 4 (red), surface reflectance of the HySpex sample pixels is limited to a very narrow reflectance domain with virtually all values below 0.025. The respective Sentinel-2 pixels exhibit low APU values. While accuracy and uncertainty are outside of the specifications for the well populated reflectance bins, precision stays within them. APU values tend to rapidly increase or decrease where surface reflectance bins are not backed by enough pixels. The lack of samples pixel in the respective bin does not allow for a clear APU statistic to be calculated, therefore those bins' APU values are unreliable and are disregarded in the analysis.

The HySpex SR bins in Band 3 (green) and 5 (red edge) are distributed in a left skewed bell shape. On top of that, Band 3 displays a separate high-frequency bin close to zero, which represents the water pixel component. In both bands, accuracy, precision and uncertainty are mainly outside of specifications, merely accuracy either touches the specification margin or drops below it at a reflectance of around 0.07. All three statistics are close together with values between 0.008 and 0.02. Between them, uncertainty, once more, performs the worst.

The HySpex SR distributions in bands 6 to 8A follow a left skewed bell shape distribution with a major peak at reflectance values between 0.015 and 0.02 and another minor one at around 0.025. Each of bands also exhibits a single high-frequency spike near zero reflectance due to the low reflectances in the water pixels. In those four bands, the Sentinel-2 APU values do not stay within their specifications, with exception of the precision value in the water SR bin. Uncertainty deviates from the specifications the most, followed by accuracy and precision. In the precision curve, dips at the peaks, i.e. the surface reflectance domain backed by the highest number of pixels, can be observed. Ignoring the near-zero bin with very small APU values, values in bands 6, 7 and 8A range from around 0.018 to around 0.05, whereas in Band 8 they extend much further to higher values of between 0.03 to 0.07.

Sentinel-2 validation figures for different landcover classes are attached the Appendix F (Figure 24- 30) and are briefly described here. All landcover types have in common, that surface reflectance values across all bands are overestimated in the Sentinel-2 data as can be seen in their respective regression plots (Figure 25 (b)). The overestimation tends to increase with larger surface reflectance values. Therefore, regression between mean HySpex and mean Sentinel-2 SR for water reveals a distinct value distribution in two cluster, due to relatively large SR values in bands 1, 2 and 3 as opposed to very low reflectance in all other bands. In terms of absolute APU values, all landcover types reveal accuracies and uncertainties mostly outside and precisions sometimes inside the specifications. Values in Band 3 (green), 4 (red) and 5 (red edge) tend to perform better and bands 6 to 8A (red edge/NIR) worse. The former bands exhibit a value range between 0.006 and 0.03 and the latter between 0.04 and 0.085 for all landcover classes, except water. Water APU values are limited to values between 0.001 and 0.005. The accuracy only falls within specifications in bands 3, 4 and 5 for water and urban area and almost falls within them for forest and grassland. Precision proves to be the best performing APU statistic. It remains within specifications across all bands for water and cropland pixels and skirts them for grassland pixels. The uncertainty closely follows the accuracy for all landcover classes, except urban area, where uncertainty is much larger. Examining relative APU values, water demonstrates particularly large values, starting at ~ 40 % accuracy and uncertainty and ~ 7% precision in the shorter wavelengths and then quickly increasing towards ~800 % accuracy, ~880 % uncertainty and ~380 % precision value in the longer wavelengths. The lowest relative APU statistics across all bands are attained by the cropland sample with maximum values below 30 % and precision values below 10 %. It is followed by grassland, urban area and forest with maximum values at ~ 38 %, ~57 % and ~92 % respectively.

To sum up, results of the Sentinel-2 surface reflectance validation show an overestimation of surface reflectance across all bands and landcovers. APU statistics are predominantly outside of the specifications, although precision does stay within desired parameter for water and cropland. The statistic with the largest values is uncertainty, followed by accuracy and precision. In high-reflectance bands, relative APU statistics perform better than in low-reflectance bands. Due to overall low surface reflectance values for water, this landcover class has extremely high relative APU values, about 10 times higher than other landcover classes. However, caution is required when interpreting the results. The tiny sample sizes of grassland and urban pixels at 60 m spatial resolution (Band 1) renders them useless for APU statistics. Furthermore, samples for cropland and water in Band 1, and grassland and urban area at 20 m spatial resolution are also not sufficiently large to produce reliable results. Most reliable are results for bands 2, 3, 4 and 8 at 10 spatial resolution in all landcover classes combined, followed by forest, water and cropland.

7. Discussion

In the following, the methods of field sampling and upscaling of field measurements are critically discussed and suggestions for improvements made. Furthermore, validation results are put into context by comparing them to results of similar studies and potential error sources explored.

7.1. Field Sampling and Upscaling Methods

Overall, the methodology of upscaling field measurements for the validation of the Sentinel-2 L2A SR product as demonstrated in this study for an area near Lake Stechlin proved to be effective. However, methods for field sampling, upscaling and identification of the geolocation error leave room for further refinement.

The measurements taken in the field campaign near Lake Stechlin do not adhere to recommendations by MUIR et al. (2011), JOHANNSEN & DAUGHTRY (2009) and SCARTH (2006) in several ways. The measurement sites are, as recommended (MUIR et al. 2011), representative and homogenous surfaces with minimal topographic variation. The measurement timing is in close synchrony with the acquisition of the remotely sensed imagery and the spatial extent of the field measurements from the Lake Stechlin campaign is sufficient, covering an area larger than 3 times 3 HySpex pixels (MUIR et al. 2011). Nevertheless, the number of measurement sites per surface does not meet the recommended five sites (MUIR et al. 2011) or more for heterogeneous surfaces (JOHANNSEN & DAUGHTRY 2009) with only a single site per surface. Moreover, sample sizes of 12 and 22 measurements over water and grass respectively are far too small considering SCARTH (2006) recommends a minimum of 300 for natural vegetation. The SVC measurements over grass also lack a clear sampling scheme. Adapting the approach by MUIR et al. (2011) to HySpex target resolution with sample size considerations (SCARTH 2006) in mind, three 3 m transects should be laid in a star-shape at 60° to one another (Figure 4) and measurements taken every 3 cm along each transect. In practice, a spacing between observations of 3 cm may not be attainable depending on the smallest possible spatial support of a measurement taken by the field spectrometer. Measurement uncertainties from human error or slightly uneven terrain may also be a problem at such small spacings. Nonetheless, a larger sample size than 22 grass samples should be achieved.

Although, the literature advocates Block Kriging for point to pixel upscaling over heterogeneous surfaces (ATKINSON & TATE 2000; CROW et al. 2012; WANG et al. 2015; ZHANG et al. 2018), this upscaling method did not work for the SVC measurements over grass as the statistical prerequisites of stationarity and spatial autocorrelation were not fulfilled. However, it does not necessarily mean an absence of stationary and spatial autocorrelation over grassland in general as spatial heterogeneity varies with scale (ATKINSON & TATE 2000; QUATTROCHI & LAM 1991). ATKINSON & TATE (2000) therefore recommend a thorough investigation of the underlying spatial variation in the data before rescaling them. In this study, the only data available to investigate the spatial variation are the SVC measurements intended to be upscaled. When preparing a semivariogram to investigate spatial dependencies of observation pairs, the spacing, extent and support used to acquire the observations function like a filter on the resulting empirical semivariogram (BLÖSCHL & SIVAPALAN 1995). With more close-knit spacing and more even coverage under the sampling scheme recommended by MUIR et al. (2011), the substantially larger number of observations may exhibit different spatial patterns in the empirical semivariograms. Consequently, theoretical semivariograms (e.g. linear, exponential, spherical or Gaussian models) may better correlate with the observations, hence better represent underlying spatial dependencies. This, in turn, could lead to the fulfillment of the statistical prerequisites and enable Block Kriging using the theoretical semivariograms to make predictions about unknown points which are then averaged over the pixel footprint. Using the more sophisticated upscaling technique of Block Kriging rather than simple averaging may produce a closer approximation of the target pixel value.

Another point to be questioned is whether this study really demands the upscaling from point to pixel values. Given that the SVC spectrometer measures reflectance within a certain field of view (SVC 2019) equating to a specific area on the ground, one can argue that with high-resolution imagery, such as the HySpex data, upscaling from plot (area) to pixel is required. The SVC measurements can only be interpreted as point values if there is a large difference between the area supporting the measurement value and that supporting the pixel value. For example, assuming that the standard field of view of 4° (SVC 2019) was used to conduct the SVC measurements from 1 m above the ground, the measurement's spatial support would amount to a circular area of around 0.07 m in diameter. Comparing the 0.07 m spatial support of the SVC measurement to the 0.8 m spatial resolution of the HySpex data, a plot to pixel upscaling method taking the proportional area of measurement to pixel support into account may have been more suitable, particularly if the spacing between measurements is small enough for multiple measurements to fit in one pixel footprint.

So far, the support of a HySpex pixel value has been equated with the spatial resolution of the HySpex data. This is, in fact, not precisely correct. As ATKINSON & TATE (2000) point out the support is characterized by the Point Spread Function (PSF) of the sensor, in this case of the HySpex sensor. The PSF describes the relative contribution from different parts of the IFOV's projection on the ground to the signal recorded at the sensor. As the ground IFOV (GIFOV) is larger than the pixel with the assigned reflectance value, contributions from neighboring pixels are included into the pixel value according to the PSF (JONES & VAUGHAN 2010). For the upscaling in this study, this would have meant that the measurements must be upscaled to the GIFOV with a weighting scheme according to the PSF instead of using the HySpex spatial resolution as the target support. This fine difference in target supports was ignored, as the HySpex data contained a geolocation error (0.267 m N, 0.533 m W) that eclipses any effect signal contributions from neighboring pixels could have had. Moreover, the pronounced spatial heterogeneity of grassland surfaces would also obscure reflectance contributions from outside the HySpex pixel footprint. Hence, for simplification the upscaling target was set to the spatial resolution of the HySpex imagery.

Furthermore, the method used for identification of the geolocation error in the HySpex data may produce erroneous results. Although the identified geolocation error in this study was plausible, finding the maximized correlation between upscaled field measurements and the respective pixels could point to incorrect results depending on the underlying spatial patterns of reflectance or uncertainty in field or remote sensing data. For example, using the correlation method on a repetitive reflectance pattern, as might occur on managed land, may lead to many geographical offsets with maximized correlation. Correlation can also be high by accident due to zoning effects, uncertainty in the acquired field data or in the upscaling method and therefore produce an ambiguous or erroneous geolocation error. On top of that, the offset steps at which the upscaling of the SVC measurements and their correlation with the HySpex pixel values was evaluated in this study also limit the accuracy of the identified geolocation error. In reality, the geolocation error could be slightly off in any direction. Hence, the uncertainty of the geolocation error is < 0.267 m (Table 4) corresponding to the offset step range (Figure 15). An alternative method to the correlation approach, is the manual marking of pixels by placing dark objects onto the ground around the measurements. However, a couple of pixels distance should be kept between dark objects and measurements to avoid contamination in the reflectance signal of the remotely sensed imagery.

7.2. Validation Results

After discussing shortcomings of the methods applied in this study, this chapter is going to evaluate the uncertainty figures from the HySpex and Sentinel-2 validations and put them into context by comparison with similar studies and evaluate potential error sources. An overview of uncertainty estimates determined in this study are given in Table 4.

Table 4. Overview of uncertainty estimates.

Surface	Data	Estimate	Entity	Value
Water	HySpex imagery	RMSE	SR	0.00129
		R^2	SR	0.95428
Grass	Raw SVC measurements (sensor changes)	Mean Spectral Uncertainty	Uncalibrated SR	0.08949
	SVC measurements	Calibration Uncertainty	SR (all bands)	0.00451
	HySpex imagery	Geolocation error	Meters	0.533 m W
		RMSE	SR	0.02739
		R^2	SR	0.99754
	HySpex geolocation error	Uncertainty	Meters	< 0.267
All	Sentinel-2 L2A SR imagery	RMSE	SR	0.02877
		R^2	SR	0.93964

The validation of the HySpex SR imagery using the upscaled field measurements revealed a mean RMSE of 0.00129 over water and 0.02739 over grass and respective mean NRMSE values of 9.287 % and 0.813 % (Table 2). The weighted equivalents over grass are 0.00241 and 0.754 %. A similar study by PFLUG (2019) also conducted near Lake Stechlin but in November 2018 found a mean RMSE of 0.0009 over water and of 0.022 over grass when comparing HySpex data with the reference MCS and SVC measurements averaged over the respective pixel footprint. In comparison, this study's mean RMSEs are marginally larger over water (~ + 0.0004) and over grass (~ + 0.005). The mean RMSE over grass weighted by the number of measurements included into the upscaled pixel value is much smaller (~ - 0.02) than the unweighted one in PFLUG (2019). The substantial difference (~ factor 10) in the grass figures may originate from the inclusion of wavelengths equivalent to Sentinel-2 bands 9, 11 and 12 into the mean RMSE figure over grass in PFLUG's (2019) study. For the RMSE over water, the same wavelengths have been used as in this study (Sentinel-2 bands 1-8A). Assessing the spectral curves (Figure 18), the two studies have matching results. Over water, the HySpex median reflectance is much larger in Sentinel-2 Band 1 due to insufficient correction of aerosol scattering and the same or lower than that of the averaged MCS measurements for bands 2 to 7. A slight deviation toward lower reflectance values can be observed for bands 8 and 8A when compared to PFLUG (2019). Over grass, both studies show smaller median reflectance values for HySpex than the averaged SVC measurements across all relevant bands (1-8A) with differences decreasing towards longer wavelengths. Altogether, both studies show a high correlation between upscaled (averaged) field measurements and the respective HySpex pixels with mean R values of 0.97686 over water and 0.99877 over grass (Table 2), compared to 0.984 and 0.998 respectively in PFLUG (2019).

