Ludwig-Maximilians-Universität München

Fakultät für Geowissenschaften

Department für Geographie

Bachelorarbeit

Wissenschaftliche Arbeit zur Erlangung des akademischen Grades Bachelor of Science

Fine Tuning of a Haze Detection Threshold in Atmospheric Correction of Space-Borne Optical Imagery

(Feinjustierung des Schwellenwerts für Haze-Detektion in der Atmosphärenkorrektur von optischen Satellitenbildern)

Verfasserin:

Lisa Duka (11414362)

Siemensallee 4, 81379 München

lisa.duka@gmx.de

Betreuer:

Prof. Dr. Lukas Lehnert

Mitbetreuung:

Maximilian Langheinrich, DLR

Datum: München, den 06.10.2020

Content

Figures

Tables

1. Introduction
1.1 Atmospheric Correction

Earth-oriented observations from space are becoming more important with each new mission, that adds more detail to the current spectral, spatial and temporal resolutions covered. The usage fields of remotely sensed data are spreading as more and more scientists and corporations use analysing methods relying on the information brought from space. From change detection to automatic land cover classifications, interdisciplinary approaches let remote sensing become more and more important for the understanding of the earth and its systems.

Atmospheric Correction (AC) plays a huge role in the accurate pre-processing of remotely sensed data. The data stream coming down from space is one of the biggest on earth, which makes the correct use and therefore the processing of this data a question of economics regarding the investment in satellites just as of quality insurance for its users worldwide. This makes the constant improvement of processing techniques of the ground segments a necessity to ensure valid end products. Inside the pre-processing chain, there are various quality levels. Data usually reaches the user in L1C status, which represents geometrically resampled, orthorectified images that contain the Top-Of-Atmosphere (TOA) radiance and have a georeference embedded. After removing all sensor-related distortion artifacts, to get to L2A, the Bottom-Of-Atmosphere (BOA) reflectance needs to be retrieved (ESA, 2020), meaning all the atmospheric absorption and scattering effects need to be assessed and then removed from the image. These effects can also be observed in air-borne imagery or in normal photography as well (Ancuti & Ancuti, 2013). A problem for space-borne imagery consists in the non-availability of measurements of the atmospheric conditions on a global scale with the temporal resolution needed to detect most atmospheric phenomena. If available, in-situ measurements represent point-measurements whereas satellite data contains measurements of an entire region. Mostly, only the information contained in the image itself is used to estimate the contribution of the atmospheric scattering or the atmospheric components like aerosols or gases relevant to the scattering. The temporal and spatial variability of the atmospheric conditions makes dynamic approaches to retrieve this said information a necessity for reliable end products.

The free satellite imagery provided by ESA or NASA usually comes as a L1C processed dataset although it is possible to obtain L2A data as well. The input data for atmospheric correction processors though is L1C, which contains the Top-of-Atmosphere radiance. This means, that the digital number (DN) values contain both surface data and atmospheric contributions. Solar radiance passes the atmosphere twice before reaching the sensor. Once from the sun to the earth and then, reflected from the earth

to the sensor. On this trajectory through the atmosphere, light can be scattered or absorbed by components found in the air.

There are four types of influences typically occurring. First, there is extinction by gases, whose concentration in the atmosphere usually is stable, such as oxygen, carbon dioxide or methane. The resulting spectra contain absorption lines varying in thickness, meaning that in the wavelength affected by the absorption, the signal is lower or completely oppressed (Ientilucci et al. 2019, 32). Molecules, depending on their three-dimensional structure and size, absorb specific wavelengths of the solar light. Figure 1 by Guanter et al. shows the relative transmission capacity of the atmosphere and the absorption windows caused by oxygen, water vapor, ozone and carbon dioxide. Compared with the atmospheric gases, that are evenly mixed in the atmosphere and are distributed over wide spatial ranges, water vapor occurrence is characterized by a huge spatial variability regarding the latitude and season. Even in a single scene, water vapor concentrations can vary, which makes the retrieval of water vapor columns on a per-pixel basis a necessity, whenever the correspondent sensing band at water vapor absorption wavelengths is available (Ientilucci et al. 2019, 33ff).
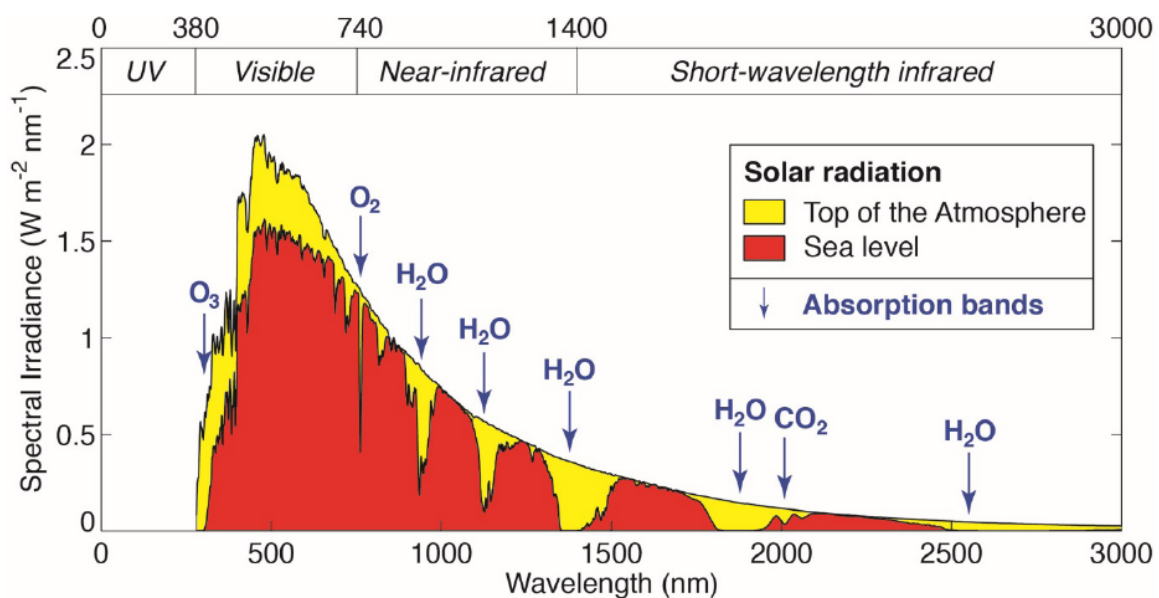
Figure 1: Atmospheric Absorption and Transmittance (Guanter et al. 2016, 2)

Furthermore, the Rayleigh and Mie Scattering models describe different size relations of the scatterer and the scattered light. Rayleigh scattering is a phenomenon usually produced by gas molecules which are a lot smaller than the wavelengths of solar incoming light. This causes an inverse proportionality of the relative amount of scattered light to the wavelength (λ^{-4}), which leads to solar radiation at shorter wavelengths being scattered much stronger than at longer ones, respectively. Mie scatterers usually have the same size as the scattered wavelength, being dust particles or other aerosols. The effect of Mie scatterers is more complicated to describe since it varies with the atmospheric conditions

6

(Chavez et al. 1988, 466). The gas molecules causing the Rayleigh scattering are mostly evenly distributed throughout the atmosphere whereas the aerosols causing the Mie scattering can be much more concentrated locally and vary in time spans of minutes to hours to days. The time span depends on the weather conditions, since precipitation can clear the sky of aerosols while windy conditions can move it spatially.

Over the decades of analysing remotely sensed data, many approaches have been developed to remove the atmospheric contribution. Logically, those most accurate are the ones that incorporate field measurements of the area of interest (AOI) at or close to the sensor overpassing time. In this summary though, the focus is laid on methods that require no external data to conduct the AC. One of the earliest methods was a simple Dark Object Subtraction (DOS) developed by Chavez et al (1988, 478). It is based in the assumption that for each band, a zero-reflectance area or pixel should be present in the scene. If there is a signal present in the data, it is assumed that this is due to atmospheric scattering of light onto the dark pixel. This DN value is then subtracted from the whole scene, band per band, to obtain the corrected atmosphere (Ientilucci et al. 2019, 35).

This method is still present in some modern AC algorithms but incorporated into other techniques to estimate the atmospheric parameters. For example, an AC-software widely used due to its implementation into Harris Corp.'s ENVI image analysis tool is the Quick Atmospheric Correction (QUAC). It uses the DOS method, hence the minimum signal per band, to derive the offset and uses spectral libraries to identify different land cover pixels. This method does only suit spectrally variable scenes and has a limited accuracy (Ientilucci et al. 2019, 38).

More complex, physics-based methods use radiative transfer (RT) functions to simulate the atmospheric conditions. A widely used model is MODTRAN (Moderate Resolution atmospheric transmission), by the U.S. Air Force. The underlying equations enable modelling of horizontally homogeneous atmospheric conditions via physical correlation of various parameters obtained from the image (Berk et al. 2014, 1).

FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes), just like QUAC is available through ENVI software. It contains a spatial convolution step to include adjacency effects, meaning that light scattered into the sensors field of view from outside the targeted area reaches the sensor and influences the reflectance values. This is done by usage of radial exponential kernels whose properties are defined by the type of scatter present, usually one for Rayleigh and one for Mie scattering (Ientilucci et al. 2019, 35 & 41).

Atmospheric Correction Now (ACORN) adds to the estimation of the water vapor column the possible distortion due to surface water contained in vegetation, which gives a more accurate estimation of the

7

column itself while not including kernels to estimate the adjacency effect as FLAASH does (Ientilucci et al. 2019, 40).

ATCOR, an AC program developed at the German Aerospace Center (DLR) (Richter & Schläpfer, 2019). It uses radiative transfer (RT) functions computed by MODTRAN just as ACORN. Like FLAASH, it includes estimations for adjacency effects as well as for bidirectional reflection distribution function (BRDF) effects. For the haze detection, an improved version of a DOS technique is used (Ientilucci et al. 2019, 41, Richter & Schläpfer, 2019, 37f).
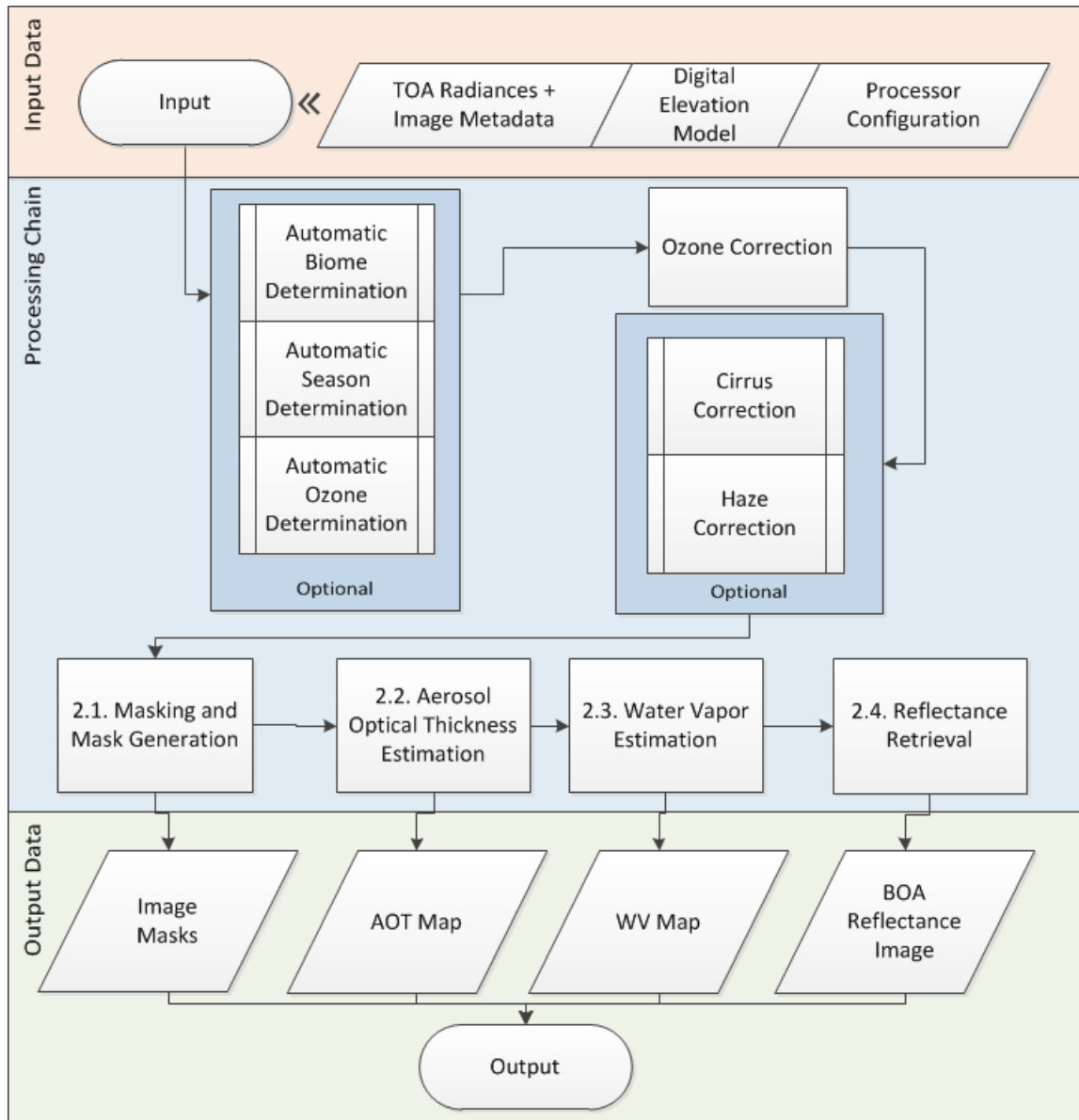
Figure 2: PACO Algorithm Workflow (de los Reyes et al. 2020)