The results from the validation of the Sentinel-2 L2A SR product are much less correlated with their HySpex reference data. Insufficient samples sizes from an uneven distribution of landcover portions in the study area disqualified APU statistics for Band 1 (60m resolution) over grassland and urban area (< 50 pixels) and render results for Band 1 (60 m resolution) over cropland and water and for bands 5, 6, 7 and 8A (20 m resolution) over grassland and urban area unreliable (< 300 pixels) (Figure 17). This leaves results from bands 2, 3, 4 and 8 at 10 m spatial resolution, from the landcover class forest and from all landcovers combined as the most reliable figures displaying the smoothest APU curves in the APU histograms (Figure 19 (e)).

Early validation of the Sentinel-2 L2A product by PFLUG et al. (2016) uses reference data from correcting the Sentinel-2 L1C product with a constant AERONET-derived AOT value in Sen2Cor. They quantified the maximum SR difference to the reference data at 0.04, which is very close to mean difference found in this study for Band 8 (Figure 19 (d)).

In comparison with results by PFLUG (2019) with an R value of 0.974 over water and 0.992 over grass for Sen2Cor performance validation using HySpex data, this study found a lower value of 0.96925 over all surfaces. Furthermore the Sentinel-2 imagery had an RMSE of 0.0287, where PFLUG (2019) achieved values of 0.005 and 0.026 over water and grass, respectively. Therefore, the Sentinel-2 L2A SR product in this study is less accurate than the one evaluated by PFLUG (2019).

The validation results published in the official Sentinel-2 L2A Data Quality Report (ESA 2019) are processed with Sen2Cor 2.5 and are based on datasets from the DOXANI et al. (2018) which include 19 AERONET sites with a variety of different landcovers from temperate, arid, equatorial forest, boreal and coastal regions. The average surface reflectance from the DOXANI et al. (2018) reference dataset in ESA (2019) climbs steadily from around 0.1 at Band 1 to 0.2 at Band 5 and reaches its maximum (bands 1-8A) of around 0.3 at Band 8A. This study appears to contain more dark targets as the average surface reflectance is much smaller starting at around 0.03 at Band 1 and staying below 0.05 until Band 4, increasing to around 0.07 at Band 5 and reaching its maximum of ca. 0.2 at Band 8A (Figure 19 (a)). The larger SR values in ESA (2019) are most likely due to inclusion of high-reflectance landcover classes such as bare soil, snow and desert. The APU curves displayed in the Data Quality Report (ESA 2019) show low accuracy values, which lie outside the specification only for Band 5. The precision and uncertainty curve run parallel to, but entirely outside the specification curve with values between ~ 0.017 and ~ 0.036 (Band 5). In this study, all three APU curves lie outside the specifications with values below 0.02 for low SR bands (1-5) and as large as 0.62 for high SR bands (6-8A), with exception of the precision values for bands 1 and 2, which scarcely remain inside the specifications. The APU values also increase non-linearly with the increase in SR, which is not the case in ESA (2019). Relative APU values in the Data Quality Report are between ~ -5 % and 22 % with accuracy mostly staying between the ± 5% boundaries, whereas relative APU values in this study are much larger with values between 17 % and 63 %. While in this study, precision performed the best, followed by accuracy and uncertainty, in ESA (2019) accuracy performed best, followed by precision and uncertainty. Overall, absolute and relative APU statistics in ESA (2019) indicate a much better Sentinel-2 L2A SR product quality than this field campaign-based study does.

Furthermore, DOXANI et al. (2018) conducted their own study using Sen2Cor 2.2 and the described reference dataset to compare different atmospheric correction processors. Their APU histogram presented for Sen2Cor Band 4 (red) compared to that of this study (Figure 19 (e)) further highlights that a larger number of dark surfaces was included in this study with most reflectance values between 0 and 0.025 as opposed between 0 and 0.25 in DOXANI et al. (2018). The total number of pixels included into the histogram also differs immensely with 30 million compared to 50000 in the Lake Stechlin dataset. Although both studies divide the SR frequency histogram into SR bins of 0.01 width, this great disparity in sample sizes lead to APU values of the most populated SR bins to be limited to 3 data points in this study as opposed to 25 in the APU histogram by DOXANI et al. (2018). In the 3 data points, precision values stay within specifications, while accuracy and uncertainty are almost within them for two SR bins, but outside for the third. Beyond the most populated SR bins, APU values are outside of the specifications. In DOXANI et al. (2018), all SR bins in

Band 4 show APU values within the specifications. Only past SR bins of ~ 0.3, the values start going over, but stay close to the specification line.

All in all, the validation of the Sentinel-2 L2A SR product using HySpex imagery from near Lake Stechlin as reference revealed a much worse performance of the Sen2Cor atmospheric correction processor than similar studies using AERONET corrected SR products as reference indicated (DOXANI et al. 2018; ESA 2019), but similar maximum SR reflectance differences as in early validation efforts (PFLUG et al. 2016). With few exceptions the APU statistics were outside the specifications. Precision, i.e. the variation around the mean difference to the reference data or repeatability, appeared to be the best performing APU statistic. Accuracy, the mean difference to the reference data, and uncertainty, the quadratic sum of accuracy and precision, showed poor results. An increase in sample size would produce smoother APU curves and more reliable results. Particularly for the subdivision into landcover-specific validation results, the number of samples pixels should exceed at least a couple of hundred.

The large differences in the Sentinel-2 L2A SR product and the HySpex reference data illustrated by discrepancies between APU values and their specifications can be attributed to a combination of different factors. The geo-referencing of the reference data should be that of the data to be validated. Moreover, surface reflectance is dependent on the reflectance properties (BRDF) under prevailing atmospheric conditions and the sun-sensor geometry at the time of acquisition. It is also influenced by the processor used for atmospheric correction as well as its inputs (DOXANI et al. 2018). Comparing these factors between Sentinel-2 and HySpex data used in this study may shed some light on potential error sources in the Sentinel-2 L2A SR product. For instance, georeferencing between HySpex and Sentinel-2 using Ground Control Points (GCPs) was not performed. Consequently, detected SR differences between HySpex and Sentinel-2 pixels could turn out to be simply SR differences between overlapping pixels corresponding to different parts of the earth's surface. Moreover, imagery from the airborne HySpex sensor at ~ 1 km altitude was acquired about 20 minutes earlier than the Sentinel-2 data at 786 km altitude and therefore has slightly different solar zenith and azimuth angles. Viewing angles also differ, which means HySpex and the MSI detect the target's signal at different angles. The expected reflectance divergence per wavelength arising from these differences in sun and viewing angles are described by the BRDF of the respective target surface. The two sensors also differ in their IFOVs leading to a ground sampling distance, i.e. spatial resolution, of 0.8 m in HySpex and 10 m, 20 m or 60 m in Sentinel-2 imagery. Although the HySpex data had been upscaled to the Sentinel-2 spatial resolution before being compared, the resulting surface reflectance will not be the same, as the real target spatial support is described by the PSF of the MSI. Another potential error source is the application of two different atmospheric correction processors. While HySpex imagery was processed using ATCOR 4 with a retrieved average AOT at 550 nm of 0.119, the Sentinel-2 L1C product was processed using Sen2Cor with a retrieved AOT at 550 nm of 0.059 in the study area. Due to the much longer atmospheric path length in signals reaching the Sentinel-2 satellite than that of signals reaching the airborne HySpex sensor, one would expect that the retrieved AOT value at 550 nm is much larger in the Sentinel-2 imagery. However, this is not the case. Instead, it is only about half. The disparity between the retrieved AOT values may be explained by a smaller proportion of DDV pixels in the Sentinel-2 scene than in the HySpex scene, impeding accurate AOT retrieval (KAUFMAN & SENDRA 2007; GASCON et al. 2017). Inaccurate AOT retrieval may have then, in turn, caused Sen2Cor to underestimate the aerosol concentration in the atmosphere and therefore overestimate surface reflectance in the Sentinel-2 L2A SR product.

8. Conclusion

Validation of the Sentinel-2 L2A SR product is essential to avoid misattributing data uncertainties to the outcome of a study based on this remote sensing imagery. Therefore, a two-part validation of the Sentinel-2 L2A SR product was conducted using field measurements and high-resolution HySpex data from near Lake Stechlin. After the quality of the HySpex imagery has been determined using the upscaled field measurements, pixel samples from the upscaled HySpex data served as reference for the validation of the Sentinel-2 L2A SR product. The upscaling was accomplished by averaging observations within the respective target pixel footprint. Finally, uncertainties were reported, the applied methods evaluated, validation results compared to those of similar studies and potential error sources identified.

The overall approach of this study has proven to be effective but could benefit from improvements regarding the sampling of field data, the upscaling strategy and the identification of the geolocation error. Field measurements should be collected using a sampling scheme that provides suitable sample sizes, support, spacing and extent to capture underlying spatial variation and enable the upscaling to a representative pixel value. A recommended sampling scheme specifically developed for upscaling is that provided by MUIR et al. (2011). Using a more advances field sampling scheme may also uncover spatial dependencies between observations and enable the more sophisticated upscaling technique of Block Kriging, rather than using the more primitive simple averaging technique. Furthermore, the identification of the geolocation error in the HySpex data by finding the best correlation with the upscaled field data may produce ambiguous or incorrect results in case of repetitive reflectance patterns, zoning effects or considerable uncertainty in the data. A better approach would be the placement of dark objects in the vicinity of the measurement site, so they can serve as Ground Control Points in the imagery of airborne or spaceborne sensors.

The validation of the Sentinel-2 L2A SR product for an area near Lake Stechlin using airborne HySpex imagery as reference revealed poor accuracy, precision and uncertainty estimates, hence point toward large uncertainties in the performance of Sen2Cor. Validation results showed APU statistics mainly outside specifications, as well as a relatively high RMSE (0.02877) and low R estimate (0.93964) and therefore fell short of those in comparable studies (DOXANI et al. 2018; ESA 2019; PFLUG 2019). Overall, an overestimation of surface reflectance is observed in the Sentinel-2 imagery. This is most likely caused by underestimating aerosol concentration in the atmosphere leading to the insufficient correction of elevated radiance from aerosol backscatter. Since the HySpex reference imagery has been validated using upscaled field measurements and found to correlate highly with the ground reference, the surface reflectance deviances in the Sentinel-2 imagery cannot be attributed to uncertainties in its reference data. Potential error sources, however, are differences in georeferencing, solar and viewing angles, IFOVs and processors as well as the retrieved AOT value at 550 nm as processing input. Moreover, the pixel sample size used for the validation is much too small to obtain valid validation statistics for some individual landcover classes and is also much smaller than in equivalent studies (DOXANI et al. 2018; ESA 2019) using a dataset from DOXANI et al. (2018).

This study highlights the complex issues arising when dealing with data of different spatial scales and different modes of acquisition. Upscaling is an important tool to overcome disparities between spatial supports of observations inherent in such data. Appropriate sampling schemes can improve the representativeness of an observation upscaled to a target pixel and make subsequent validation or other applications more reliable. The validation of HySpex and Sentinel-2 imagery near Lake Stechlin can be further expanded to the validation of the Landsat 8 L2A SR product, as it was also acquired during the field campaign on May 4th, 2018.

Acknowledgments: I would like to thank my supervisors Bringfried Pflug (DLR), Dr. Magdalena Main-Knorn (DLR), Dr. David Frantz (HU Berlin) and Prof. Dr. Patrick Hostert (HU Berlin) for the kind support through the process of preparing this thesis. A big thank also to the team at the DLR in Berlin-Adlershof for taking me in and to the staff of the Geomatics Lab who patiently answered all my questions.

References

ACHARYA, R. (2017): Interaction of waves with medium. In: ACHARYA, R. (ed.): Satellite signal propagation, impairments and mitigation. London. p. 57–86.

ANDERSON, M. C., NEALE, C. M. U., LI, F., NORMAN, J. M., KUSTAS, W. P., JAYANTHI, H. & CHAVEZ, J. (2004): Upscaling ground observations of vegetation water content, canopy height, and leaf area index during SMEX02 using aircraft and Landsat imagery. Remote Sensing of Environment **92** (No. 4). p. 447–464.

ANDREWS, D. L. (2017): Rayleigh Scattering and Raman Effect, Theory. In: LINDON, J. C., TRANTER, G. E.& KOPPENAAL, D. W. (ed.): Encyclopedia of Spectroscopy and Spectrometry p. 924–930.

ATKINSON, P. M. (2013): Downscaling in remote sensing. International Journal of Applied Earth Observation and Geoinformation **22**. p. 106–114.

ATKINSON, P. M. & TATE, N. J. (2000): Spatial Scale Problems and Geostatistical Solutions. A Review. The Professional Geographer **52** (No. 4). p. 607–623.

BACCINI, A., FRIEDL, M. A., WOODCOCK, C. E. & ZHU, Z. (2007): Scaling Field Data to Calibrate and Validate Moderate Spatial Resolution Remote Sensing Models. Photogrammetric Engineering & Remote Sensing **73** (No. 8). p. 945–954.