PACO, the python-based atmospheric correction algorithm by DLR (de los Reyes et al. 2020, 2f) is the software used in this work for all parts of the study. The algorithms in PACO are inherited from ATCOR, which means that statements made for PACO in this work also apply to the code of ATCOR. Just as its

predecessor ATCOR, it uses MODTRAN to compute Look-Up-Tables (LUT) with RT functions for different seasons. They are convolved with the sensor spectral response functions and contain parameters considering observational properties, such as the sun zenith angle (dependent on the season and latitude), the relative sun-sensor azimuth angle, the observation off-nadir angle (tilt from the shortest line from sensor to earth surface) and ground elevation (input through a Digital Elevation Model (DEM)). Furthermore, atmospheric parameters like the aerosol optical thickness (AOT) the Water Vapor Column and the correspondent aerosol model and the sensor-specific band central wavelengths are stored in the LUT's . As shown in figure 2, PACO is an algorithm organized in modules, that can be variably implemented in any processing chain. This is an advantage of the usage of python as an object-oriented language, which makes it easy to modularize the code. It uses L1C-processed data and sensor-specific configurations plus a DEM, if available, as input and is designed as a multi-sensor tool which is currently embedded in the ground segment of DESIS and EnMAP missions (Alonso et al. 2019, 9). Figure 2 illustrates the relevance of the Cirrus and Haze Correction as main correction steps for retrieving the BOA reflectance although they are optional since the correspondent phenomena are not present in every satellite image.

1.2 Research Purpose

The focus of this work concentrates on the atmospheric haze phenomenon. Haze as a common image pollution is a known effect for any kind of photography. It originates in the scattering of light in the atmosphere due to water vapor or aerosol particles (Makarau et al. 2014, 5895). In remote sensed data it is essential to correct the effect of haze since the data is processed later to retrieve valuable information about the surface. Compared to thick clouds, that limit the possibility to retrieve surface data, haze is mostly characterized as transparent enough to still retrieve the information from the surface below.

At first, the workflow of the module is explained. The haze module, just as many other modules from PACO are inherited from the DLR-developed ATCOR program. For ATCOR, two methods are available with the Haze Optimized Transform (HOT) always being executed but limited to visible range channels, while the calculation of a Haze Thickness Map (HTM) provides more accurate results and is currently embedded in PACO (Richter & Schläpfer 2019, 102, Makarau et al. 2014, 5895). The haze model describes the signal reaching the sensor L_{sensor} as an addition of the surface component L_0 and the haze contribution HR (Makarau et al. 2014, 5895f).

$$L_{sensor} = L_0 + HR \qquad\qquad (1)$$

To subtract the correct amount of haze contributed, an HTM is calculated. The calculation is based in a Dark Object Subtraction Method (ref. Chapter 1.1), where dark pixels such as shadows are used to determine the thickness of both haze and aerosols. An extrapolated band (usually from two bands from the blue area of the spectrum) is used for the dark pixel search. This is done to avoid overcorrection, which can occur when using only one blue band. Later, a band specific HTM is calculated with the help of a regression coefficient. This ensures that the decreasing influence of Rayleigh-scattering for the longer wavelengths is considered while dehazing. Before the dehazing is done, a binary mask is employed to label hazy and haze-free regions. This is done by taking an additional HTM with larger window size than the band specific ones. By thresholding, the mask is created. The threshold t_{HTM} is calculated as the following with m(HTM) being the mean of the HTM, hs being haze sigma and σ being the standard deviation of the HTM.

$$t_{HTM} = m(HTM) + hs * \sigma(HTM) \qquad\qquad (2)$$

In the current parametrisation of PACO, the haze sigma value is hardcoded as 0. This means, that the standard deviation of the HTM values has no influence on the sensitivity for flagging of haze (Makarau et al. 2014, 5896ff). This work intends to calibrate the threshold of the haze detection module via evaluation of the effects of different haze sigma parameters on the dehazing module and choose,

which parametrisation could improve the performance of the dehazing. Furthermore, the question of a possible scene-dependency of haze sigma and therefore of the threshold regarding the scene incidence and solar zenith angles will be addressed.

2. Data and Methodology

2.1 Sensor Data

2.1.1 Analysis Data

The Landsat datasets, satellite data distributed by NASA and the United States Geological Survey (USGS), are acquired from a near polar, sun-synchronous orbit at an altitude of 705km for Landsat-8 (Roy et al. 2014, 156). The sensors bands are summarized with their correspondent wavelength ranges in table 2. The sensor has a relatively narrow field of view, with the maximum viewing angle being 15° (Roy et al. 2017, 26).

For atmospheric correction, the panchromatic band such as both thermal infrared bands are not used. PACO temporally removes them and the panchromatic band during processing from the datasets that are evaluated for this study. This leaves eight bands to begin with for the dehazing analysis and evaluation. The cirrus band was excluded for all calculations though, due to an expectancy of very low values. It usually records the signal of high-altitude clouds but in case of no presence of clouds, due to the water vapor absorption, only very low ground signals reach the sensor. These very low values can easily create outlier effects when divided in the statistical evaluation.

For Landsat-8, 15 scenes were downloaded from the USGS server. The scenes are chosen from a pool of already studied scenes from the PACO team to facilitate the search for hazy scenes with enough dark dense vegetation (DDV) pixels. The presence of DDV pixels is essential for the first, experimental part of the study, which will be further explained in section 2 of this chapter, along with the reason for the parametrization of haze sigma. The Landsat-8 metadata is summarized in table 1. The sun zenith angle is season-dependent and describes the elevation of the sun with reference to the local zenith. The scene incidence angle can also be described as the off nadir viewing angle, thus describing the sensors shift from the shortest line down to earth.

Scene (Continent)	Date and Time (UTC) of Acquisition	Solar Zenith Angle [deg]	Scene Incidence Angle [deg]	DDV's % at Haze Sigma=10
Bodensee (C. Europe)	19.07.2014: 10:10:12	30.947	0	6.2
Alps (C. Europe)	17.10.2017: 09:58:15	58.223	0	11.6
Berlin (C. Europe)	12.10.2015: 09:56:23	61.378	0	19.4
Irkutsk (N. Asia)	19.08.2015: 03:58:03	41.273	0	18.8

Kyiv (E. Europe)	16.04.2013: 08:51:00	42.289	0	6.5
Sodankyla (N. Europe)	31.05.2013: 09:54:03	46.411	0	24.5
Timisoara (E. Europe)	29.03.2017: 09:20:52	45.832	0	5.0
Manacapuro (S. America)	10.08.2015: 14:12:36	33.884	0	45.7
Baeksa (E. Asia)	19.05.2016: 02:10:38	24.665	0	10.5
Montelimar (W. Europe)	14.07.2013: 10:25:28	28.157	0	3.5
Wallops (N. America)	06.04.2016: 15:40:18	36.267	0	6.6
Penang (S. E. Asia)	27.02.2014: 03:34:39	32.289	0	11.5
Pokhara (C. Asia)	02.04.2014: 04:54:44	31.774	0	10.5
Strzyzow (E. Europe)	12.03.2014: 09:26:48	56.081	0	8.0
Mingo (N. America)	07.11.2013: 16:38:02	33.975	0	22.7

Table 1: Metadata for Landsat-8 tested scenes

The Sentinel data made available via ESAs Sentinel Hub is acquired from a sun-synchronous orbit at 786 km of altitude. The orbit provides a local overpassing time of 10:30 am, which is comparably close to the Landsat-8 overpassing time. The Sentinel System consists of two twin satellites to enable a shorter revisit time, Sentinel A and B (Drusch et al. 2012, 27f). The MSI (multi-spectral instrument) has a wider field of view than Landsat-8 of about 20.6° (Roy et al. 2017, 26). For the data needed for this analysis, from the 13 bands available with Sentinel-2, the cirrus band and the water vapor band are excluded from the calculations. The reasons are the same as for Landsat-8, as described above.

In table 2 the band characteristics of both sensors are displayed (edited from Roy et al. 2014, 157 & Gavilán et al. 2019, 4). For Sentinel 2, the wavelengths of twin A are displayed, since the ones from B only marginally differ and in the tested data set, only the scene of Camaguey, Cuba is taken by Sentinel-B. Regarding Sentinel-2s varying ground resolution per sensor (ref. table 2), PACO can be run with an option to interpolate the 60m and 10m resolution bands to 20m. In this analysis, for all Sentinel-2 data, only the data cube merged to a GSD (Ground Sampling Distance) of 20m was used.

Landsat-8 OLI				Sentinel-2 MSI			
Temporal Resolution:	16 Days			Temporal Resolution:	5 Days		
Bands		[µm]	GSD [m]	Bands		[µm]	GSD [m]
B1:	Coastal	0.435–0.451	30	B1:	Coastal	0.433–0.453	60
B2:	Blue	0.452–0.511	30	B2:	Blue	0.458–0.523	10
B3:	Green	0.533–0.590	30	B3:	Green	0.543–0.578	10
B4:	Red	0.636–0.673	30	B4:	Red	0.650–0.680	10
				B5:	Vegetation Red Edge	0.698–0.713	20
				B6:	Vegetation Red Edge	0.733–0.748	20
				B7:	Vegetation Red Edge	0.773–0.793	20
B8:	NIR	0.767–0.908	10				
B5:	NIR	0.85–0.88	30	B8a:	NIR	0.848–0.881	20
B9:	WV	0.931–0.958	60				
B9:	Cirrus	1.363–1.384	30	B10:	Cirrus	1.338–1.414	60
B6:	SWIR	1.567–1.651	30	B11:	SWIR	1.539–1.681	20
B7:	SWIR	2.107–2.294	30	B12:	SWIR	2.072–2.312	20
B8:	PAN	0.503–0.676	15				
B10:	TIRS	10.60–11.19	100				
B11:	TIRS	11.50–12.51	100				

Table 2: Sensor Band specifications for Landsat-8 & Sentinel-2 (edited from Roy et al. 2014, 157 & Gavilán et al. 2019, 4)

In the first part of the presented study, 13 scenes of Sentinel A and one scene of Sentinel-B are analysed. This has no specific reason, as before with the Landsat dataset the scenes were chosen from a randomly created pool of globally spread scenes, already processed by the PACO Developers Team at DLR. The scene's metadata is summarized in the following table 3. The solar zenith angles in both datasets show approximately the same range, varying from 20° up to 60°. The scene incidence angles though show, that for Sentinel-2, no straight downline nadir view is available in any of the scenes. Three scenes are marked with a * due to an incompleteness in the testing range, that will be further explained in chapter 3.1.

Scene Name (continent)	Date and Time (UTC) of Acquisition	Solar Zenith Angle [deg] [0;70]	Incidence Angle [deg]	DDV Pixels at Haze Sigma = 10 [%]
Israel '15 S2A (Asia)	28.11.2015: 08:37:32	55.074	3.33	3.9
Israel '16 S2A (Asia)	07.03.2016: 08:29:08	41.810	3.33	4.7
France '16 S2A (Europe)	15.02.2016: 10:32:11	59.104	7.08	5.5
Appalachian S2A (N. Am.)	23.12.2016: 16:17:02	61.739	2.89	3.1
Camaguey S2B (C. America) *	22.06.2018: 15:55:29	17.487	7.33	4.6

Bamenda S2A (C. Africa)	23.12.2016: 09:34:12	36.378	6.28	32.3
Luang Namtha S2A (S.E. Asia)	22.09.2016: 03:45:32	39.546	8.57	16.9
Taipei (E. Asia)	27.03.2019: 02:25:51	29.298	2.24	13.1
Tallahassee (N. Am.)	02.05.2017: 16:13:51	21.583	4.58	39.3
Yekaterinburg (N. Asia)	17.03.2017: 07:26:21	58.915	8.27	15.8
Thompson Farm (N. Am.)	25.09.2016: 15:40:02	45.110	8.52	14.0
Portglenone (N.W. Europe) *	08.04.2017: 11:33:21	48.295	9.25	7.6
Palgrunden (N. Europe)	23.05.2017: 10:40:31	38.573	6.96	24.1
Itajuba (S. America) *	21.02.2017: 13:12:41	30.384	1.13	8.8

Table 3: Metadata for Sentinel-2 tested scenes (= incomplete test range)*

2.1.2 Validation Data

In addition to the 29 scenes, that are used for the explorative study part, four DESIS Scenes were used for the validation part. DESIS is a hyperspectral spectrometer developed by DLR and currently operating on the International Space Station (ISS). It has a spectral sampling distance of 2.55 nm across the visible and infrared wavelengths from 400 to 1000 nm with a spatial resolution of 30m. It is the first of a fleet of spectrometers to be installed on the ISS. Contrary to the orbits of Landsat and Sentinel, which are in monitoring orbits, the orbit of the ISS is not sun-synchronous but revisiting times are within a few days, which allows for seasonal or diurnal change detection. In addition, DESIS now delivers data to the user only after ordering it for a specific task (Alonso et al. 2019, 2ff).

DESIS Validation Scenes			
Scene	Date and Time of Acquisition (UTC)	Solar Zenith Angle [deg]	Scene Incidence Angle [deg]
Berlin	22.08.2019: 10:09:03	43.322	29.90
Bavarian Forest 1	26.06.2019: 09:36:53	31.651	4.24
Bavarian Forest 2	27.06.2019: 05:33:37	69.1	15.89
Bavarian Forest 3	29.06.2019: 08:44:57	38.536	22.19

Landsat-8 and Sentinel-2 Validation Scenes			
L8 Montelimar Reference Scene	05.07.2013: 10:31:39	62.964	0
S2 Berlin Reference Scene	22.08.2019: 10:10:31	43.231	9.83

Table 4: Metadata for DESIS, L8 and S2 validation scenes

The last two datasets used for validation purposes are a Sentinel-2 scene of Berlin, correspondent to the DESIS scene of Berlin and a Landsat-8 scene of Montelimar, taken 9 days previous to the hazy Montelimar scene included in part one of this study (ref. table 2). Their Metadata is shown in table 4.

2.2 Methodology
2.2.1 DDV-Approach

As the first part of the study, multiple haze sigma parameter values are to be tested and their results evaluated. The second part of the study is a validation attempt of the prior results.

Figure 3: World Map with tested scenes

In figure 3, the scenes tested in the first part of the study are displayed in their global context. The distribution of scenes should be as widespread as possible, to cover a maximum of different land surfaces inside the criteria for the test population. Apart from Antarctica and Australia, which due to the approach's requirement of a presence of Dark Dense Vegetation are not represented, a good distribution of the test scenes is achieved.

In the experimental part, the test scenes described in chapter 2.1 are processed 6 times with different haze sigma values. To be able to observe more clearly the effects of the dehazing module, a full AC, including the compensation for aerosols is not conducted, instead, the option to only dehaze is selected in the PACO configurations. The tested values are [-0.5, -0.35, -0.25, 0, 0.25, 0.5]. In the early phase of this part, the -0.35 value was added due to a first processing test and a following evaluation of results that pointed into the direction of -0.5 and -0.25 giving the better results. Within the first runs with the mentioned range of parametrisation, a general effect of positive vs. negative values was

detected. The smaller the number (negative, respectively) the more sensitive is the haze flag threshold. An approach to test the processed scenes was developed and designed to be working without any external data inputs. Some are necessary later in the validation step though where a clear scene counterpart serves as external validation. The first assumption for this approach is the spectral similarity of pixels classified as Dark Dense Vegetation(DDV). Since the classification as such pixel type already depends on certain spectral characteristics, the assumption is quite simple. DDV pixels are defined in PACO like in ATCOR code via a threshold using the SWIR or NIR bands and their correlation to the red band (Richter & Schläpfer 2019, 47). PACO uses these pixels to estimate the AOT (which is determined for 550nm) and visibility index, so a DDV mask is already available thanks to the PACO workflow (de los Reyes et al. 2020, 3). The second assumption is that the variance of the spectral characteristics is the lowest in spatially conjunct areas, mostly forests, here referred to as patches. While, as already mentioned, the patches are all similar due to their classification as DDV pixels, the assumption here states that one patch, is spectrally coherent. Most forests do not have a huge variance in tree species present and therefore no relevant differences in spectra. Along seasonal changes, this of course is not true. But since individual scenes are considered, one can expect, as shown in figure 4 that for one patch the spectral profile will be very similar and coherent in patterns. Therefore, in figure 4, two spectra are shown from the same patch one (part one and two), compared to another, spatially disconnected patch two. The wavelength for figure 4 is given in µm. Each patch is first considered individually regarding the ratio before calculating weighted mean ratios for the scenes.

Figure 4: Spectral Profiles of DDV-Patches from Bodensee L-8 scene

The requirement for a valid patch does not only include the coverage with DDV pixels but also the presence of areas flagged as haze and areas flagged as clear by the haze mask threshold. A scheme for the valid patch criteria is presented in figure 5. To make sure, that less outliers get into the analysis

18

and to ensure that a patch contains enough pixels for a statistically conclusive mean value , only patches that have a minimum of 100 pixels in both hazy and clear parts are defined as valid. In two cases, Israel 2016 and Penang, the minimum pixel amount is decreased to 80 pixels due to a very low number of patches valid.
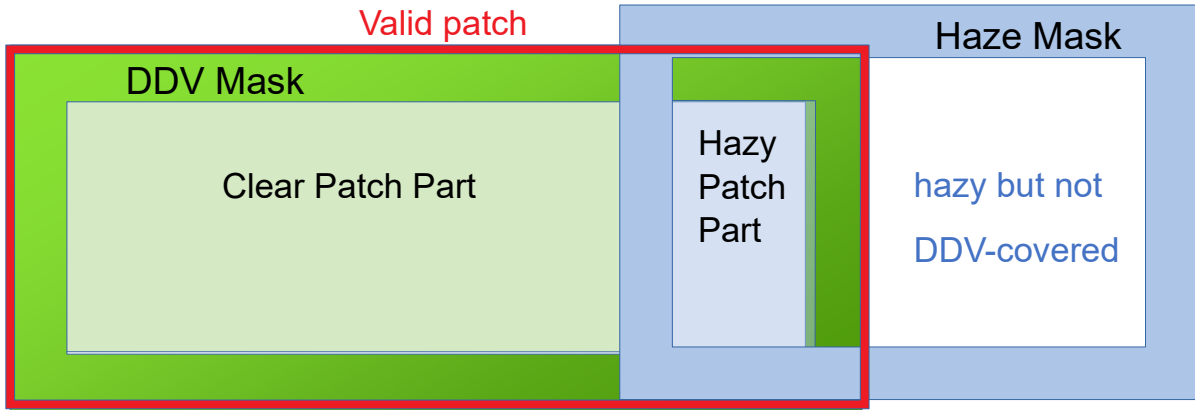

Figure 5: Scheme of Valid Patch Criteria

The DDV mask is obtained by running the full AC once with a very high haze sigma value. Through experiments, it was found that the value 10 leads to zero percent of haze flagged in all the scenes. This is necessary because PACO removes pixels from the DDV-mask if they are also defined as hazy in the Haze-Mask. The classification of the PACO algorithm is designed in such way, that all pixels are defined with one property. Therefore, with the haze sigma value of 10, it is made sure, that the DDV-Mask covers all the scenes DDVs since no haze is flagged. This allows to find patches with both hazy and clear parts later. After creating the DDV-mask for each scene, they are run once with each haze sigma value to be tested, with the configuration set to only dehaze and not to perform a full AC.

The scenes are then individually analysed with Jupyter notebook scripts using the numerical and scientific computing libraries numpy and scipy such as gdal library, which is an assistant module for working with geospatial, multi-dimensional image data. They are also the only external libraries, PACO relies on, to work (de los Reyes et al. 2020, 6) . The dehazed TOA outputs together with the scenes DDV mask, serve as an input to the scripts, to only evaluate the effects of the dehazing module. A full AC would compensate all aerosols present and therefore hide the concrete effects of haze sigma changes in the haze detection performance.

Due to the spectral similarity assumption, which is applied for each patch individually (meaning, two independent patches should vary more than one continuous patch), an evaluation of the dehazing module is possible by introducing the spectral ratio.

$$R_{p,b,r} = \frac{Mean(Clear)_{p,b,r}}{Mean(Hazy)_{p,b,r}} \tag{3}$$

This value takes the mean spectral value of the clear patch part, divided by the mean value for the hazy part. Speaking of 'runs' in this context always means processing of the test scenes, and 'parametrization', with the different haze sigma values described in the beginning of this chapter. So, this value is calculated for each patch (p), in each band(b) and for each run(r) with a different haze sigma. The patch number ranges from [1; $N_{validpatches}$], the band number depends on the sensor, being b=7 for Landsat-8 data and b=11 for Sentinel-2 scenes. Therefore, a maximum of six runs is possible covering the whole range of tested haze sigma values. The ratio resembles the similarity of a hazy and a clear part inside a patch. The closer to 1 this value approximates, the more accurate this patch is dehazed. If the binary decision of the haze distribution (Haze Flag) is correct, one would expect matching spectra from the hazy and clear patch parts. Since the haze sigma parametrization has a direct influence on the percentage of pixels flagged as haze, one would assume that the optimal choice for haze sigma leads to the spectral ratio closest to 1.

To get results for each scene, a mean ratio over all patches of a scene is calculated the following way, with $N_{validpatches}$ being the number of valid patches for each scene.

$$R_{b,r} = \frac{\sum_{i=1}^{N_{validpatches}} R_{p,b,r}}{N_{validpatches}} \qquad (4)$$

To calculate a conclusive error, the standard deviation of $R_{b,r}$ is calculated in the same step and then used to calculate a standard error of the mean (SEM) (Walpole et al. 2012, 277).

$$\sigma(R)_{b,r} = \sqrt{\frac{\sum_{i=1}^{N_{validpatches}}(R_{p,b,r} - R_{b,r})^2}{N_{validpatches}}} \qquad (5)$$

$$SEM(R)_{b,r} = \frac{\sigma(R)_{b,r}}{\sqrt{N_{validpatches}}} \qquad (6)$$

This SEM considers the distribution around R=1 with including the standard deviation, while also including the number of valid patches per scene. Thus, a higher standard deviation from the mean ratio increases the SEM just as a low number of patches available. These values are calculated scene-wise, for each band and each run of each scene.

It is necessary to include $N_{validpatches}$ into the calculations since not only is it scene-dependent regarding the distribution and occurrence of DDVs, but it is also changing with the tested haze sigma values. As previously described, the influence of a varying haze sigma can be found in the thresholds sensitivity to flag a pixel as haze or not. Depending on the combination of the spatial distribution of DDVs, their overall percentage and the spatial distribution of haze in the scene, the number of patches matching the criteria as valid, can change drastically. The more clustered the haze in an image is, the lower is the number of valid patches at high haze flagging due to the fact, that valid patches can only be found on

the edges of the hazy areas. Image examples of this behaviour can be found in chapter 3.1. When evenly distributed, reduced haze flagging is expected to produce lower numbers of valid patches.