BILL, R. (2010): Principles of Geo-Information Systems [German original: Grundlagen der Geo-Informationssysteme]. 5. ed. Berlin. 804 p.

BLÖSCHL, G. & SIVAPALAN, M. (1995): Scale issues in hydrological modelling. a review. Hydrological Process **9**. p. 251–290.

BRADY, W. W., MITCHEL, J. E., BONHAM, C. D. & COOK, J. W. (1995): Assessing the power of the point-line transect to monitor changes in plant basal cover. Journal of Range Management **48** (No. 2). p. 187–190.

CEOS (2019): Committee on Earth Observation Satellites. https://www.radcalnet.org/#!/ (15 April 2019).

CLIFF, A. & ORD, J. K. (1973): Spatial Autocorrelation. Monographs in spatial and environmental systems analysisLondon. 178 p.

CROW, W. T., BERG, A. A., COSH, M. H., LOEW, A., MOHANTY, B. P., PANCIERA, R., ROSNAY, P. de, RYU, D. & WALKER, J. P. (2012): Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. Reviews of Geophysics **50** (No. 2). p. 3675.

CROW, W. T., RYU, D. & FAMIGLIETTI, J. S. (2005): Upscaling of field-scale soil moisture measurements using distributed land surface modeling. Advances in Water Resources **28** (No. 1). p. 1–14.

CZAPLA-MYERS, J., BOUVET, M. & WENNY, B. (2016): The Radiometric Calibration Network (RadCalNet). A Global Calibration and Validation Test Site Network.

DARK, S. J. & BRAM, D. (2007): The modifiable areal unit problem (MAUP) in physical geography. Progress in Physical Geography **31** (No. 5). p. 471–479.

DAVIDSON, S., SANTOS, M., SLOAN, V., REUSS-SCHMIDT, K., PHOENIX, G., OECHEL, W. & ZONA, D. (2017): Upscaling CH4 Fluxes Using High-Resolution Imagery in Arctic Tundra Ecosystems. Remote Sensing **9** (No. 12). p. 1227.

DAVIS, F. W., QUATTROCHI, D. A., RISS, M. K., LAM, N. S.-N., WALSH, S. J., MICHAELSEN, J. C., FRANKLIN, J., STOW, D. A., JOHANNSEN, C. J. & JOHNSTON, C. A. (1991): Environmental analysis using integrated GIS and remotely sensed data. Some research needs and priorities. Photogrammetric Engineering and Remote Sensing **57**. p. 689–697.

DEUTSCHES ZENTRUM FÜR LUFT-UND RAUMFAHRT (DLR) (2015): HySpex Product Guide. OpAiRS - Aiborne Remote Sensing. 1.1. ed. 12 p.

DOXANI, G., VERMOTE, E., ROGER, J.-C., GASCON, F., ADRIAENSEN, S., FRANTZ, D., HAGOLLE, O., HOLLSTEIN, A., KIRCHES, G., LI, F., LOUIS, J., MANGIN, A., PAHLEVAN, N., PFLUG, B. & VANHELLEMONT, Q. (2018): Atmospheric Correction Inter-Comparison Exercise. Remote Sensing **10** (No. 2).(352), 1-18.

EMDE, C., BURAS-SCHNELL, R., KYLLING, A., MAYER, B., GASTEIGER, J., HAMANN, U., KYLLING, J., RICHTER, B., PAUSE, C., DOWLING, T. & BUGLIARO, L. (2016): The libRadtran software package for radiative transfer calculations (version 2.0.1). Geoscientific Model Development **9** (No. 5). p. 1647–1672.

EUROPEAN SPACE AGENCY (ESA) (2015): Sentinel-2 User Handbook. ESA Standard Document. 1. ed. 64 p.

EUROPEAN SPACE AGENCY (ESA) (2017): COPE-GSEG-EOPG-TN-15-0007 (Technical Document). Sentinel-2 Spectral Response Functions. 3.0. ed. 5 p.

EUROPEAN SPACE AGENCY (ESA) (2018): S2 MPC. L1C Data Quality Report. 26. ed. 37 p.

EUROPEAN SPACE AGENCY (ESA) (2019): Copernicus (Overview). https://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Overview3 (2 February 2019).

EUROPEAN SPACE AGENCY (ESA) (2019): S2 MPC. Level 2A Data Quality Report. 9. ed. 37 p.

EUROPEAN SPACE AGENCY (ESA) (2019): Sentinel Online (Validation). https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/validation (2 February 2019).

EUROPEAN SPACE AGENCY (ESA) (2019): Sentinel Online (SENTINEL-2 Radiometric Resolutions). https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric (12 April 2019).

FARR, T. G., ROSEN, P. A., CARO, E., CRIPPEN, R., DUREN, R., HENSLEY, S., KOBRICK, M., PALLER, M., RODRIGUEZ, E., ROTH, L., SEAL, D., SHAFFER, S., SHIMADA, J., UMLAND, J., WERNER, M., OSKIN, M., BURBANK, D. & ALSDORF, D. (2007): The Shuttle Radar Topography Mission. Reviews of Geophysics **45** (No. 2). p. 1485.

FOTHERINGHAM, A. S. & WONG, D. W. S. (1991): The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. Environment and Planning A **23** (No. 7). p. 1025–1044.

GASCON, F., BOUZINAC, C., THÉPAUT, O., JUNG, M., FRANCESCONI, B., LOUIS, J., LONJOU, V., LAFRANCE, B., MASSERA, S., GAUDEL-VACARESSE, A., LANGUILLE, F., ALHAMMOUD, B., VIALLEFONT, F., PFLUG, B., BIENIARZ, J., CLERC, S., PESSIOT, L., TRÉMAS, T., CADAU, E., BONIS, R. de, ISOLA, C., MARTIMORT, P. & FERNANDEZ, V. (2017): Copernicus Sentinel-2A Calibration and Products Validation Status. Remote Sensing **9** (No. 584). p. 1–81.

GEHLKE, C. E. & BIEHL, K. (1934): Certain Effects of Grouping upon the Size of the Correlation Coefficient in Census Tract Material. Journal of the American Statistical Association **29** (No. 185A). p. 169–170.

GONZÁLEZ-ROGLICH, M. & SWENSON, J. J. (2016): Tree cover and carbon mapping of Argentine savannas. Scaling from field to region. Remote Sensing of Environment **172**. p. 139–147.

GOODCHILD, M. F. (2011): Scale in GIS. An overview. Geomorphology **130** (No. 1-2). p. 5–9.

HOLBEN, B. N., ECK, T. F., SLUTSKER, I., TANRÉ, D., BUIS, J. P., SETZER, A., VERMOTE, E., REAGAN, J. A., KAUFMAN, Y. J., NAKAJIMA, T., LAVENU, F., JANKOWIAK, I. & SMIRNOV, A. (1998): AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization. Remote Sensing of Environment **66** (No. 1). p. 1–16.

ISLAM, T., KOKHANOVSKY, A. A., WANG, J. & HU, Y. (2018): Remote sensing of aerosols, clouds and precipitation. First edition. ed. Amsterdam, Netherlands. 11 p.

JING, X., LEIGH, L., TEIXEIRA PINTO, C. & HELDER, D. (2019): Evaluation of RadCalNet Output Data Using Landsat 7, Landsat 8, Sentinel 2A, and Sentinel 2B Sensors. Remote Sensing **11** (No. 5). p. 541.

JOHANNSEN, C. J. & DAUGHTRY, C. S. T. (2009): Surface Reference Data Collection. In: WARNER, T. A., NELLIS, M. D.& FOODY, G. M. (ed.): The SAGE Handbook of Remote Sensing. 1 Oliver's Yard, 55 City Road London EC1Y 1SP. p. 244–256.

JONES, H. G. & VAUGHAN, R. A. (2010): Remote sensing of vegetation. Principles, techniques, and applications. First edition. ed. Oxford. 12 p.

JOSHI, C., MOHANTY, B. P., JACOBS, J. M. & INES, A. V. M. (2011): Spatiotemporal analyses of soil moisture from point to footprint scale in two different hydroclimatic regions. Water Resources Research **47** (No. 1). p. 127.

KAUFMAN, Y. J. & SENDRA, C. (2007): Algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery. International Journal of Remote Sensing **9** (No. 8). p. 1357–1381.

KING, A. W. (1991): Translating models across scales. In: TURNER, M. G. (ed.): Quantitative methods in landscape ecology. The analysis and interpretation of landscape heterogeneity. New York, NY. Ecological studies **82.**

KOKALY, R. F., CLARK, R. N., SWAYZE, G. A., LIVO, K. E., HOEFEN, T. M., PEARSON, N. C., WISE, R. A., BENZEL, W. M., LOWERS, H. A., DRISCOLL, R. L. & KLEIN, A. J. (2017): USGS Spectral Library Version 7. U.S. Geological Survey Data Series 1035. 61 p.

KOTCHENOVA, S. Y. & VERMOTE, E. F. (2007): Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data Part II. Homogeneous Lambertian and anisotropic surfaces. Applied Optics **46** (No. 20). p. 4455.

KOTCHENOVA, S. Y., VERMOTE, E. F., MATARRESE, R. & KLEMM, J. F. J. (2006): Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data Part I. Path radiance. Applied Optics **45** (No. 26). p. 6762–6774.

LANNOY, G. J.M. de, HOUSER, P. R., VERHOEST, N. E.C., PAUWELS, V. R.N. & GISH, T. J. (2007): Upscaling of point soil moisture measurements to field averages at the OPE3 test site. Journal of Hydrology **343** (No. 1-2). p. 1–11.

LEITÃO, P. J., SCHWIEDER, M., PÖTZSCHNER, F., PINTO, J. R. R., TEIXEIRA, A. M. C., PEDRONI, F., SANCHEZ, M., ROGASS, C., VAN DER LINDEN, S., BUSTAMANTE, M. M. C. & HOSTERT, P. (2018): From sample to pixel. Multi-scale remote sensing data for upscaling aboveground carbon data in heterogeneous landscapes. Ecosphere **9** (No. 8).e02298.

LIANG, S., LI, X. & WANG, J. (2012): Atmospheric Correction of Optical Imagery. In: LIANG, S., LI, X.& WANG, J. (ed.): Advanced Remote Sensing. San Diego, United States. p. 111–126.

LILLESAND, T. M., KIEFER, R. W. & CHIPMAN, J. W. (2015): Remote sensing and image interpretation. 7. ed. Hoboken, NJ. 719 p.

LOEW, A. & SCHLENZ, F. (2011): A dynamic approach for evaluating coarse scale satellite soil moisture products. Hydrology and Earth System Sciences **15** (No. 1). p. 75–90.

MCCOY, R. M. (2005): Field methods in remote sensing. New York. 159 p.

MIRALLES, D. G., CROW, W. T. & COSH, M. H. (2010): Estimating Spatial Sampling Errors in Coarse-Scale Soil Moisture Estimates Derived from Point-Scale Observations. Journal of Hydrometeorology **11** (No. 6). p. 1423–1429.

MORAN, P. A. P. (1948): The Interpretation of Statistical Maps. Journal of the Royal Statistical Society. Series B (Methodological) **10** (No. 2). p. 243–251.

MUIR, J., SCHMIDT, M., TINDALL, D., TREVITHICK, R., SCARTH, P. & STEWART, J. B. (2011): Field measurement of fractional ground cover. A technical handbook supporting ground cover monitoring for Australia. Canberra, A.C.T. 48 p.

NASA GODDARD SPACE FLIGHT CENTER (2019): AERONET (Mission). https://aeronet.gsfc.nasa.gov/ (22 February 2019).

NICODEMUS, F. E., RICHMOND, J. C. & HSIA, J. J. (1977): Geometrical Considerations and Nomenclature for Reflectance.

OPENSHAW, S. (1983): The Modifiable Areal Unit Problem. Concepts and Techniques in Modern Geography **38**.

OPENSHAW, S. (1984): Ecological fallacies and the analysis of areal census data. Environment and Planning A **16**. p. 17–31.

OPENSHAW, S. & TAYLOR, P. (1979): A million or so correlation coefficients: three experiments on the modifiable areal unit problem. Statistical Applications in the Spatial Sciences p. 127–144.

O'SULLIVAN, D. & UNWIN, D. J. (2010): Chapter 2. The Pitfalls and Potential of Spatial Data. In: O'SULLIVAN, D.& UNWIN, D. J. (ed.): Geographic Information Analysis. Hoboken, NJ, USA. p. 33–53.

PFLUG, B. (2019): Validation of HySpex and Sentinel-2 Imagery by Upscaling Field Measurements over Water and Grass. (31 March 2019). Personal Communication with: Themann, Britta. Berlin.

PFLUG, B., MAIN-KNORN, M., BIENIARZ, J., DEBAECKER, V. & LOUIS, J. (2016): Early Validation of Sentinel-2 L2A Processor and Products. In: OUWEHAND, L. (ed.): Proceedings of Living Planet Symposium 2016. 9-13 May 2016, Prague, Czech Republic. Noordwijk, The Netherlands. ESA SP **740 (August 2016).**

QUATTROCHI, D. A. & LAM, N. S.-N. (1991): Perspectives on integrating multiscale multitemporal remote sensing data with geographic information systems. In: STAR, J. L. (ed.): The Integration of Remote Sensing and Geographic Information Systems. Baltimore, MD. p. 151–166.