To further compare the performance of each run, a weighted mean over all bands of one scene is calculated with w_b being the weighting factor for each band, $\overline{R_r}$ the mean weighted ratio and $\Delta\overline{R_r}$ the mean weighted error (Taylor 1988, w. p.).

$$w_b = \frac{1}{SEM_b{}^2} \qquad (7)$$

$$\overline{R_r} = \frac{\sum_{i=1}^{n_{Bands}} w_b * R_{b,r}}{\sum_{i=1}^{n_{Bands}} w_b} \qquad (8)$$

$$\Delta\overline{R_r} = \sqrt{\frac{1}{\sum_{i=1}^{n_{bands}} w_b}} \qquad (9)$$

$\overline{R_s}$ and its respective error are calculated over all scenes for each run to obtain mean results for each haze sigma parametrization containing all tested scenes. The weighting factor here takes the mean weighted error for each scene, calculated in the previous step. In the same way as equations 10-12, but once calculated per sensor (with n= 15 for Landsat, and n=11 for Sentinel's complete scenes), the error matrices found in chapter 3.1 are calculated per band and per run, over the sensors scenes.

$$w_s = \frac{1}{\Delta\overline{R_r}^2} \qquad (10)$$

$$\overline{R_s} = \frac{\sum_{i=1}^{n=29} w_s * \overline{R_r}}{\sum_{i=1}^{n=29} w_s} \qquad (11)$$

$$\Delta\overline{R_s} = \sqrt{\frac{1}{\sum_{i=1}^{n=29} w_s}} \qquad (12)$$

The values calculated are used to assess the effects of different haze sigma on the dehazing performance. The goal is to determine an improved haze sigma to be used in PACO and to explore a possible scene-dependency of the optimal parametrization. Due to the limited amount of time for this study, no explicit testing for scene-dependency was conducted as the results did not point toward any specific parameter being scene-dependent. Therefore, the best haze sigma value from the experimental part of this study will be implemented and tested in the validation part as a static parameter.

2.2.2 Inter Sensor Validation

To evaluate dehazing results optimally, a scene with field measurements during the overpassing time would be the best choice. Since one would have to predict hazy conditions and such data is not easy to acquire without efforts, a different method was chosen. For the inter sensor validation, two images of the same area are compared in their BOA reflectance values. The underlying assumption here is,

that the ground reflectance does not change during the short interval of those two sensors overpassing. The interval consists of roughly one minute and a half, which leaves the probability of a changed atmosphere very low.

First, a DESIS image of the scene is fully atmospheric corrected, including the dehazing. Then, a Sentinel-2 image of the same area within few minutes is corrected with the same configuration. Both are run through the PACO-Algorithm with the haze sigma value resulting from the DDV-Analysis in part one of the study.

After running the scenes, they are loaded into ENVI and geographically linked. Then, with help of the classification map of PACO, a location with haze present is chosen. An individual window size is chosen for each scene, since the spatial resolution of a pixel varies, being 20m for Sentinel-2 and 30m for DESIS. Therefore, the window size for the DESIS-Scene is set to 2x2 and the one for Sentinel-2 to 3x3. The mean spectra are extracted via a built-in function of ENVI and plotted together. Since DESIS has a shorter range of wavelengths covered, the Sentinel-2 scene is not displayed further than 1µm.

2.2.3 Temporal Validation

For the temporal analysis, first, a scene combination taken out of the work of Makarau et al. 2014 is taken to compare both the dehazed TOA radiance and the BOA reflectance values. The haze-contaminated scene is already used for the explorational part one of this study. The clear scene counterpart is taken 9 days before, which provides the opportunity to compare different atmospheric conditions while still giving consistency in terms of surface changes. For the Bavarian Forest DESIS scenes, the same analysis is carried out. Those scenes are taken within four days, presenting a very high temporal resolution compared to the Montelimar dataset. The DESIS dataset is provided by the neighbouring center of remotely sensed data, the DFD.

After running the scenes with the respective new haze sigma from part one of this study, the dehazed TOAs are compared for both hazy and clear image areas. To conduct this, the corresponding scenes are loaded into ENVI, geographically linked, and combined with the correspondent classification map of the hazy scene. In case of Bavarian Forest, all three haze maps are loaded and only areas are taken that present the same classification in all three images. A trade-off is made here since the areas' homogeneity in terms of haze-flag, out of which the window is taken, cannot be assured. The 10x10 pixel window used to retrieve the spectral profiles for validation itself, contains homogeneous pixels.

3. Results

3.1 Single Scene-Based Results

Figure 6: Results of DDV-Approach summarized
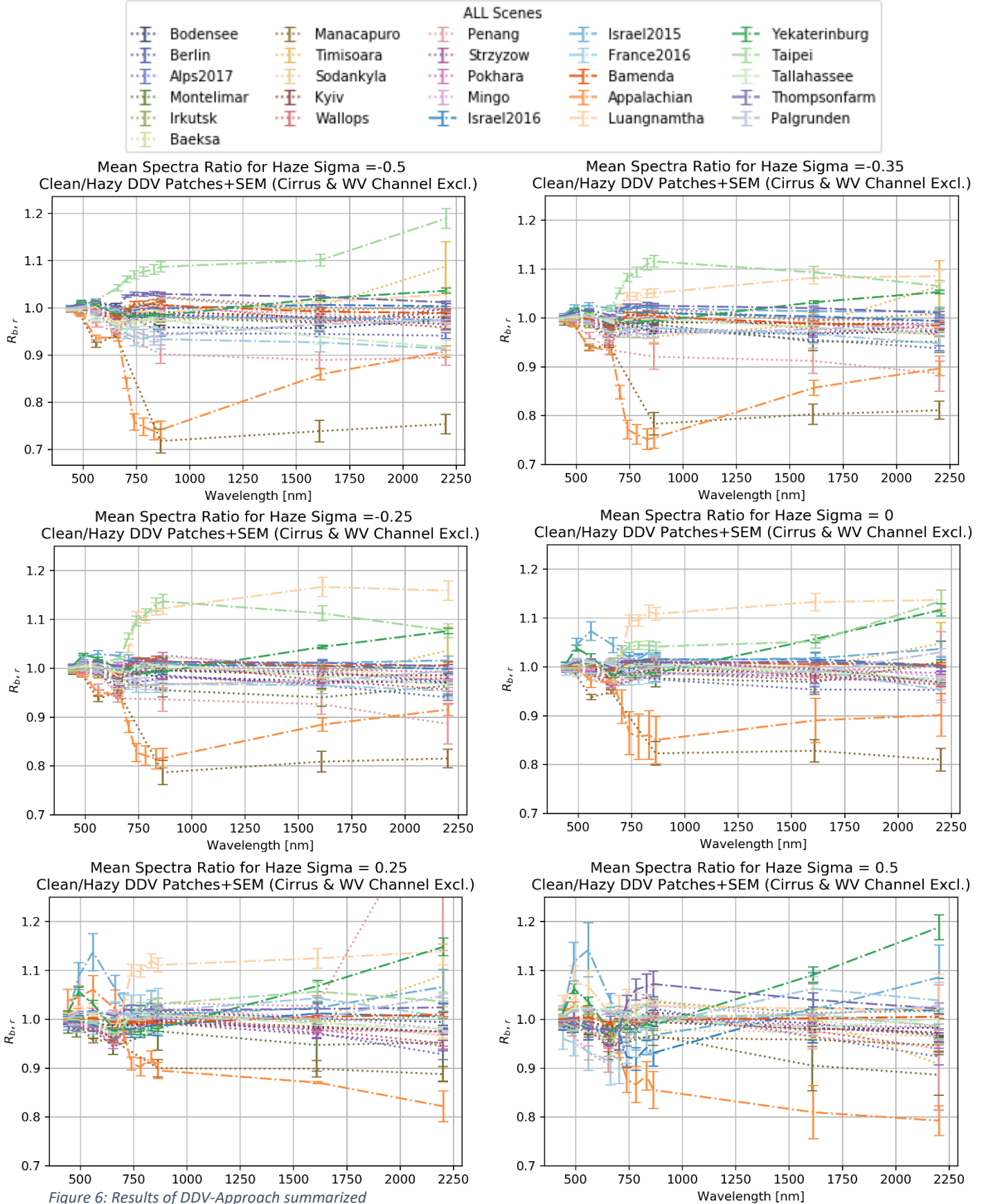
In figure 6, an overview over all tested scenes and PACO-Runs is presented. $R_{b,r}$ is plotted with the corresponding SEMs as error bars. To distinguish between sensors, different line styles are chosen with the dotted lines representing the Landsat-8 datasets and the dash-dotted ones representing Sentinel-2 data. To first evaluate the results, the plot over all scenes with parametrisation of haze sigma = 0 is visually compared with the other plots. The scenes Itajuba, Camaguey and Portglenone (marked with * in the metadata tables in chapter 2.1.1) are shown separately afterwards since it was not possible to obtain results for haze sigma = -0.5, due to the valid patches criteria. For those scenes, the parameter -0.5 flags too much haze to find any valid patches that also contain clear areas.

For the current value of 0, most scenes cluster around $R_{b,r}$=1 with a ± 5% variance. Obvious outliers are Appalachian and Manacapuro, with a tendency of 5-20% towards undercorrection. Luang Namtha shows an opposite behaviour with tendency for overcorrection of 8-15%. Yekaterinburg and Taipei are performing well in the visible range but show a tendency towards overcorrection of 10-15% for the SWIR bands. When comparing with haze sigma=0.25 in the visible range, the tested scenes show a stronger deviation from $R_{b,r}$=1, especially Israel 2016. The L8-scene Penang shows an exceptional high error of 30% for the SWIR band at 2250 nm. The plot for haze sigma=0.5 shows a still increasing variance comparing with 0 and 0.25. In the visible range, an average deviation of ± 10-15% can be observed, which is increasing with the wavelength to ± 20% deviation from $R_{b,r}$=1.

Turning to the negative tested values, a ± 3-7% deviation around $R_{b,r}$=1 can be observed for the visible range. The vegetation red edge (VRE) bands of the Sentinel-2 scenes show two kinds of deviations. For Luang Namtha and Taipei, this consists in a 10-15% chance of overcorrection while in Appalachian, the opposite deviation can be observed ranging from 10-20% towards undercorrection.

The same outlier tendency as observed at -0.25 can be found for Taipei and Appalachian, though reduced for Luang Namtha, which shows a 5-10% tendency for overcorrection now. Apart from the outliers, in the visible range values distribute around $R_{b,r}$=1 with a deviation up to ±5%. In the NIR-SWIR range, not considering the outliers as well, a ±10-15% distribution can be observed.

The overcorrection tendency for the VRE bands for Luang Namtha and Taipei are further reduced, now below 10%. On the opposite side, Appalachian shows a 22-26% of undercorrection in those bands. Looking at the SWIR band, Timisoara presents a large error bar with a 5-14% chance of overcorrection. The deviations in the visible range look like the ones for haze sigma=-0.35. But turning towards longer wavelengths, a slightly increasing deviation tendency ±12-15% can be observed.

For the incomplete test scenes without patches for haze sigma = -0.5, the plots are shown in figure 7. Here, the plots for 0 show a 2-5% chance of overcorrection in the visible range, which increases in the VRE bands and SWIR band up to 5-15% of overcorrection. The tendency for overcorrection increases

24

with positive haze sigma values, presenting almost 30% chance of overcorrection for the vegetation red edge bands at haze sigma = 0.5.

Decreasing the haze sigma to negative values seems to switch this tendency, at least for the scenes Portglenone and Itajuba. Camaguey still presents an overcorrection chance of 5-8% for -0.25, increasing up to 10% for the VRE bands at -0.35. Itajuba shows a drop for the VRE bands that is comparable to Appalachian Scene behaviour. One possible explanation for the inconsistent results in
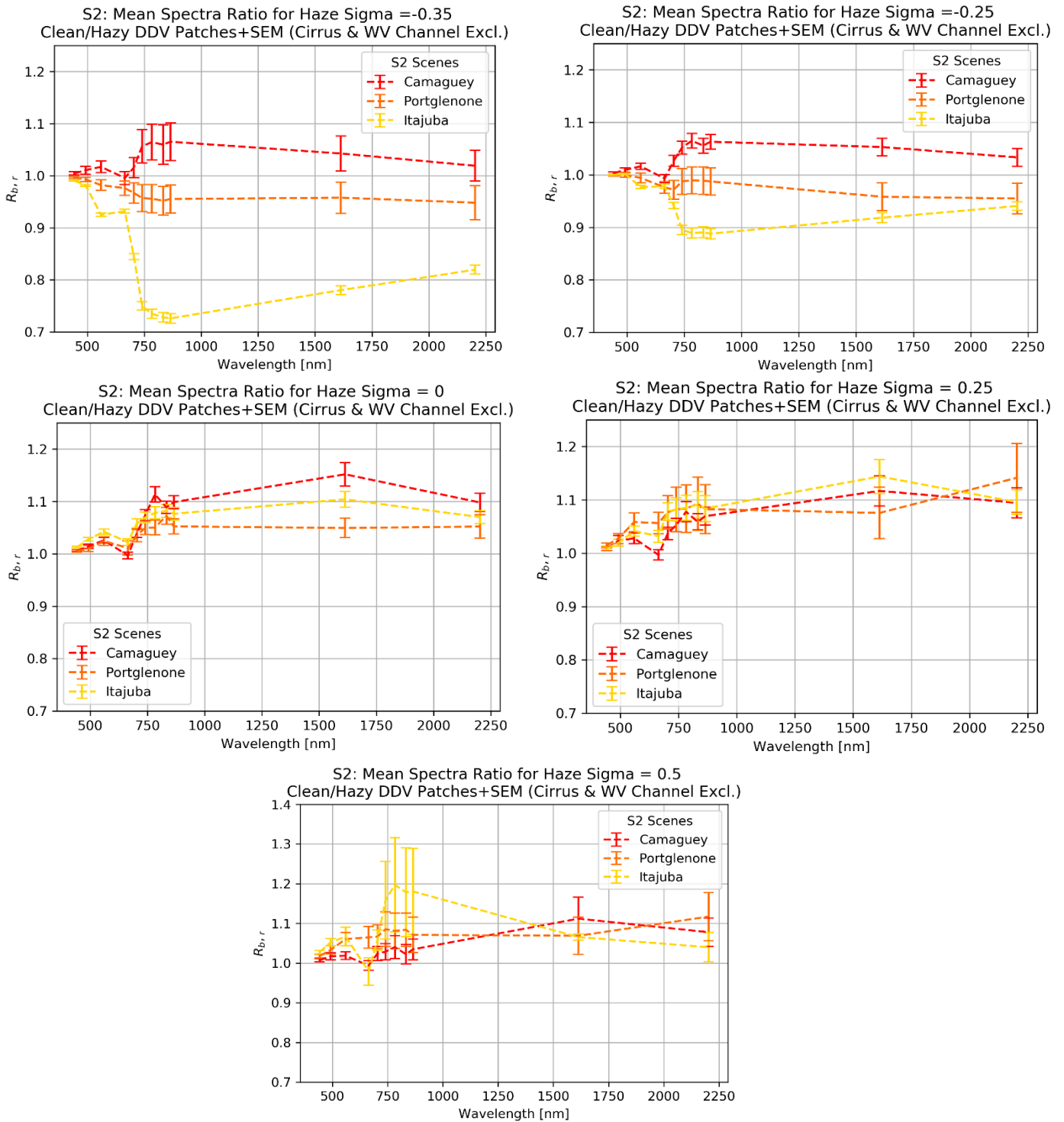

Figure 7: Results for incomplete scenes summarized

25

the VRE bands of Sentinel-2 compared with the Landsat-8 scenes might be the influence of the BRDF effects for those wavelengths. According to Roy et al. (2017, 33) Sentinel-2A images are stronger influenced by BRDF effects due to the sensors wider field of view and varying scene incidence angles. This influence also seems to increase with the wavelength, being 0.035 for blue, 0.057 for red and 0.078 to about 0.1 for NIR to SWIR wavelengths (Roy et al. 2017, 25). As shown in the metadata tables for both sensors (ref. Chapter 2), Sentinel Data shows a variety of off-nadir views while the Landsat Data consistently provides nadir-view. Yekaterinburg and Luang Namtha both have incidence angles higher than 8° (8.27° and 8.57°, respectively). Both show outlying behaviour, especially in the positive haze sigma range. Thompson Farm though, with an incidence angle of 8.52° shows no exceptional behaviour in any run, comparable to Portglenone with an incidence angle of 9.25° which as well has consistent results. Itajuba on the other hand, having an incidence angle of 1.13 presents a very outlying behaviour. Concluding, BRDF effects might have an influence on the Sentinel-2 results but cannot be identified solely as cause of the outliers in VRE bands.

Regarding the higher errors for the positive haze sigma tests, they can partly be explained with the approaches' criteria themselves. As shown in Chapter 2, the SEM includes the number of valid patches
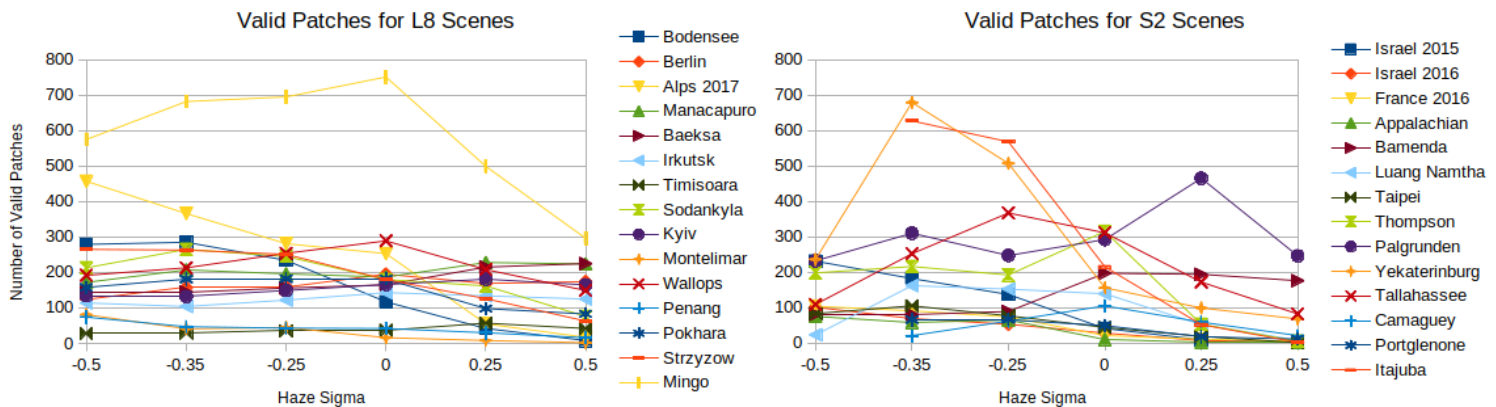

Figure 8: Valid Patches for all tested scenes

as denominator. So, a low number of valid patches for one test run leads to higher SEMs for the corresponding haze sigma value. But as it is shown in figure 8, the number of valid patches to be found does not show an overall tendency to decrease with increasing haze sigma values. As already mentioned in chapter 2.2.1, the number of patches available for analysis is dependent on two factors. First, the spatial distribution of DDVs and second, the spatial distribution of haze.

Alps 2017, with an evolution of valid patches constantly and strongly decreasing with increasing haze sigma values, presents the following DDV and Haze Mask, as shown in figure 9 on the left side.
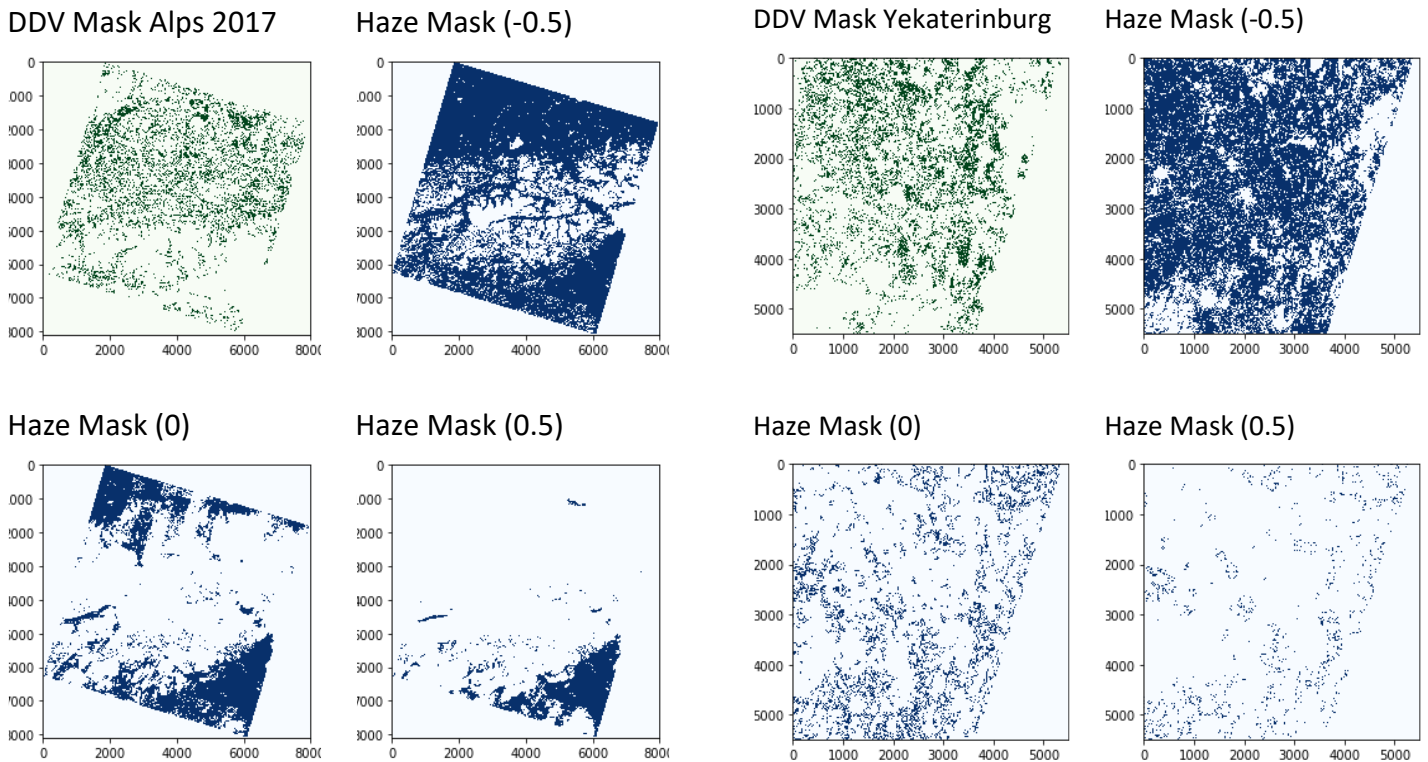

Figure 9: DDV and Haze masks for Alps and Yekaterinburg

The first image coloured green presents the DDV mask and the blue coloured ones the Haze Masks from haze sigma = (-0.5, 0, 0.5). The darker green represents DDVs present, whereas the dark blue represents haze present. The four images on the left belong to Alps 2017. The haze distribution shows a rather clustered behaviour, and especially with increasing the haze sigma, the spatial clustering of the haze becomes more obvious. In addition, the hazy areas for haze sigma = 0 and = 0.5 are matching with the areas of least dense DDV occurrence, which explains the drastic decrease in valid patches.

Yekaterinburg, on the right side of figure 9, shows an evolution of patches rather different, although the DDV maps are comparable considering distribution just as the percentage of DDVs varies from 11.6% (Alps) to 15.8% (Yekaterinburg). This is due to the way more evenly distributed haze which facilitates finding a lot of patches that have both enough hazy and enough clear pixels.

For Landsat-8, the average valid patch number per run is 2659 valid patches, with 3117 as a maximum patch number for the run of -0.35 and a minimal patch number of 1665 for 0.5. For the Sentinel-2 scenes, the number is lower, partly due to the non-availability of -0.5 results of the three scenes mentioned above. Here, an average patch number of 1802 is available with a maximum of 2918 for -0.35 and a minimum of 631 for 0.5. In total this adds up to an average of 4462 patches being analysed per run, with a maximum of 6035 at -0.35 and the minimum at 2296 for 0.5.