SALMIVAARA, A., PORKKA, M., KUMMU, M., KESKINEN, M., GUILLAUME, J. & VARIS, O. (2015): Exploring the Modifiable Areal Unit Problem in Spatial Water Assessments. A Case of Water Shortage in Monsoon Asia. Water **7** (No. 12). p. 898–917.

SCARTH, P. (2006): State of the paddock. Monitoring condition and trend in ground cover across Queensland. In: SCARTH, P., BYRNE, M., DANAHER, T., HASSET, R., CARTER, J.& TIMMERS, P. (ed.): 13 ARSPC. The 13th Australasian Remote Sensing and Photogrammetry Conference : Earth Observation, from Science to Solutions. Perth, WA.

SCHAEPMAN-STRUB, G., SCHAEPMAN, M. E., PAINTER, T. H., DANGEL, S. & MARTONCHIK, J. V. (2006): Reflectance quantities in optical remote sensing—definitions and case studies. Remote Sensing of Environment **103** (No. 1). p. 27–42.

SCHLÄPFER, D., BOREL, C. C., KELLER, J. & ITTEN, K. I. (1998): Atmospheric Precorrected Differential Absorption Technique to Retrieve Columnar Water Vapor. Remote Sensing of Environment **65** (No. 3). p. 353–366.

SCHMIDT, M., TINDALL, D., SPELLER, K., SCARTH, P. & DOUGALL, C. (2010): Ground cover monitoring with satellite imagery in agricultural areas and improved pastures. final report to the Bureau of Rural Sciences, State of Queensland.

SHI, Y., WANG, J., QIN, J. & QU, Y. (2015): An Upscaling Algorithm to Obtain the Representative Ground Truth of LAI Time Series in Heterogeneous Land Surface. Remote Sensing **7** (No. 10). p. 12887–12908.

SPECTRA VISTA CORPORATION (SVC) (2019): Field Portable Spectroradiometers. https://www.spectravista.com/our-instruments/hr-1024i/specifications/ (17 April 2019).

TOBLER, W. R. (1970): A Computer Movie Simulating Urban Growth in the Detroit Regio. Economic Geography **46**. p. 234–240.

UNITED STATES GEOLOGICAL SURVEY (2019): Spectral Characteristics Viewer. https://landsat.usgs.gov/spectral-characteristics-viewer (10 April 2019).

WANG, J., GE, Y., HEUVELINK, G. & ZHOU, C. (2015): Upscaling In Situ Soil Moisture Observations to Pixel Averages with Spatio-Temporal Geostatistics. Remote Sensing **7** (No. 9). p. 11372–11388.

WESTERN, A. W. & BLÖSCHL, G. (1999): On the spatial scaling of soil moisture. Journal of Hydrology **217**. p. 203–224.

WESTERN, A. W., GRAYSON, R. B. & BLÖSCHL, G. (2002): Scaling of Soil Moisture. A Hydrologic Perspective. Annual Review of Earth and Planetary Sciences **30** (No. 1). p. 149–180.

WONG, D. W. (2009): Modifiable Areal Unit Problem. In: KITCHIN, R. (ed.): International Encyclopedia of Human Geography. Amsterdam. p. 169–174.

WU, H. & LI, Z.-L. (2009): Scale issues in remote sensing. A review on analysis, processing and modeling. Sensors (Basel, Switzerland) **9** (No. 3). p. 1768–1793.

ZANDLER, H., BRENNING, A. & SAMIMI, C. (2015): Potential of Space-Borne Hyperspectral Data for Biomass Quantification in an Arid Environment. Advantages and Limitations. Remote Sensing **7** (No. 4). p. 4565–4580.

ZHANG, X., ZUO, W., ZHAO, S., JIANG, L., CHEN, L. & ZHU, Y. (2018): Uncertainty in Upscaling In Situ Soil Moisture Observations to Multiscale Pixel Estimations with Kriging at the Field Level. ISPRS International Journal of Geo-Information **7** (No. 1). p. 33.

Appendix A: Atmospheric Scattering and Absorption

Atmospheric effects altering the intensity and spectral composition of the sensed radiation include atmospheric scattering and absorption (LILLESAND et al. 2015). Atmospheric scattering refers to the diffusion of incident radiation by particles in the atmosphere in an unpredictable manner (LIANG et al. 2012; LILLESAND et al. 2015). These particles include aerosols, which are solid or liquid particles suspended in air, and gas molecules, such as water vapor, both dramatically changing with time and space (LIANG et al. 2012).

Aerosols may originate from volcanic eruptions, ground dust, sandstorms, forest fires or anthropogenic activities. Their concentration decreases exponentially in the troposphere while they remain relatively stable in the stratosphere due to convection and turbulence (LIANG et al. 2012). Their scattering effect differs depending on the wavelength of the incident electromagnetic waves and the particles' diameter (ISLAM et al. 2018; JONES & VAUGHAN 2010; LIANG et al. 2012).

Specifically, radiation interacts with atmospheric particles or molecules smaller in diameter (Rayleigh scatter) (ANDREWS 2017), with a similar diameter (Mie scatter) (ACHARYA 2017) and much larger in diameter (nonselective scatter) (LIANG et al. 2012; LILLESAND et al. 2015) than the radiation's wavelength. Rayleigh scatter occurs on gases like oxygen, nitrogen or water vapor and mostly affects visible light, i.e. short wavelengths (LIANG et al. 2012). Longer wavelengths, on the other hand, are prone to be influenced by Mie scatter which is caused by dust and aerosols in the atmosphere (LIANG et al. 2012). Nonselective scatter affects all visible and near-to mid-infrared wavelengths and is caused by water droplets of 5 to 100 µm in size making fog and clouds appear white on remote sensing imagery (LILLESAND et al. 2015).

Radiation energy is not only scattered, but also absorbed by the atmosphere's constituents, most importantly by water vapor, carbon dioxide and ozone, and lost for detection (LILLESAND et al. 2015). The radiation is thereby converted from electromagnetic energy into molecular excited oscillation energy. Water vapor has strong absorption bands in the infrared, carbon dioxide predominantly at 2.7, 4.3 and 15 µm and ozone mostly in the ultraviolet domain (LIANG et al. 2012). The wavelength ranges at which the atmosphere is most transmissive are called *atmospheric windows* (LILLESAND et al. 2015). Remote sensing systems are typically designed to focus on exactly those wavelength domains to avoid atmospheric attenuation (JONES & VAUGHAN 2010).

Appendix B: Atmospheric Correction Algorithms

To improve approximation of BOA reflectance, atmospheric correction algorithms are constantly evolving. One very basic approach of atmospheric correction is the removal of haze by dark object subtraction, which uses surface features with no reflectance, such as deep clear water in the near infrared domain, to isolate the path radiance, which can subsequently be subtracted from all image pixels (CHAVEZ 1988). However, as this approach is typically applied uniformly throughout the image, it may ignore an uneven spatial distribution of haze. On top of that, haze removal alone may still leave other atmospheric effects in the signal (LILLESAND et al. 2015). Therefore, more sophisticated atmospheric correction algorithms are needed to better correct atmospheric effects. The use of radiative transfer models of atmospheric scattering and absorption have much potential for large area application (DINER et al. 2005; GORDON 1997; VERMOTE et al. 2002). This concept is adopted for atmospheric correction of major earth observation instruments' imagery, such as that of Moderate Resolution Imaging Spectroradiometer (MODIS) (Terra/Aqua) (VERMOTE et al. 2006), Thematic Mapper (TM)/Enhanced Thematic Mapper Plus (ETM+)/Operational Land Imager (OLI) (Landsat) (MASEK et al. 2006; VERMOTE et al. 2016) and MSI (Sentinel-2) (GASCON et al. 2017). A radiative transfer model (RTM) establishes a characterization of the atmospheric impacts based on physical principles, so the effects can then be removed from the imagery. Therefore, RTMs are critical for the translation of TOA radiance to BOA surface reflectance. Additional information on local atmospheric conditions or surface elevation are often required to feed the radiative transfer model (LIANG et al. 2012).

Hence, an integral part of atmospheric correction is the estimation of atmospheric parameters. The main challenge thereby are the highly dynamic temporal and spatial changes in water vapor and aerosol concentration (LIANG et al. 2012). Effects of water vapor are less problematic, since sensors operate in particularly transmissive atmospheric windows. However, these do not reduce effects from aerosol scattering. Aerosol concentration is typically expressed through Aerosol Optical Thickness (AOT). A variety of methods exist for AOT estimation. The earliest and most common approach exploiting spectral characteristics is the dark target method. It takes advantage of bands only minimally affected by aerosols to determine the surface reflectance, from which the AOT of more aerosol-sensitive bands can be deduced. The standard dark target is Dense Dark Vegetation (DDV) (GASCON et al. 2017; LIANG et al. 2012) and the traditionally used domains are the mid-infrared (MIR) or shortwave-infrared (SWIR), where AOT is calculated using the linear relationship between the 2.1 µm band and the red and blue band in dark target pixels (KAUFMAN et al. 1997a; KAUFMAN et al. 1997b). If a 2.1 µm band is not available, the linear relationship between dense vegetation reflectance in the 0.85 µm near-infrared domain and the red band can be exploited (RICHTER et al. 2007). Examples of the application of the dark target method to estimate AOT are ATCOR4 (RICHTER & SCHLÄPFER 2016a), Sen2Cor (GASCON et al. 2017) and the atmospheric correction routines for MODIS (VERMOTE ET AL. 2002), Landsat Thematic Mapper (TM) (RICHTER 1996) and the Medium-Resolution Imaging Spectrometer (MERIS) (SANTER et al. 1999). Other methods estimating AOT include the use of time-series, imagery from multiple angles, spatial information for histogram matching of clear and hazy regions and the use of polarization information (LIANG et al. 2012). Water vapor (WV) demonstrates the most pronounced absorption features at 940 nm and 1140 nm (LIANG et al. 2012). For the estimation of atmospheric column WV content, LIANG et al. (2012) list a multitude of different algorithms all taking advantage of a WV absorption channel. (LIANG et al. 2012). For instance, the ratio between wide and narrow band radiances at 940 nm can be exploited to retrieve a surface reflectance-independent column water vapor estimate for the atmospheric path (FROUIN et al. 1990). Another important method, of which an adapted version is used for Sentinel-2 WV estimation, is the Atmospheric Pre-corrected Differential Absorption (APDA) method. It is a combination of an iterative partial atmospheric correction and a differential absorption technique (SCHLÄPFER et al. 1998).

Due to the computational complexity of radiative transfer equations, solving them on-the-fly would take a long time. To speed up the atmospheric correction process, radiative transfer equations are solved for a range of different AOT, water vapor and other inputs and stored in Look-Up Tables (LUTs). These LUTs are then readily available during processing for different types of atmospheric conditions occurring in the image (LIANG et al. 2012). The most common codes for atmospheric radiative transfer modelling are the Moderate-Resolution Atmospheric Transmittance and Radiance Code, short MODTRAN (BERK et al. 2008), and Second Simulation of the Satellite Signal in the Solar Spectrum, short 6S (VERMOTE et al. 2006). MODTRAN allows for the calculation of a wide range of effects, offers a selection of seven land-based BRDF models and is not limited to Lambertian surfaces (see Appendix C). Its output is simulated apparent radiance (LIANG et al. 2012; BERK et al. 2008). MODTRAN based software packages include the atmospheric correction modules FLAASSH, ACORN and ATCOR (LIANG et al. 2012). The latter is a module capable of correcting interference from changes in weather and solar elevation angles and, given a Digital Elevation Model (DEM) as input, capable of correcting images with and without terrain (RICHTER & SCHLÄPFER 2016b; RICHTER & SCHLÄPFER 2016a). The other common algorithm for radiative transfer modeling is the widely used 6S code, compiled with the programming language FORTRAN (LIANG et al. 2012). Applicable to imagery from airborne and spaceborne sensors, it accurately models the atmospheric composition and absorption using Lambertian surfaces. The 6S code is also behind the LUTs calculation for the MODIS atmospheric correction algorithm, which additionally requires atmospheric ancillary data and aerosol information (VERMOTE et al. 2006; VERMOTE et al. 1997; KOTCHENOVA et al. 2006). An adjusted version of this atmospheric correction algorithm has been successfully applied to Landsat data in the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) Project (MASEK et al. 2006).

Appendix C: Radiation Terminology

To address the criticism brought forward by SCHAEPMAN-STRUB et al. (2006) that the terminology for radiation quantities is often misused in the remote sensing literature, this paper will thoroughly define all relevant terms. When dealing with radiation quantities, it is important to distinguish angular characteristics of downwelling and upwelling radiation. All radiation quantities are measured within conical or hemispherical beam. Incoming sunlight is a beam of light with hemispherical extent, i.e. an opening angle of 2π [sr]. It consists both of direct sunlight and sunlight scattered by the atmosphere and surface features causing anisotropic (direction-dependent), diffuse illumination. The ratio of diffuse to direct irradiance is wavelength-dependent and therefore impacts directional effects differently across the spectrum (SCHAEPMAN-STRUB et al. 2006). A categorization of incident and reflected light by their angular properties into 9 conceptual cases, of which only 2 are measurable, is provided by NICODEMUS ET AL. (1977). Figure 20 illustrates principle radiation components of a remote sensing signal.