L8	Band 1(442.982)	Band 2 (482.589)	Band 3 (561.335)	Band 4 (654.608)	Band 5 (864.571)	Band 6 (1609.091)	Band 7 (2201.249)
-0.5	0.9950 ± 0.0454	0.9934 ± 0.0492	0.9885 ± 0.0578	0.9777 ± 0.0630	0.9774 ± 0.0752	0.9667 ± 0.0741	0.9639 ± 0.0763
-0.35	0.9949 ± 0.0454	0.9930 ± 0.0494	0.9884 ± 0.0578	0.9752 ± 0.0628	0.0628 ± 0.0733	0.9733 ± 0.0720	0.9668 ± 0.0745
-0.25	0.9950 ± 0.0461	0.9926 ± 0.0499	0.9985 ± 0.0579	0.9753 ± 0.0622	0.9885 ± 0.0726	0.9759 ± 0.0715	0.9666 ± 0.0746
0	0.9953 ± 0.0471	0.9929 ± 0.0505	0.9894 ± 0.0589	0.9769 ± 0.0632	0.9935 ± 0.0716	0.9812 ± 0.0713	0.9711 ± 0.0766
0.25	0.9950 ± 0.0481	0.9914 ± 0.0524	0.9894 ± 0.0615	0.9725 ± 0.0663	1.0026 ± 0.0747	0.9854 ± 0.0732	0.9749 ± 0.0802
0.5	0.9942 ± 0.0503	0.9919 ± 0.0543	0.9892 ± 0.0630	0.9719 ± 0.0696	1.0081 ± 0.0763	0.9849 ± 0.0751	0.9659 ± 0.0814

Table 5: Error matrix per band and haze sigma for L8

Another observation made regarding the errors is, that in most cases, except for some outliers in the positive haze sigma runs like Israel 2016, the scenes tend to have lower errors in the visible bands and higher errors in the infrared bands. In table 5, the error matrix per band and run is presented over all the Landsat-8 scenes and in table 6 over all Sentinel-2 scenes except for the ones mentioned above, that do not have results on haze sigma = -0.5. The correspondent central wavelengths are given in nanometers for each band. In figure 10, a graphic plot of table 6 can be observed.
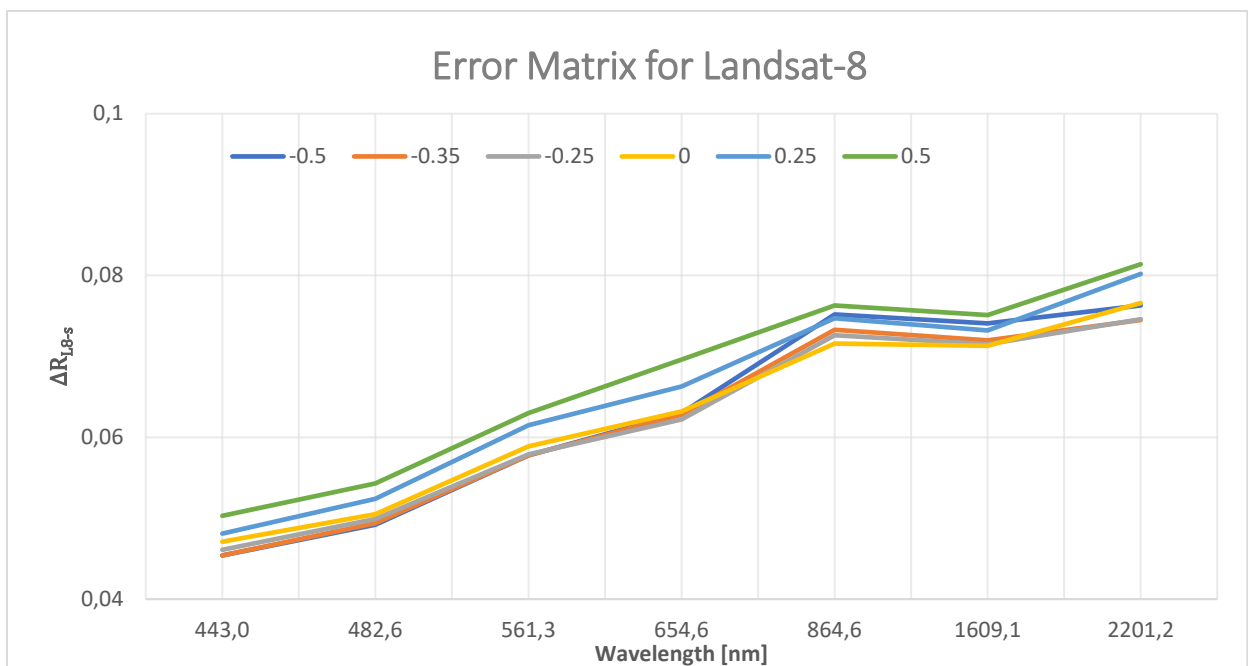

Figure 10: Error Matrix for Landsat-8 Data, graphic plot

Not only are the deviations from the mean values consistently increasing with the wavelength, but also the mean values themselves are closer to $R_{b,r}=1$ for shorter wavelengths. While for Landsat Band 1, they differ in an order of magnitude of ± 0.002, for Band 7 they vary around ± 0.009. Thus, the different parametrizations of haze sigma seem to have a higher error potential in the NIR-SWIR bands than in the visible range. For Sentinel-2, similar tendencies are observed in table 6 with a graphic plot of the same data shown in figure 11. An increasing error is found for each parametrization with increasing wavelengths, horizontally comparing the errors in one line. The same effect in a weaker order of magnitude can be found in the vertical line, looking at one band and different parametrizations. Comparing with the Landsat-8 table, the both the vertical and the horizontal effects are stronger for Sentinel-2 Data.

S2	Band 1 (442.7)	Band 2 (492.4)	Band 3 (559.8)	Band 4 (664.6)	Band 5 (704.1)	Band 6 (740.5)	Band 7 (782.8)	Band 8 (832.8)	Band 9 (864.7)	Band 10 (1613.7)	Band 11 (2202.4)
-0.5	0.9986 ± 0.0527	1.0010 ± 0.0638	0.9969 ± 0.0735	0.9858 ± 0.0776	0.9839 ± 0.0831	0.9807 ± 0.0897	0.9795 ± 0.0912	0.9807 ± 0.0922	0.9795 ± 0.0919	0.9895 ± 0.0916	0.9951 ± 0.0925
-0.35	1.0 ± 0.0513	1.0039 ± 0.0623	1.0025 ± 0.0707	0.9933 ± 0.0733	0.9963 ± 0.0798	0.9976 ± 0.0862	0.9974 ± 0.0876	0.9999 ± 0.0887	0.9988 ± 0.0886	1.0042 ± 0.0885	1.0036 ± 0.0892
-0.25	1.0006 ± 0.0534	1.0054 ± 0.0649	1.0049 ± 0.073	0.9933 ± 0.0753	1.0 ± 0.0814	1.0037 ± 0.0874	1.0039 ± 0.0888	1.0073 ± 0.09	1.0064 ± 0.0897	1.011 ± 0.09	1.0115 ± 0.0911
0	1.0011 ± 0.0553	1.0088 ± 0.0676	1.0096 ± 0.0754	0.9964 ± 0.0777	1.0032 ± 0.0821	1.0063 ± 0.0881	1.0064 ± 0.0896	1.0103 ± 0.0908	1.0079 ± 0.0904	1.0147 ± 0.0913	1.0224 ± 0.0945
0.25	1.0037 ± 0.0615	1.0146 ± 0.0757	1.0155 ± 0.081	0.9984 ± 0.0834	1.0007 ± 0.0848	1.0005 ± 0.0891	1.0018 ± 0.0913	1.0074 ± 0.0921	1.0036 ± 0.0921	0.9989 ± 0.0887	1.0246 ± 0.1002
0.5	1.003 ± 0.0706	1.0168 ± 0.0853	1.0152 ± 0.0878	0.993 ± 0.092	0.9876 ± 0.0924	0.9873 ± 0.0988	0.9877 ± 0.1009	0.9902 ± 0.1012	0.9902 ± 0.1024	1.0097 ± 0.1044	1.015 ± 0.1084

Table 6: Error Matrix per band and haze sigma for S2

The higher deviations from the mean with longer wavelengths can be explained with the before mentioned relation between wavelength and BRDF effects as shown by Roy et al. 2017 due to different scene incidence angles. For example, at 442 nm (blue band for both sensors) Landsat data shows an average error of ±0.045 whereas the Sentinel Data shows one of ± 0.055 with both values being taken from the haze sigma = 0 parametrization. The same tendency for higher error in Sentinel Bands can be observed for the other parametrizations as well. But since the Landsat scenes all have a scene incidence angle of 0 and still present an error increasing in wavelength, respectively smaller than the Sentinel-2 scenes' error, it can only partly explain the behaviour observed here. Another possible reason might be a spectral smoothing effect, which would be stronger in shorter wavelengths as those suffer more from haziness and therefore would produces less variance in mean reflection values.
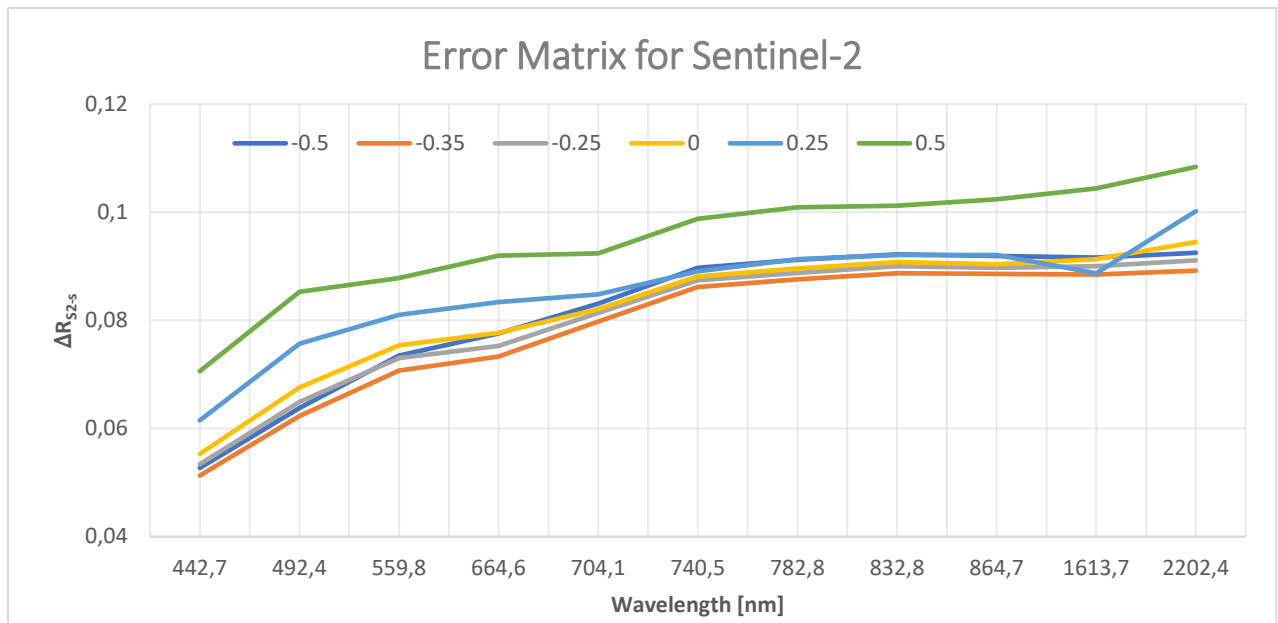
Figure 11: Error Matrix of Sentinel-2 Data, graphic plot

To choose a new haze sigma value, which can be further tested and validated, for each sensor and band, the lowest average error (yellow) and best $R_{b,r}$ value (green) are marked. As seen also in the plots above, a clear conclusive result cannot be obtained. Though most bands show a tendency towards negative values, there are outliers present. For Sentinel-2, the outliers with better $R_{b,r}$ values are the VRE bands, as already described from visually analysing. For Landsat-8, the longer the wavelength is, the better performing positive haze sigma values are. At this point already should be mentioned the low number of tested scenes, which is limiting the studies conclusiveness.

Since the matrixes and the plots' results are pointing into the negative direction, apart from outliers, a scene and run-wise calculation of a weighted R was performed and the haze sigma value, which together with their errors show the minimal deviation from R=1 was filled into the table 7.

According to each scene-specific run of haze sigma values, most show the best concentration around $R_{b,r}$=1 for -0.35. It must be mentioned here though, that for some scenes, -0.35 was not the best choice, but still showed an improvement in terms of the ratio compared to the current hardcoded haze sigma value 0. This trade-off was done to be able to work with one haze sigma in the validation part, since no scene dependency of haze sigma regarding the scene incidence angle or the sun zenith angle were found by comparing the angles from the metadata tables to the results. One reason for the inconclusiveness regarding scene-dependency is the mere number of tested scenes, which is not enough to precisely neglect scene-dependencies. On the other hand, only the incidence angle and solar zenith angle are evaluated and compared with the haze sigma chosen for the correspondent scene, whereas the optimal haze sigma parametrization could also depend on other factors such as land surface coverage or presence of water bodies.