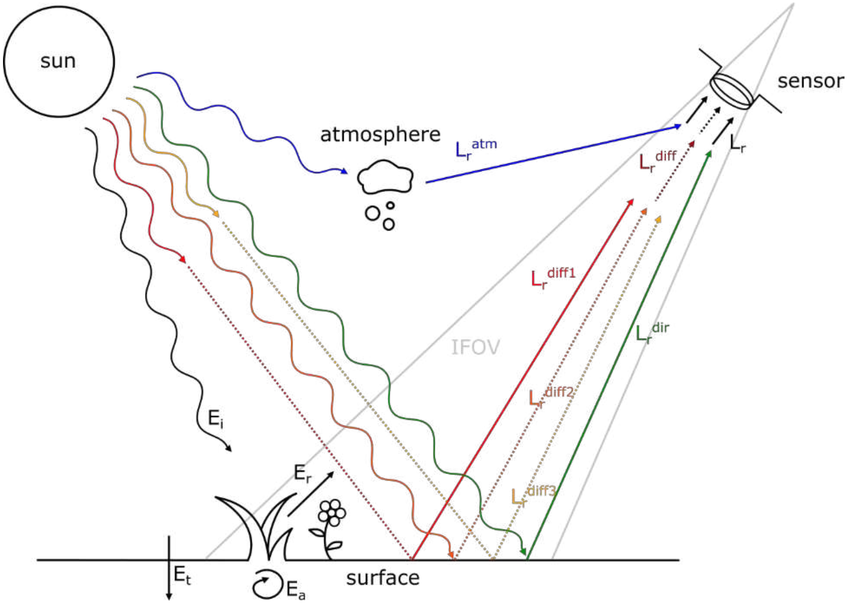


Figure 20. Components of radiation in the signal detected by a sensor (by author).

As mentioned before, the electromagnetic incident energy, or *irradiance* (E_i), reaching the earth's surface is split into *reflected* (E_r), *absorbed* (E_a) and *transmitted* (E_t) *energy*, all of which are a function of wavelength (λ) (8) (LILLESAND et al. 2015). Consequently, the reflected energy amounts to the incoming irradiance without the absorbed and transmitted component (9). The reflected portion of the irradiance from an object is used to describe the object's reflectance characteristics and is referred to as *spectral reflectance* ($\rho_{Si,Sr,\lambda}$) (10). This unitless quantity is dependent on the angular distribution (zenith and azimuth) of all incoming and reflected energy and expressed either as an interval between 0 to 1 or in percent (LILLESAND et al. 2015; SCHAEPMAN-STRUB et al. 2006). Thereby, the zenith and azimuth angles of the reflected energy represent a solid angle cone corresponding to the sensor's Instantaneous Field Of View (IFOV). Spectral reflectance is often illustrated in a spectral reflectance curve (SCHAEPMAN-STRUB et al. 2006). A typical spectral reflectance curve for a given surface is known as the surface's spectral signature.

The signal recorded at the sensor is called *spectral radiance* (L_r) and is also wavelength and IFOV angle dependent. Measured in Watts per area, steradian and wavelength, it is the sum of the radiance scattered by the atmosphere into the IFOV without ever reaching the earth's surface (L_r^{atm}), also known as *atmospheric path radiance*, the radiance of direct sunlight reflected by the surface without any interference (L_r^{dir}) and diffuse radiance (L_r^{diff}) (11). Diffuse radiance is composed of light scattered by the atmosphere on its downward path before touching the surface (L_r^{diff1}), on its upward path after being reflected by the surface (L_r^{diff2}) and by light being scattered before and after being reflected from the surface (L_r^{diff3}) (12). To convert at-sensor radiance to reflectance, the atmospheric path radiance and all three diffuse components ($L_r^{atm}, L_r^{diff1}, L_r^{diff2}, L_r^{diff3}$) need to be subtracted. However, many atmospheric correction algorithms (Schaepman-Strub et al. 2006), neglect the correction for downwelling diffuse light scattered before contact with the surface, which then directly reaches the sensor's IFOV (L_r^{diff1}). While these relationships are valid for homogenous surfaces, remote sensing data also needs to be corrected for the adjacency effect when working with high-resolution data. For field measurements on the ground, the atmospheric path radiance and the two diffuse radiance components reflected from the surface ($L_r^{atm}, L_r^{diff2}, L_r^{diff3}$) are close to zero and no correction is necessary (Schaepman-Strub et al. 2006).

Another related term, particularly relevant for field spectrometry, is the unitless *reflectance factor* (R). It is defined as the radiance of a target surface divided by the radiance of an ideal and diffuse standard surface using the same beam geometry, wavelength range and illumination conditions (13). Spectralon panels are often used as such standard surfaces, also called Lambertian surfaces (Schaepman-Strub et al. 2006). Lambertian surfaces are assumed to scatter radiance in all direction equally (isotropic) regardless of illumination conditions (Philpot & Philipson 2012). For the calibration of field spectrometer data, the measured reflectance (ρ) is calibrated by multiplying it with the reflectance factor of the white reference panel used in the field (14) (Schaepman-Strub et al. 2006).

The closest representation of an object's reflectance properties is provided by the surface's *bidirectional reflectance distribution function* (*BRDF*). It describes how reflectance of a scattered incident beam of light varies for all combinations of illumination and viewing angles within a hemisphere per wavelength (15) (Lillesand et al. 2015; Schaepman-Strub et al. 2006). The BRDF of a Lambertian surface is the reflectance multiplied by π due to its isotropic nature (Philpot & Philipson 2012). While a BRDF can only be approximated and not measured as it is expressed as a ratio of infinitesimal quantities (Nicodemus et al. 1977), it enables the derivation of other useful quantities through integration over the desired angles (Schaepman-Strub et al. 2006).

Irradiance

$$E_{i,\lambda} = E_{r,\lambda} + E_{a,\lambda} + E_{t,\lambda} \tag{8}$$

$$E_{r,\lambda} = E_{i,\lambda} - \left[E_{a,\lambda} + E_{t,\lambda}\right] \tag{9}$$

Reflectance

$$\rho_{Si,Sr,\lambda} = \frac{E_{r,\lambda}}{E_{i,\lambda}} \tag{10}$$

Radiance

$$L_r = L_r^{atm} + L_r^{dir} + L_r^{diff} \tag{11}$$

$$L_r^{diff} = L_r^{diff1} + L_r^{diff2} + L_r^{diff3} \tag{12}$$

Reflectance Factor

$$R = \frac{L_{r,target}}{L_{r,standard}} \tag{13}$$

$$\rho_{Si,Sr,\lambda,target} = L_{r,target} * R \tag{14}$$

BRDF

$$BRDF_\lambda = f_r(S_i, S_r, \lambda) = \frac{\Delta L_r(S_i, S_r, \lambda)}{\Delta E_i(S_i, \lambda)} \tag{15}$$

where

$\lambda = wavelength\ of\ radiation\ [nm]$

$E = irradiance\ [\frac{W}{m^2}]$

$\rho_{Si,Sr,\lambda} = spectral\ reflectance\ [unitless]$

$L_r = radiance\ [\frac{W}{m^2 * sr * nm}]$

$L_r^{atm} = atmospheric\ path\ radiance\ [\frac{W}{m^2 * sr * nm}]$

$S\ = distribution\ of\ direction\ of\ radiation\ (zenith, azimuth)[rad, rad]$

$R\ = reflectance\ factor\ [unitless]$

$BRDF\ = bidirection\ reflectance\ distribution\ function$

$f = function$

$i = incident$

$r = reflected$

$a = absorbed$

$t = transmitted$

$atm = atmospheric$

$dir = direct$

$diff = diffuse$

Appendix D: SVC Data Uncertainty from Sensor Changes

The provided SVC reflectance data over grass were acquired using three temperature-dependent sensors, each working in a different wavelength domain. At the domain boundaries (~980 nm, ~1930 nm), sudden data jumps occur due to different temperature sensors, as illustrated in Figure 21 (b). In this study, these jumps cannot be corrected, as a characterization of their temperature-dependencies was unavailable. Instead, the jumps are noted as data uncertainty. To quantify the magnitude of this uncertainty, the difference between before- and after-jump averages of 10 measurements are calculated. The uncertainty in the raw SVC measurements of 0.08949 represents the mean of the two data jumps in Figure 22. However, the natural slope of the spectral signature of grass immediately after the first data jump most likely caused an overestimation of the respective data uncertainty.

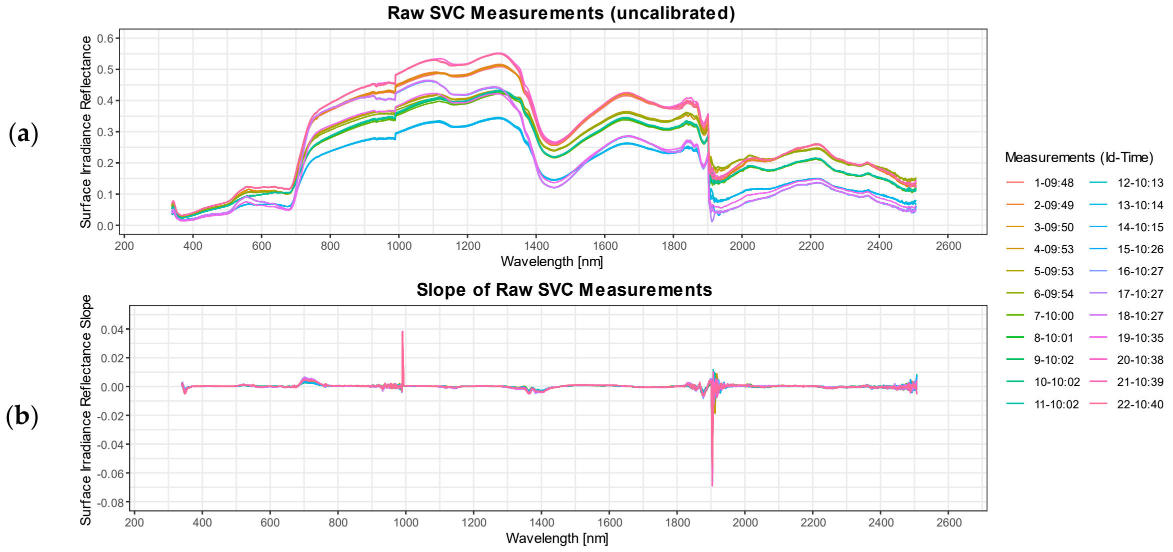


Figure 21. Data jumps in the SVC measurements: (a) raw SVC spectra, (b) slope of raw spectra; both displaying data jumps at ~980nm and ~1930nm due to the use of three different temperature-dependent sensors.

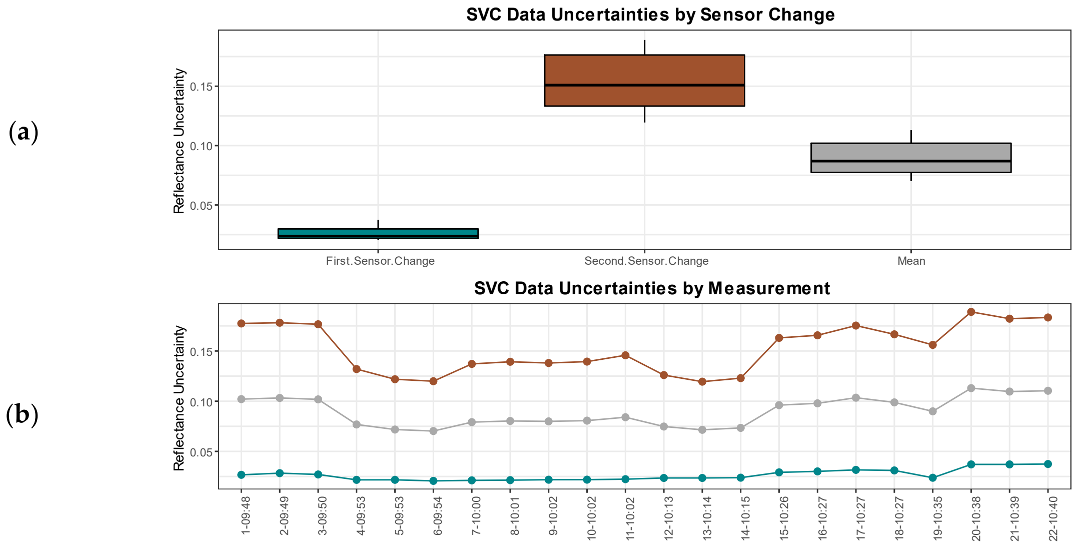


Figure 22. Spectral uncertainty in the raw SVC spectra: (a) by sensor change; (b) by measurement.