30

Scenes L8 (Continent)	Scene Incidence Angle [deg]	New Haze Sigma (Weighted Ratio ± Error)
Bodensee (C. Europe)	0	-0.35 (0.9956 ± 0.0005)
Alps (C. Europe)	0	-0.35 (0.9929 ± 0.0006)
Berlin (C. Europe)	0	-0.35 (0.9900 ± 0.0009)
Irkutsk (N. Asia)	0	-0.35 (1.0020 ± 0.0005)
Kyiv (E. Europe)	0	-0.35 (0.9931 ± 0.0008)
Sodankyla (N. Europe)	0	-0.35 (0.9956 ± 0.0006)
Timisoara (E. Europe)	0	0.25 (0.9980 ± 0.0021)
Manacapuro (S. America)	0	-0.35 (0.9903 ± 0.0005)
Baeksa (E. Asia)	0	-0.35 (0.9896 ± 0.0007)
Montelimar (W. Europe)	0	-0.35 (0.9848 ± 0.0015)
Wallops (N. America)	0	-0.35 (0.9888 ± 0.0009)
Penang (S. E. Asia)	0	0 (0.9925 ± 0.0021)
Pokhara (C. Asia)	0	-0.35 (0.9923 ± 0.0005)
Strzyzow (E. Europe)	0	-0.35 (0.9930 ± 0.0004)
Mingo (N. America)	0	-0.35 (0.9946 ± 0.0004)

Table 7: Resulting haze sigma per scene for L8

As already mentioned in chapter 2.1, the Landsat-8 data are all acquired with nadir-view, which makes them better comparable than the Sentinel-2 scenes that show varying scene incidence angles.

Scenes S2 (continent)	Scene Incidence Angle [deg]	New Haze Sigma (Weighted Ratio ± Error)
Israel '15 S2A (Asia)	3.33	-0.35 (0.9900 ± 0.0008)
Israel '16 S2A (Asia)	3.33	-0.35 (0.9956 ± 0.0005)
France '16 S2A (Europe)	7.08	-0.35 (0.9929 ± 0.0006)
Appalachian S2A (N. Am.)	2.89	0 (0.9950 ± 0.0021)
Camaguey S2B (C. America) *	7.33	-0.35 (1.0082 ± 0.0027)
Bamenda S2A (C. Africa)	6.28	-0.35 (0.9848 ± 0.0015)
Luang Namtha S2A (S.E. Asia)	8.57	-0.35 (0.9896 ± 0.0007)
Taipei (E. Asia)	2.24	-0.35 (0.9899 ± 0.0005)
Tallahassee (N. Am.)	4.58	-0.35 (0.9956 ± 0.0006)

Yekaterinburg (N. Asia)	8.27	-0.35 (0.9903 ± 0.0006)
Thompson Farm (N. Am.)	8.52	-0.35 (0.9931 ± 0.0008)
Portglenone (N.W. Europe) *	9.25	-0.35 (0.9943 ± 0.0016)
Palgrunden (N. Europe)	6.96	-0.35 (0.9888 ± 0.0009)
Itajuba (S. America) *	1.13	-0.25 (0.9951 ± 0.0007)

Table 8: Resulting haze sigma per scene for S2

As seen in the table 9 below, though most scenes are presenting a good result for haze sigma = -0.35, the differences between the parametrisation options are not significant. Although the value calculated in this table includes all the outliers for each haze sigma run, even lower errors would be expected to produce a similar order of magnitude regarding the different mean weighted ratios. To consider the scenes different error bars, equation 10 (ref. chapter 2.2.1) includes as a weighting factor the inverse of the error calculated per scene and run. This provides that scenes with higher errors, as well as outlying scene runs with higher errors are weighted low in the calculation of $\overline{R_s}$ and its respective error.

Haze Sigma	$\overline{R_s}$ $\pm \Delta\overline{R_s}$
-0.5	0.9991 ± 0.0211 [0.978: 1.0202]
-0.35	1.0013 ± 0.0202 [0.9811: 1.0215]
-0.25	1.0031 ± 0.0228 [0.9803: 1.0259]
0	1.0054 ± 0.0216 [0.9838: 1.027]
0.25	1.0053 ± 0.0156 [0.9897: 1.0209]
0.5	1.0081 ± 0.0211 [0.987: 1.0292]

Table 9: Weighted Ratio and Error per haze sigma over all scenes

From this table it becomes obvious that the order of magnitude in changes due to haze sigma variation is quite low, far from statistically significant.

3.2 Inter-Sensor Validation

For the inter sensor validation, a scene in the north of Berlin is chosen. Both scenes were processed with haze sigma = -0.35. As stated in the metadata table in chapter 2.2.2, the scenes have an overpassing time difference of approximately one minute. A different window size for obtaining the spectra was chosen, since both sensors have different ground sampling distances. Figure 12 shows the L1C input for PACO on the left and the fully atmospheric corrected BOA on the right for DESIS, whereas

Spectral Profile
■ location

Figure 12: DESIS Berlin L1C input (left), dehazed TOA (middle and BOA (right)

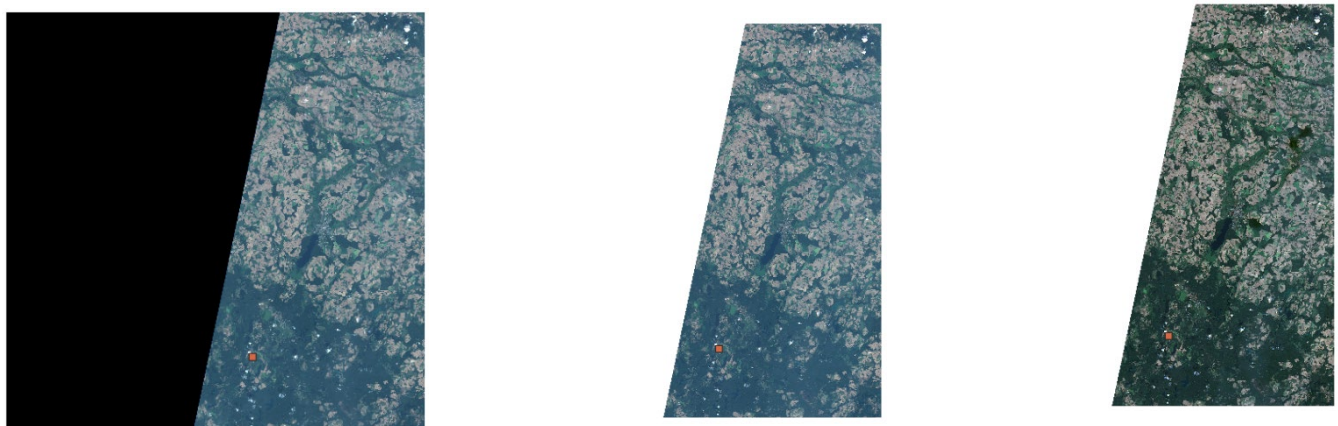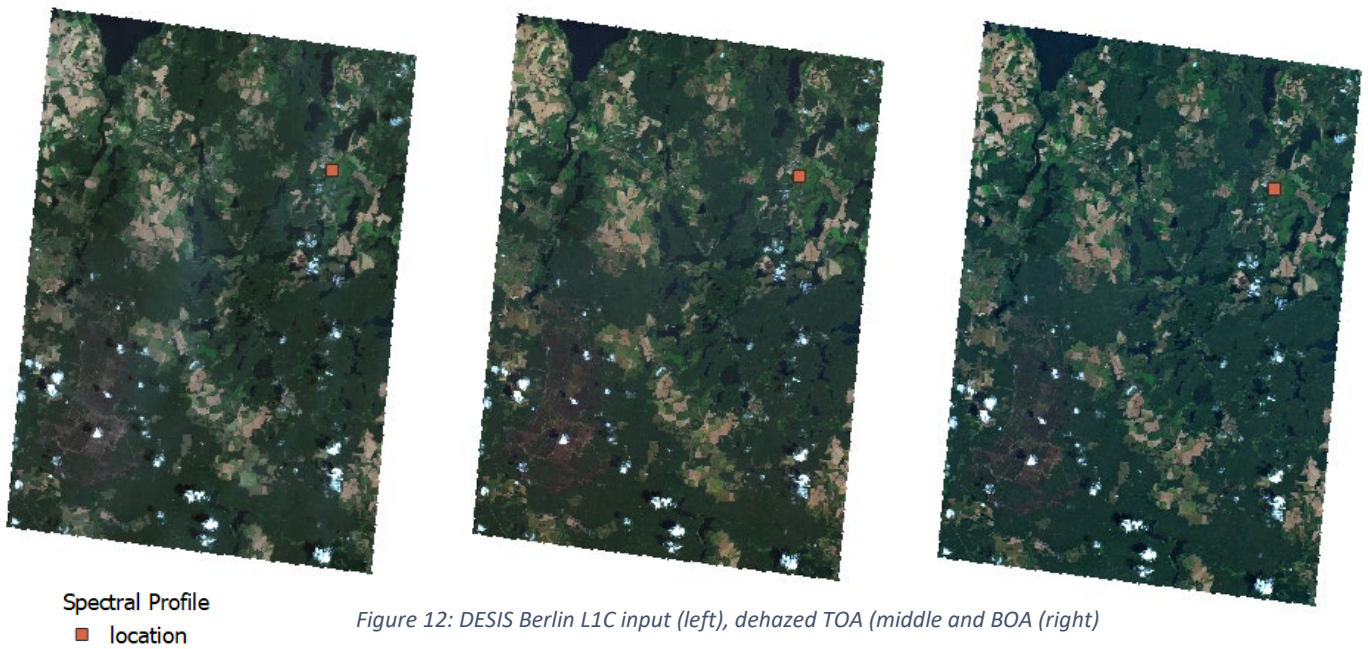*Figure 13: Sentinel-2 Berlin L1C (left), dehazed TOA (middle) and BOA (right)*

figure 13 shows the same combination for Sentinel-2. Figure 12 only shows visible haze in the center of the image. Over an area of brownish fields, a slight effect of the dehazing can be observed. In figure 13, the Sentinel-2 image containing overlaps with the DESIS image presents a more even haze distribution.

In figure 14 the BOA reflectance profiles of both sensors are compared over a hazy location chosen to be as homogeneous as possible. While for the visible range, the reflectance values of both scenes show almost overlapping curves, the DESIS scene has increasingly higher reflectance from around 700 nm.
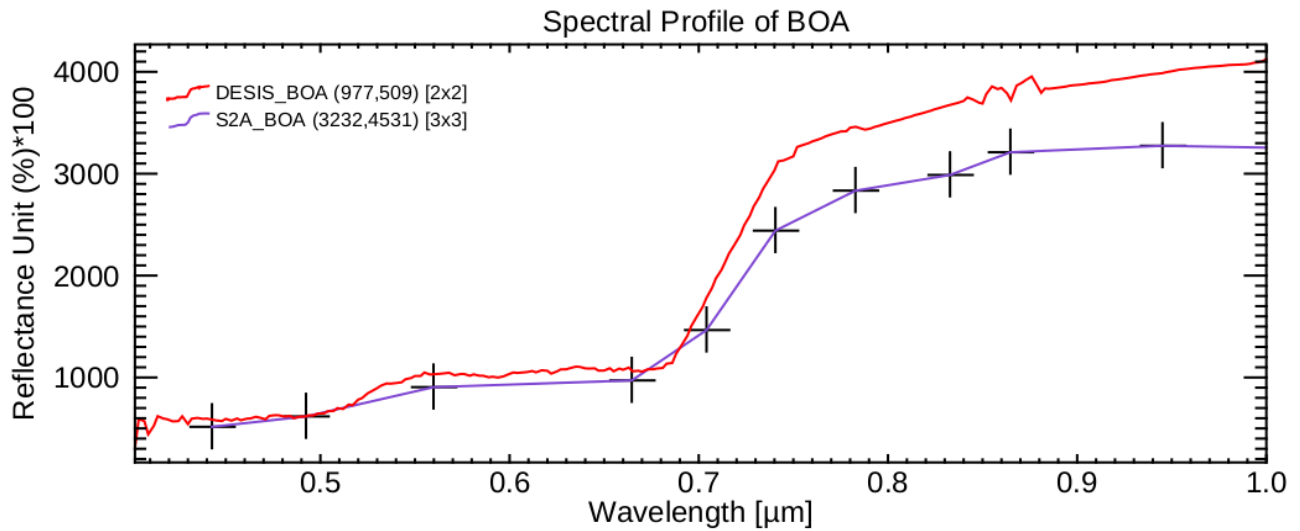

Figure 12: Spectral Profile of BOA for Hazy Area in Berlin

The variables shown in table 10, estimated for each scene during PACOs processing, may be possible causes for the difference in BOAs. The WV column is higher though for the Sentinel-2 scene as for the DESIS scene. Neither the AOT values can explain the difference in BOA reflectance for the NIR range as for a higher AOT, a lower reflectance value would be assumed. Also, since the plot shown here presents fully atmospheric corrected scenes, the aerosol effects should not be so visible anymore as well as one would expect a decrease in their influence with increasing wavelength due to scatter function properties (ref. chapter 1). The same goes for the water vapor absorption, which after being estimated from the correspondent bands should be corrected.