Appendix E: Upscaling Attempt using Block Kriging

Block Kriging is closely related to Ordinary Kriging, but instead of interpolating for unknown points, it includes an additional step in which all interpolated points are averaged over a rectangular area. For the calculation of the weights, information on the nature of the spatial dependency or autocorrelation in the original data are required. A variogram provides such information by plotting semivariance (16) between point pairs against their distance to one another (lag) (BILL 2010). In the literature, the use of the term *variogram* is inconsistent, sometimes it designates a covariance plot (ATKINSON & TATE 2000), sometimes a semivariance plot (CROW et al. 2012). To clear up any confusion, the term *semivariogram* will be adopted here to designate plotted semivariance. For Block Kriging to work, the data needs to fulfil three prerequisites: isotropy, stationarity and spatial autocorrelation (BILL 2010). Isotropy refers to a direction-independent spatial dependency (BILL 2010), stationarity to a constant variable mean independent of measurement site (ATKINSON & TATE 2000) and spatial autocorrelation to the similarity of observations according to a spatial pattern (O'SULLIVAN & UNWIN 2010). Firstly, to uncover this spatial pattern, an empirical semivariogram describing the changes in semivariance between two known observations with increasing distance to one another is established. This requires the calculation of distances between each of the observation and all other observations and each point pair's semivariance. Furthermore, the distances are subdivided into lags, distance classes, for which a mean semivariance per lag is determined (17). Then, a theoretical semivariogram (model) is fitted to the mean semivariance per lag representing the rule-set of the underlying spatial autocorrelation. Three main parameter characterize the theoretical semivariogram: the nugget, the range and the sill. The nugget designates the basic variation in very close observation points and may originate from a sampling error, the range is defined as the distance at which spatial dependency between observations ceases to exist and the sill represents the maximum semivariance between two very distant observations (BILL 2010; WESTERN et al. 2002). If a sill exists, the prerequisite of stationarity is fulfilled (WESTERN et al. 2002). To acquire the value of an unknown point, the observed semivariances from the empirical semivariogram and the predicted semivariances of an unknown point from the theoretical semivariogram are used to calculate each observation's weight in the interpolation process ((18)-(19)). The semivariance and prediction matrices are expanded to ensure unbiasedness, i.e. the sum of all weights being equal to 1. For Block Kriging, the values of a whole grid of evenly distributed unknown points within the area of interest need to be solved. Subsequently, all points within the desired support are averaged to produce the Block Kriging estimate (BILL 2010).

Semivariance per observation pair
(Matheron 1965)

$$\gamma(x_i, x_i + h) = \frac{1}{2}[z(x_i) - z(x_i + h)]^2 \tag{16}$$

Mean semivariance per lag
(Bill 2010)

$$\gamma(h) = \frac{1}{2n}\sum_{i=1,n}[z(x_i) - z(x_i + h)]^2 \tag{17}$$

Block Kriging weights
(Bill 2010)

$$z(x_0) = \sum_{i=1,n}\lambda_i \cdot z(x_i) \quad with \quad \sum_{i=1,n}\lambda_i = 1 \tag{18}$$

Calculation of Block Kriging weights
(Bill 2010)

$$A \cdot \begin{bmatrix} \lambda \\ \phi \end{bmatrix} = b \Leftrightarrow A^{-1} \cdot b = \begin{bmatrix} \lambda \\ \phi \end{bmatrix} \tag{19}$$

$$\begin{bmatrix} \gamma_{1,1} & \cdots & \gamma_{1,i} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{i,1} & \cdots & \gamma_{n,n} & 1 \\ 1 & \cdots & \gamma_{n,n} & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \gamma_{1,0} \\ \vdots \\ \gamma_{n,0} \\ 1 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \phi \end{bmatrix}$$

where

$\gamma =$ semivariance
$h =$ separation distance $=$ lag
$i =$ index
$z =$ value
$x =$ observation
$x_0 =$ uknown point
$\lambda =$ weight to estimate value of unknown point
$n =$ number of observations within lag
$A^{-1} =$ inverse of the expanded semivariance matrix of all observation combinations
$b =$ expanded prediction matrix of unknown point
$\phi =$ Lagrange multiplier

$$RMSE = \sqrt{\frac{1}{n}\sum[\gamma_P(x_i, x_i + h) - \gamma_O(x_i, x_i + h)]^2} \tag{20}$$

$$R^2 = \left(\frac{\sum[(\gamma_O - \hat{\gamma}_O) \cdot (\gamma_P - \hat{\gamma}_P)]}{\sqrt{\sum[(\gamma_O - \hat{\gamma}_O)^2] \cdot \sum[(\gamma_P - \hat{\gamma}_P)^2]}}\right)^2 \tag{21}$$

where

$\gamma_P =$ predicted semivariance
$\gamma_O =$ observed semivariance
$\hat{\gamma} =$ mean semivariance
$x =$ observation
$i =$ index
$n =$ number of observations within lag
$h =$ separation distance $=$ lag

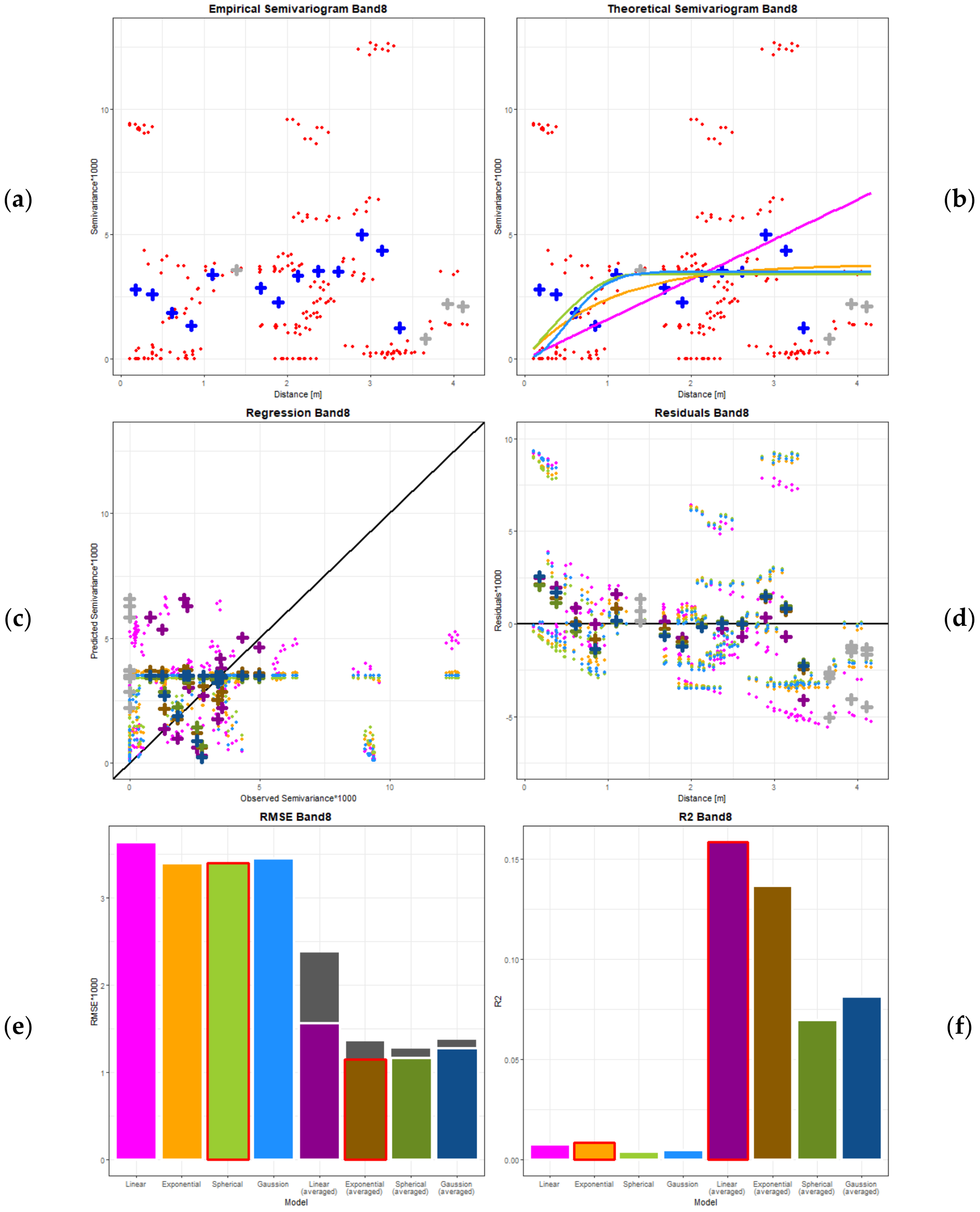


Figure 23. Semivariogram analysis: (a) Empirical semivariogram; (b) theoretical semivariogram; (c) regression plot; (d) residual plot; (e) RMSE plot; (f) R2 plot; Band8 show weak fits across all models (Linear, Exponential, Spherical, Gaussian) with the best-fitting model marked in red.

In this study, the support to which the reference data from the field shall be upscaled are the HySpex pixel footprints covering the measurements. Since Block Kriging relies on spatial variation, it cannot be applied to the MCS measurements whose coordinates have been altered to avoid proximity to the boat. Consequently, only empirical semivariograms for all bands of the SVC measurements were established and linear, exponential, spherical and Gaussian models visually fitted to the mean semivariances per lag. To determine their goodness-of-fit, a regression and residual plot, as well as the RMSE (20) and R^2 (21) of the empirical and theoretical semivariogram have been included, as shown in Figure 23. This was performed using all mean semivariances and all lags, with exclusion of underrepresented lags containing less than 20 % of observations (grey in Figure 23). However, even when excluding underrepresented lags, all models in all bands revealed large RMSEs between 0.000026 (Band1, Gaussian) and 0.0016 (Band8 and Band8A, linear) and extremely low R-squared values between around 0.001 (Band7, spherical) and 0.35 (Band2 and Band4, spherical) indicating insufficient model fits. On top of this, nugget, range and sill were not clearly visible in any of the bands. Without a clear sill, there is no stationarity.

$$z_i = x_i - \hat{x} \tag{22}$$

$$S_0 = \sum_i \sum_j w_{ij} \tag{23}$$

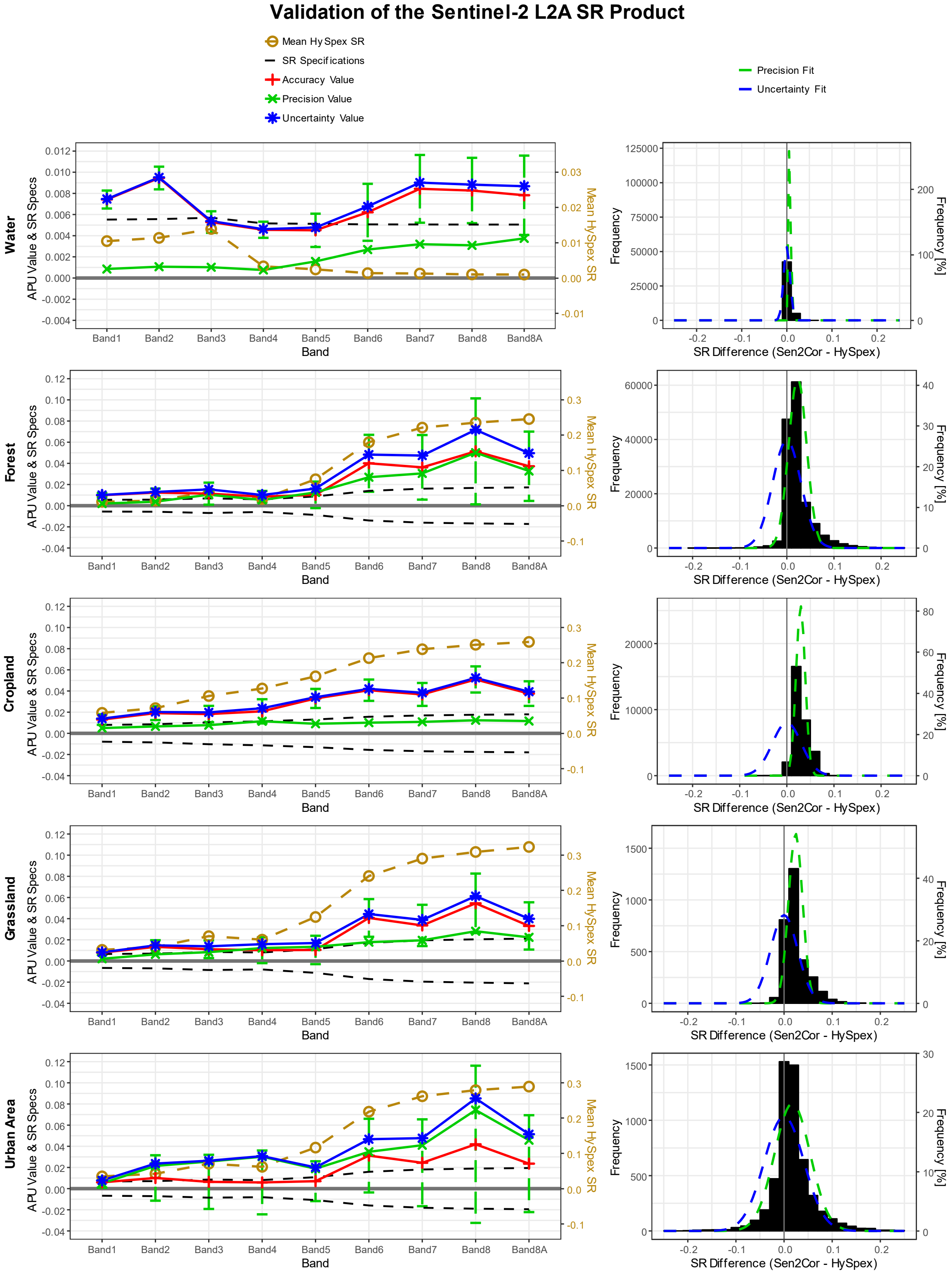
$$I = \frac{\sum_i \sum_j w_{ij} z_i z_j / S_0}{\sum_i z_i^2 / n} \tag{24}$$

where

$i = observation\ location$
$x = observation$
$\hat{x} = observation\ mean$
$z = value\ difference\ to\ mean$
$w_{ij} = elements\ of\ the\ spatial\ weights\ matrix$
$n = number\ of\ observations$

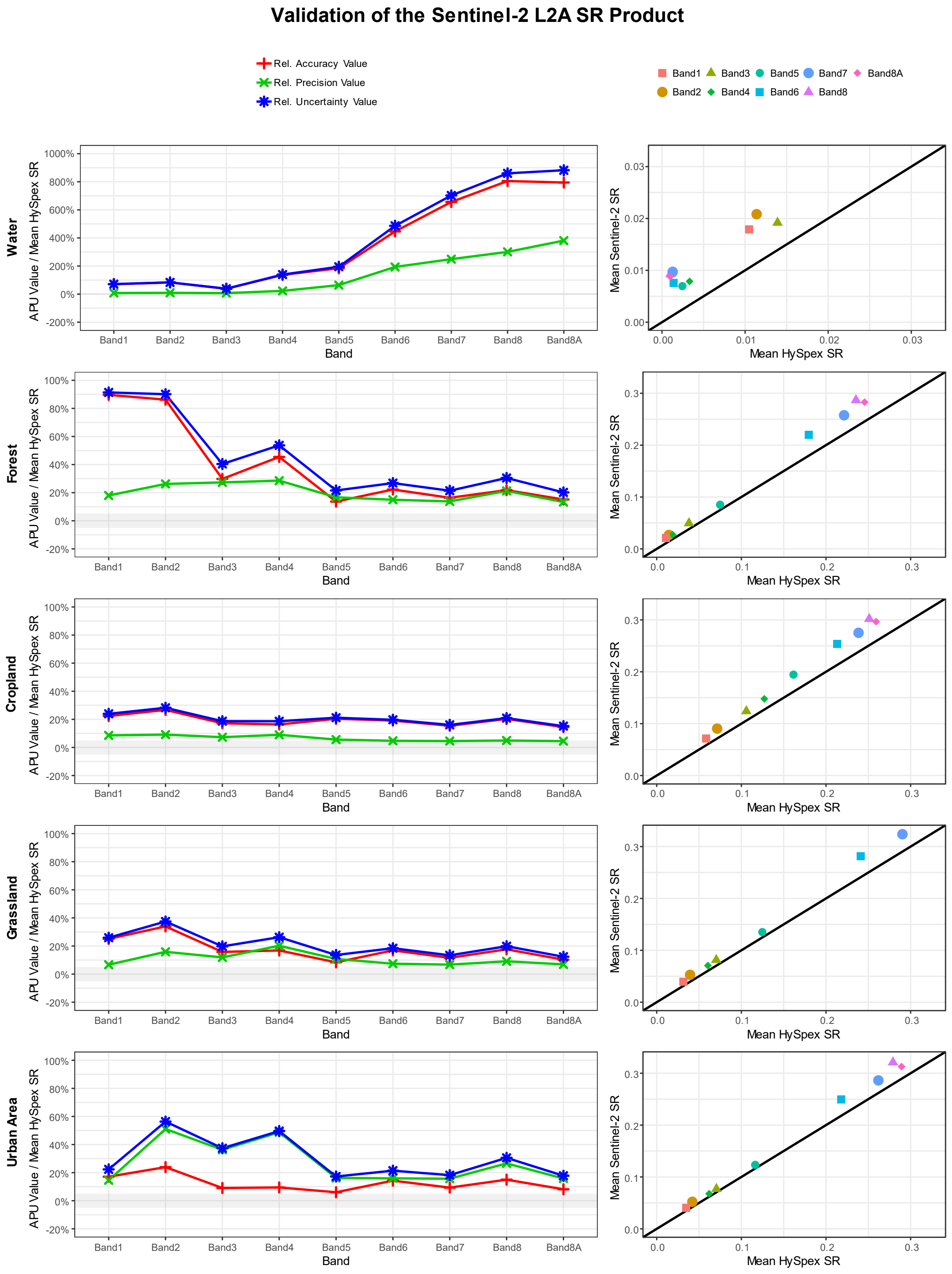
The Moran's I values calculated using the opensource software GeoDa confirm the weak spatial autocorrelation in each band ((22)-(24)). Therefore, the SVC measurements fail to fulfil two of the three prerequisites, stationarity and spatial autocorrelation and cannot be upscaled using Block Kriging.

Appendix F: Sentinel-2 L2A SR Product Validation Supplements



(**a**) (**b**)

Figure 24. Validation of the Sentinel-2 L2A SR for different classes (absolute APU, SR difference): (a) APU plots; (b) surface reflectance difference histograms.



(a) (b)

Figure 25. Validation of the Sentinel-2 L2A SR for different classes (relative APU, regression): (a) relative APU plots; (b) mean SR regression plots.

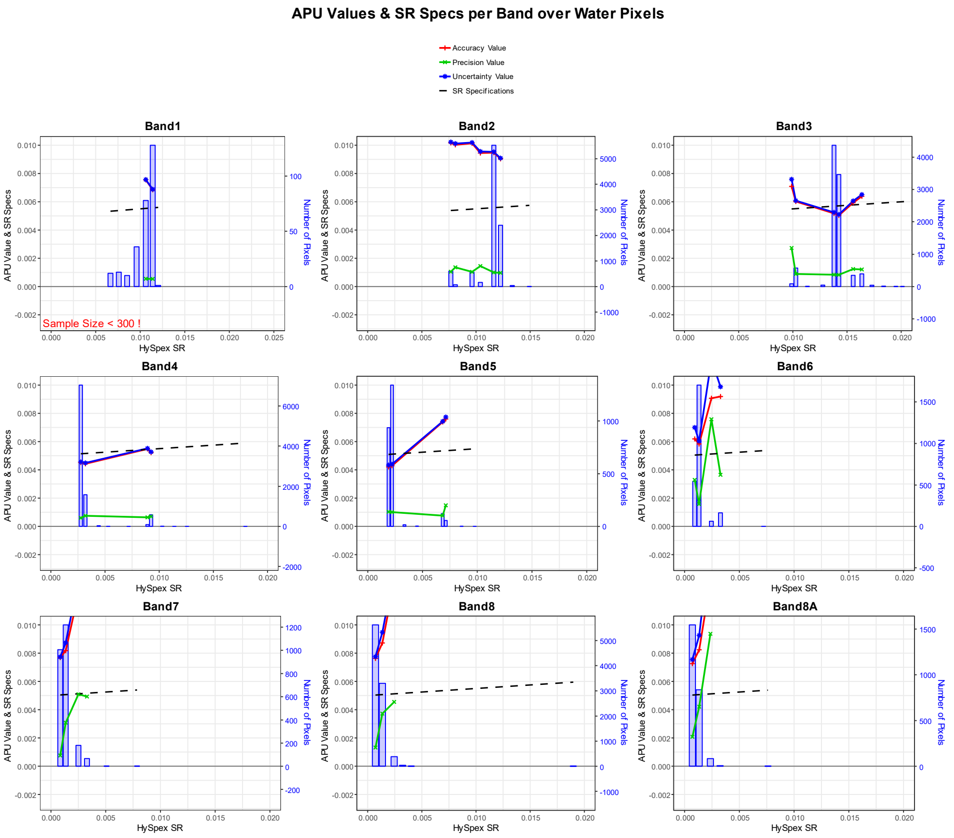


Figure 26. APU values and surface reflectance specifications per band over water.

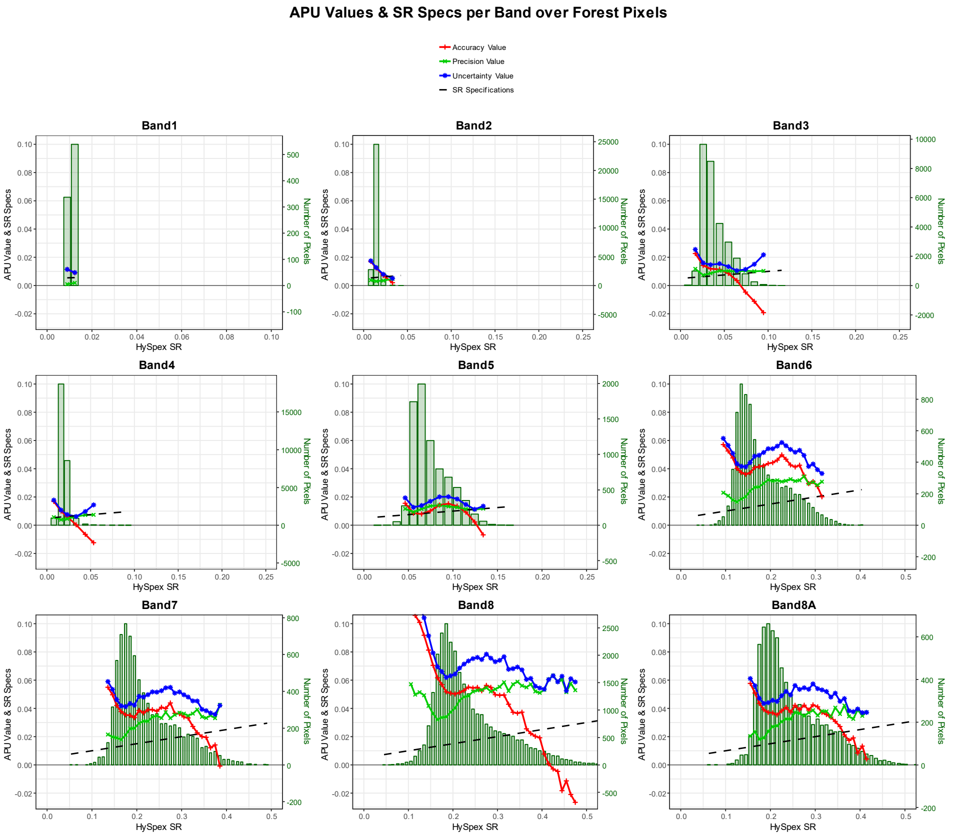


Figure 27. APU values and surface reflectance specifications per band over forest.

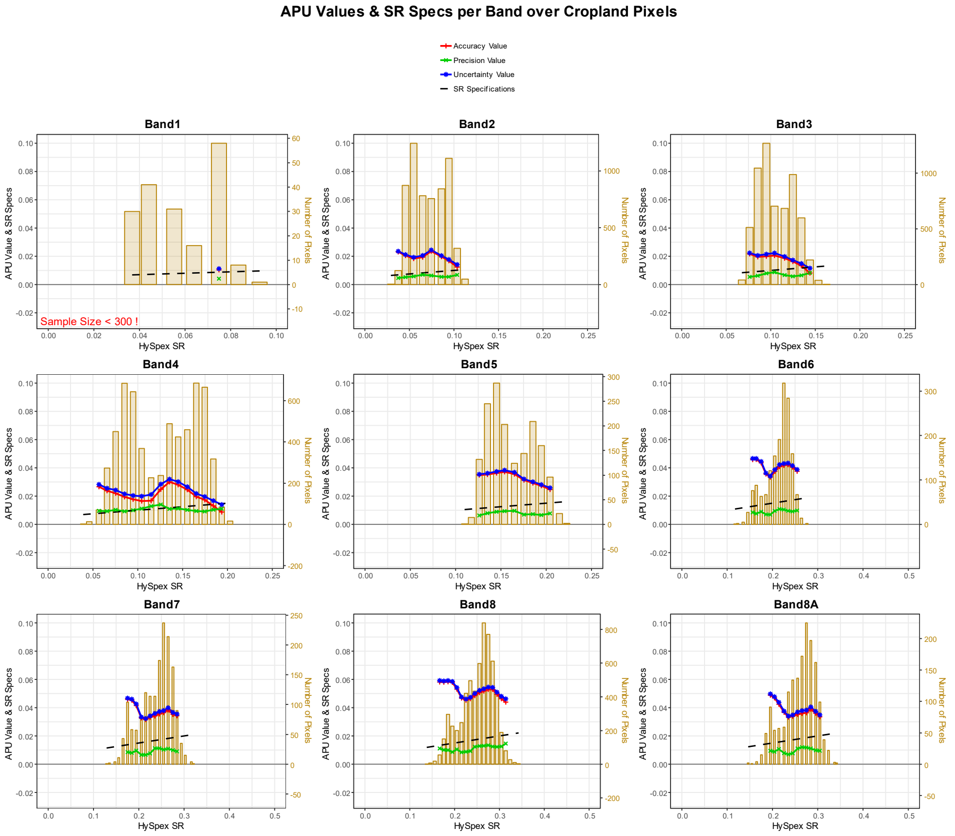


Figure 28. APU values and surface reflectance specifications per band over cropland.

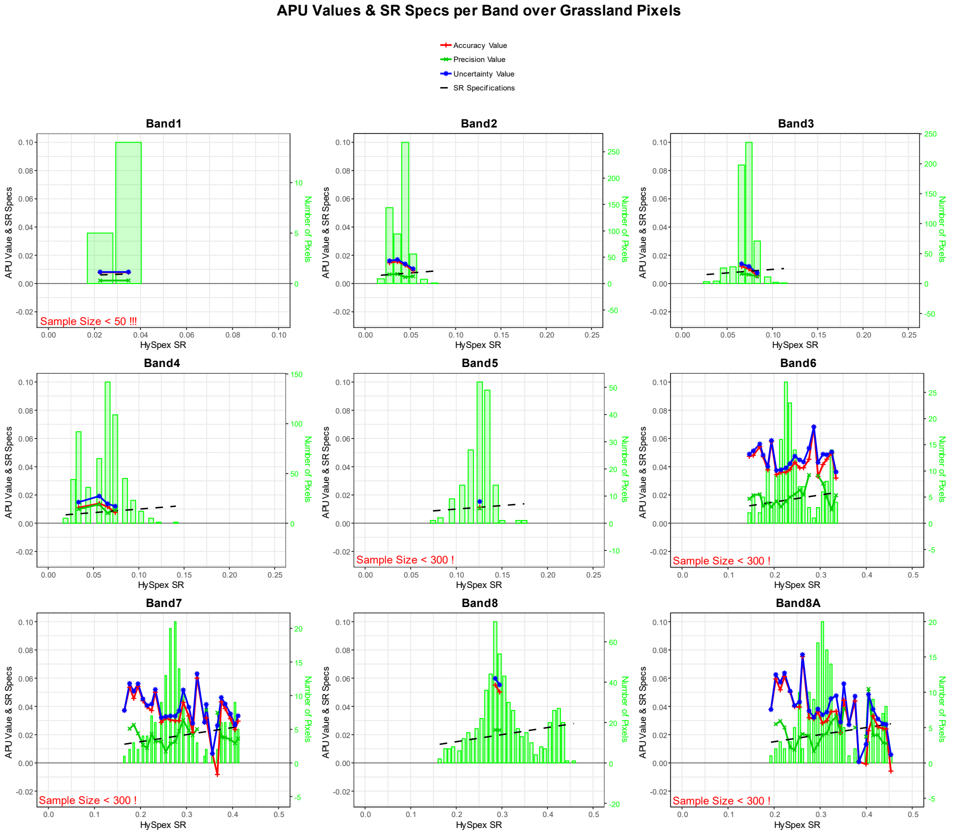


Figure 29. APU values and surface reflectance specifications per band over grassland.

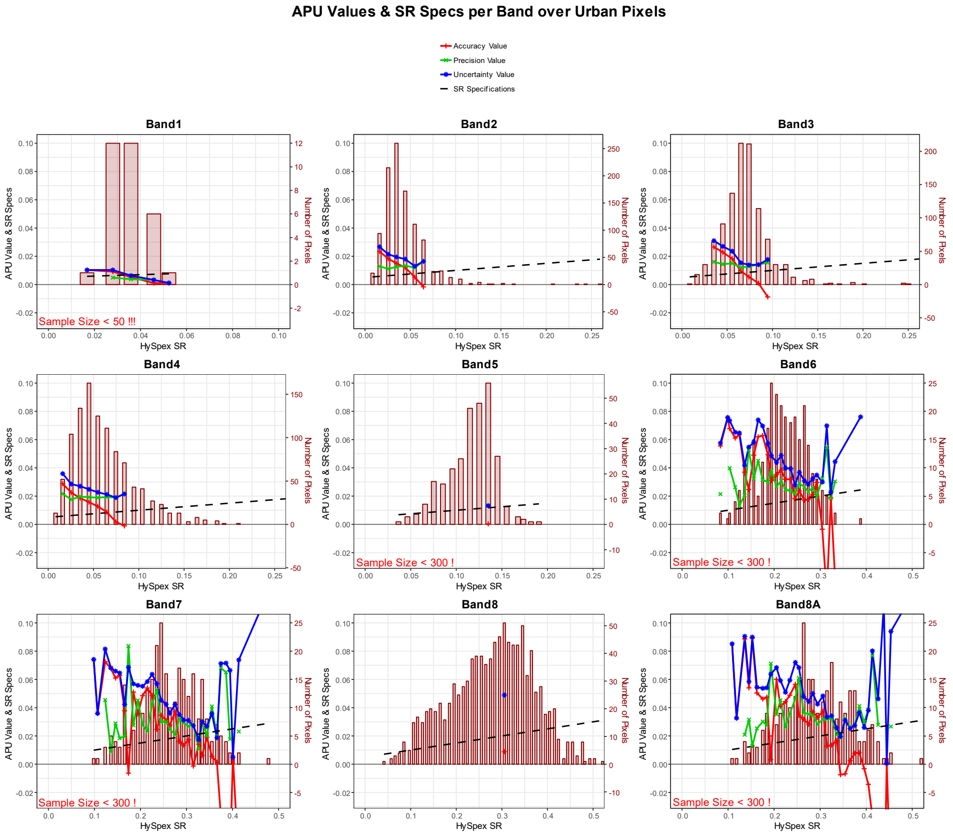


Figure 30. APU values and surface reflectance specifications per band over urban area.

Appendix References

Acharya, R. (2017): Interaction of waves with medium. In: Acharya, R. (ed.): Satellite signal propagation, impairments and mitigation. London. p. 57–86.

Andrews, D. L. (2017): Rayleigh Scattering and Raman Effect, Theory. In: Lindon, J. C., Tranter, G. E.& Koppenaal, D. W. (ed.): Encyclopedia of Spectroscopy and Spectrometry p. 924–930.

Atkinson, P. M. & Tate, N. J. (2000): Spatial Scale Problems and Geostatistical Solutions. A Review. The Professional Geographer **52** (No. 4). p. 607–623.

Berk, A., Anderson, G. P., Acharya, P. K. & Shettle, E. P. (2008): MODTRAN 5.2.0.0 User's Manual. 4 Fourth Avenue, Burlington, MA.

Bill, R. (2010): Principles of Geo-Information Systems [German original: Grundlagen der Geo-Informationssysteme]. 5. ed. Berlin. 804 p.

Chavez, P. S. (1988): An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sensing of Environment **24** (No. 3). p. 459–479.

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., Rosnay, P. de, Ryu, D. & Walker, J. P. (2012): Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. Reviews of Geophysics **50** (No. 2). p. 3675.

Diner, D. J., Martonchik, J. V., Kahn, R. A., Pinty, B., Gobron, N., Nelson, D. L. & Holben, B. N. (2005): Using angular and spectral shape similarity constraints to improve MISR aerosol and surface retrievals over land. Remote Sensing of Environment **94** (No. 2). p. 155–171.

Frouin, R., Deschamps, P.-Y. & Lecomte, P. (1990): Determination from Space of Atmospheric Total Water Vapor Amounts by Differential Absorption near 940 nm. Theory and Airborne Verification. Journal of Applied Meteorology **29** (No. 6). p. 448–460.

Gascon, F., Bouzinac, C., Thepaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., Languille, F., Alhammoud, B., Viallefont, F., Pflug, B., Bieniarz, J.,

CLERC, S., PESSIOT, L., TREMAS, T., CADAU, E., BONIS, R. de, ISOLA, C., MARTIMORT, P. & FERNANDEZ, V. (2017): Copernicus Sentinel-2A Calibration and Products Validation Status. Remote Sensing **9** (No. 584). p. 1–81.

GORDON, H. R. (1997): Atmospheric correction of ocean color imagery in the Earth Observing System era. Journal of Geophysical Research: Atmospheres **102** (No. D14). p. 17081–17106.

ISLAM, T., KOKHANOVSKY, A. A., WANG, J. & HU, Y. (2018): Remote sensing of aerosols, clouds and precipitation. First edition. ed. Amsterdam, Netherlands. 11 p.

JONES, H. G. & VAUGHAN, R. A. (2010): Remote sensing of vegetation. Principles, techniques, and applications. First edition. ed. Oxford. 12 p.

KAUFMAN, Y. J., TANRÉ, D., REMER, L. A., VERMOTE, E. F., CHU, A. & HOLBEN, B. N. (1997a): Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. Journal of Geophysical Research: Atmospheres **102** (No. D14). p. 17051–17067.

KAUFMAN, Y. J., WALD, A. E., REMER, L. A., GAO, B.-C., LI, R.-R. & FLYNN, L. (1997b): The MODIS 2.1-µm channel-correlation with visible reflectance for use in remote sensing of aerosol. IEEE Transactions on Geoscience and Remote Sensing **35** (No. 5). p. 1286–1298.

KOTCHENOVA, S. Y., VERMOTE, E. F., MATARRESE, R. & KLEMM, J. F. J. (2006): Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data Part I. Path radiance. Applied Optics **45** (No. 26). p. 6762–6774.

LIANG, S., LI, X. & WANG, J. (2012): Atmospheric Correction of Optical Imagery. In: LIANG, S., LI, X.& WANG, J. (ed.): Advanced Remote Sensing. San Diego, United States. p. 111–126.

LILLESAND, T. M., KIEFER, R. W. & CHIPMAN, J. W. (2015): Remote sensing and image interpretation. 7. ed. Hoboken, NJ. 719 p.

MASEK, J. G., VERMOTE, E. F., SALEOUS, N. E., WOLFE, R., HALL, F. G., HUEMMRICH, K. F., GAO, F., KUTLER, J. & LIM, T.-K. (2006): A Landsat Surface Reflectance Dataset for North America, 1990–2000. IEEE Geoscience and Remote Sensing Letters **3** (No. 1). p. 68–72.

MATHERON, G. (1965): Regionalized Variables and their Estimation [French original: Les Variables Régionalisées et Leur Estimation]. Paris.

MORAN, P. A. P. (1948): The Interpretation of Statistical Maps. Journal of the Royal Statistical Society. Series B (Methodological) **10** (No. 2). p. 243–251.

NICODEMUS, F. E., RICHMOND, J. C. & HSIA, J. J. (1977): Geometrical Considerations and Nomenclature for Reflectance.

O'SULLIVAN, D. & UNWIN, D. J. (2010): Chapter 2. The Pitfalls and Potential of Spatial Data. In: O'SULLIVAN, D.& UNWIN, D. J. (ed.): Geographic Information Analysis. Hoboken, NJ, USA. p. 33–53.

PHILPOT, W. D. & PHILIPSON (2012): Remote Sensing Fundamentals. Cornell University.

RICHTER, R. (1996): A spatially adaptive fast atmospheric correction algorithm. International Journal of Remote Sensing **17** (No. 6). p. 1201–1214.

RICHTER, R. & SCHLÄPFER, D. (2016a): Atmospheric / Topographic Correction for Airborne Imagery. (ATCOR-4 User Guide, Version 7.0.3, March 2016). Wil, Switzerland.

RICHTER, R. & SCHLÄPFER, D. (2016b): Atmospheric / Topographic Correction for Satellite Imagery. (ATCOR-2/3 User Guide, Version 9.0.2, March 2016). Wil, Switzerland.

RICHTER, R., SCHLÄPFER, D. & MÜLLER, A. (2007): An automatic atmospheric correction algorithm for visible/NIR imagery. International Journal of Remote Sensing **27** (No. 10). p. 2077–2085.

SANTER, R., CARRERE, V., DUBUISSON, P. & ROGER, J. C. (1999): Atmospheric correction over land for MERIS. International Journal of Remote Sensing **20** (No. 9). p. 1819–1840.

SCHAEPMAN-STRUB, G., SCHAEPMAN, M. E., PAINTER, T. H., DANGEL, S. & MARTONCHIK, J. V. (2006): Reflectance quantities in optical remote sensing—definitions and case studies. Remote Sensing of Environment **103** (No. 1). p. 27–42.

SCHLÄPFER, D., BOREL, C. C., KELLER, J. & ITTEN, K. I. (1998): Atmospheric Precorrected Differential Absorption Technique to Retrieve Columnar Water Vapor. Remote Sensing of Environment **65** (No. 3). p. 353–366.

VERMOTE, E., JUSTICE, C., CLAVERIE, M. & FRANCH, B. (2016): Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sensing of Environment **185**. p. 46–56.

VERMOTE, E., TANRÉ, D., DEUZÉ, J. L., HERMAN, M., MORCRETTE, J. J. & KOTCHNOVA, S. Y. (2006): 6S User Guide Version 3. Second Simulation of a Satellite Signal in the Solar Spectrum - Vector (6SV). 55 p.

VERMOTE, E. F., EL SALEOUS, N., JUSTICE, C. O., KAUFMAN, Y. J., PRIVETTE, J. L., REMER, L., ROGER, J. C. & TANRÉ, D. (1997): Atmospheric correction of visible to middle-infrared EOS-MODIS data over land surfaces. Background, operational algorithm and validation. Journal of Geophysical Research: Atmospheres **102** (No. D14). p. 17131–17141.

VERMOTE, E. F., EL SALEOUS, N. Z. & JUSTICE, C. O. (2002): Atmospheric correction of MODIS data in the visible to middle infrared. First results. Remote Sensing of Environment **83** (No. 1-2). p. 97–111.

WESTERN, A. W., GRAYSON, R. B. & BLÖSCHL, G. (2002): Scaling of Soil Moisture. A Hydrologic Perspective. Annual Review of Earth and Planetary Sciences **30** (No. 1). p. 149–180.

ERKLÄRUNG

Ich erkläre, dass ich die vorliegende Arbeit nicht für andere Prüfungen eingereicht, selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe. Sämtliche fremde Quellen inklusive Internetquellen, Grafiken, Tabellen und Bilder, die ich unverändert oder abgewandelt wiedergegeben habe, habe ich als solche kenntlich gemacht. Mir ist bekannt, dass Verstöße gegen diese Grundsätze als Täuschungsversuch bzw. Täuschung geahndet werden.

Berlin, den

Unterschrift