Since the DESIS scene has an incidence angle almost three times higher than the Sentinel-2 scene, a possible explanation could be the BRDF influence as described in the previous section for longer wavelengths.

For Location: 764585.8411, 5909376.5543 (UTM 32 N)	AOT @ 550 nm)	WV [cm]	Scene Incidence Angle [deg]
DESIS	0.224	1.346	29.90
S2	0.088	1.569	9.83

Table 10: AOT, WV and Scene Incidence for Berlin Validation scenes

3.3 Temporal Validation

As described in Chapter 2.2.3, the scene combination for Montelimar, France was used in a previous study to evaluate the performance of the dehazing algorithm of ATCOR (Makarau et al. 2014). Therefore, the characterisation of the reference scene as clear is assumed as correct. Since both have the same ground resolution, the same window size must be applied.
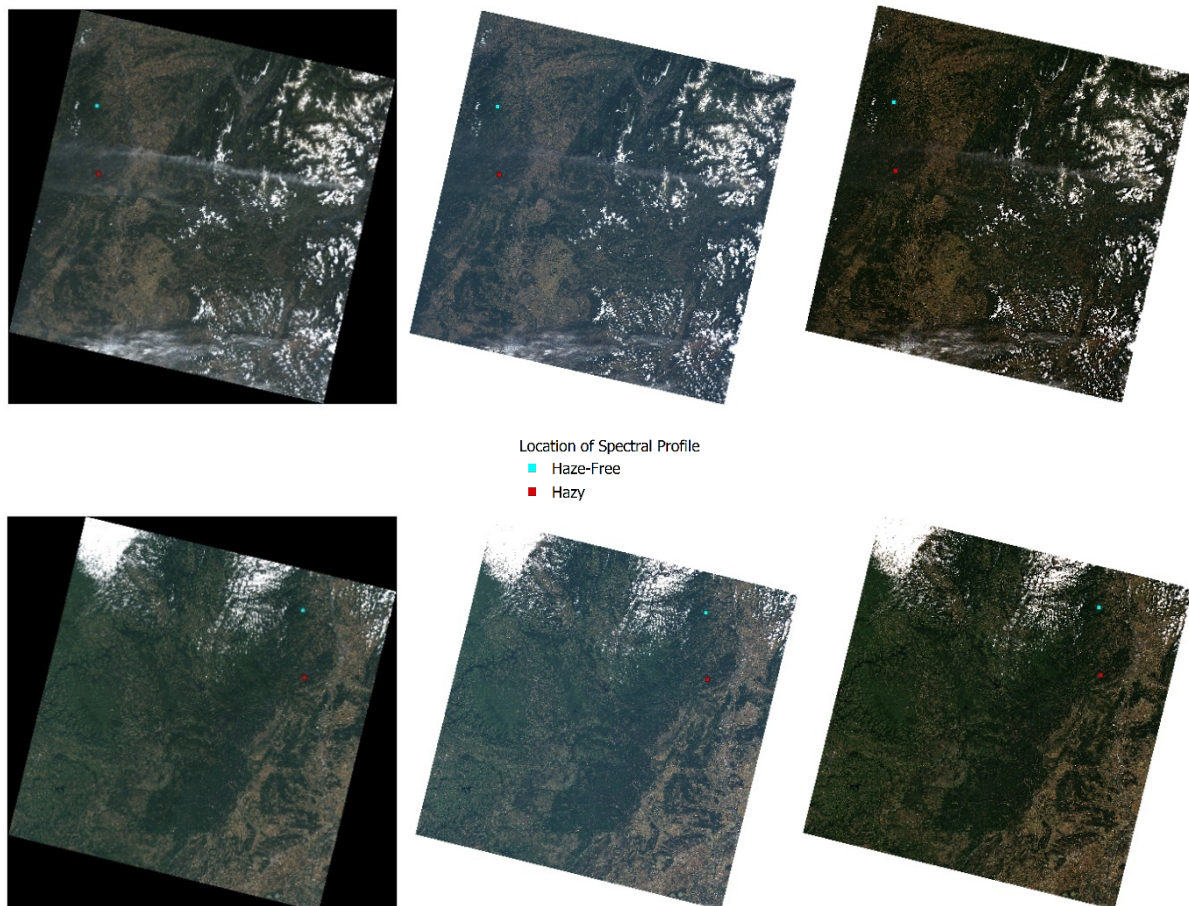

Figure 13: Montelimar: First Row: 14.07.13 L1C (left), dehazed TOA (middle) and BOA (right); Second Row: 05.07.13 Reference Image, L1C(left), dehazed TOA (middle) and BOA (right)

Figure 15 shows the hazy Montelimar scene of the 14th of July 2013 in the upper row as well as the clear reference image from 05.07.13 in the lower row. On the very left, the L1C input is displayed, the middle image shows the dehazed toa and the image on the right the BOA reflectance output.

In figure 16, the L1C Radiances of both scenes are compared with the dehazed TOA ('-z', describes the configuration option of PACO only to dehaze) of the hazy scene. First, a pixel window of size 10x10
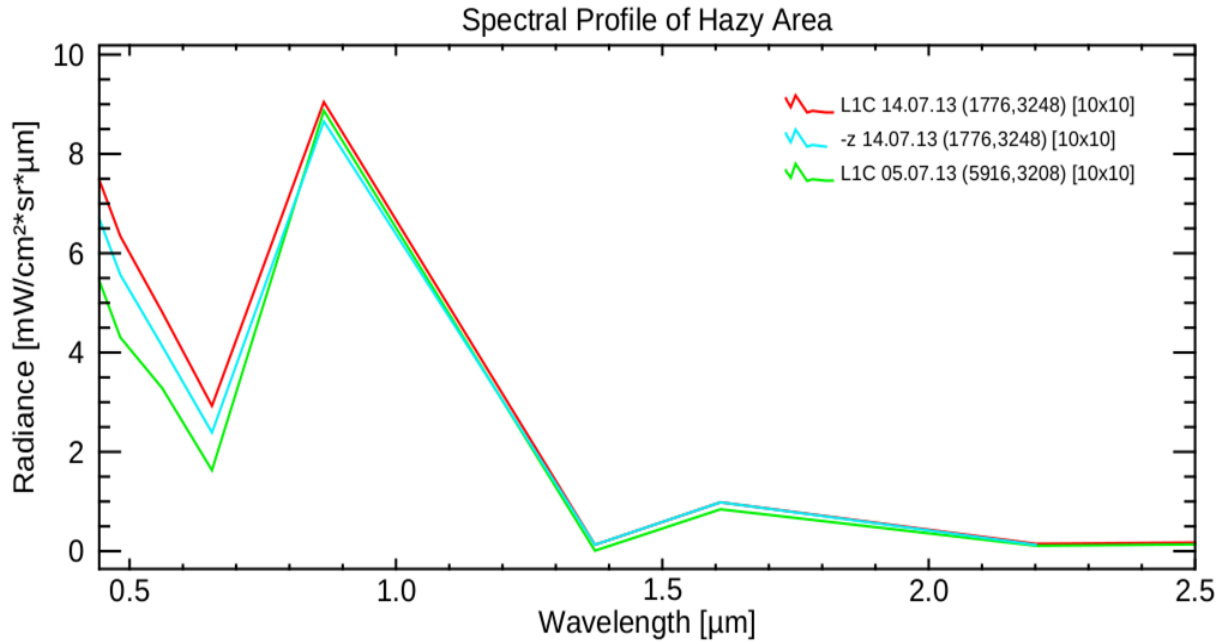

Figure 14: Spectral Profile for Hazy Area in Montelimar

pixels inside a homogeneous area flagged as hazy in the haze mask of the 14th July scene is chosen. The pixel coordinates in the legend represent the first pixel chosen before creating the window. In this case, it is not the exact center pixel of the window (the center is between pixels), but it is right next to the center. In the visible range, the TOA of both scenes are diverging while matching in the NIR-SWIR range. The dehazed TOA seems to be spectrally consistent in terms of shape. Residual atmospheric rayleigh-scattering seems to produce the difference in the dehazed TOA of the hazy image compared to the reference one, since the mismatching is reduced in longer wavelengths, which is characteristic for this kind of scattering (ref. chapter 1). Conducting the full AC though, it is observed that the BOA reflectance of both scenes is not the same as seen in figure 17. Although one would assume that the surface is not changing much within 9 days, this cannot be guaranteed without field measurements. The difference in the BOA also, contradictory to the TOA difference shown in figure 16, is higher in the NIR-SWIR range. Since in the case of Landsat-8 all scenes have the same incidence angle, the different BOA reflections are most probably due to a change in land cover which increases the reflectance for the 14th of July.
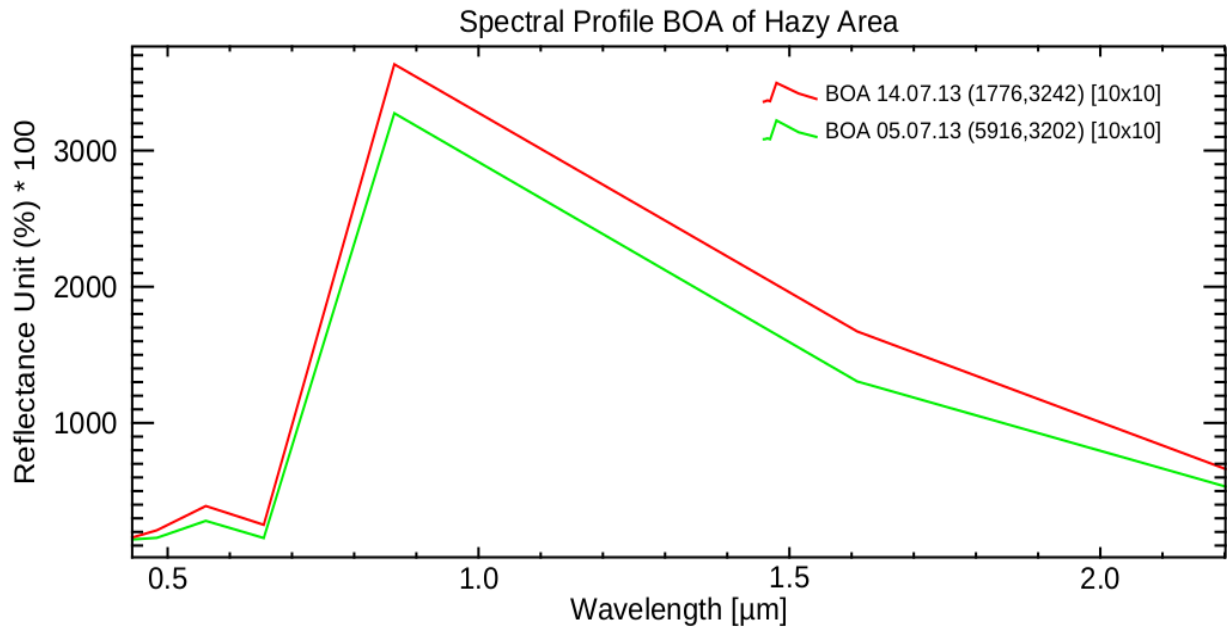
Figure 15: Spectral Profile of BOA for hazy area in Montelimar

For a haze-free location, the same plots are shown in figure 18. The L1C of the 14th of July has a slightly lower radiance value in the 400-650 nm range than for the hazy location but is spectrally consistent with the dehazed TOA for the same day. Since no dehazing is performed, the only difference shown is a lower radiance for the L1C of the 5th of July reference. Most probably, the difference is caused by residual haze that was not flagged as such due to the parametrization.
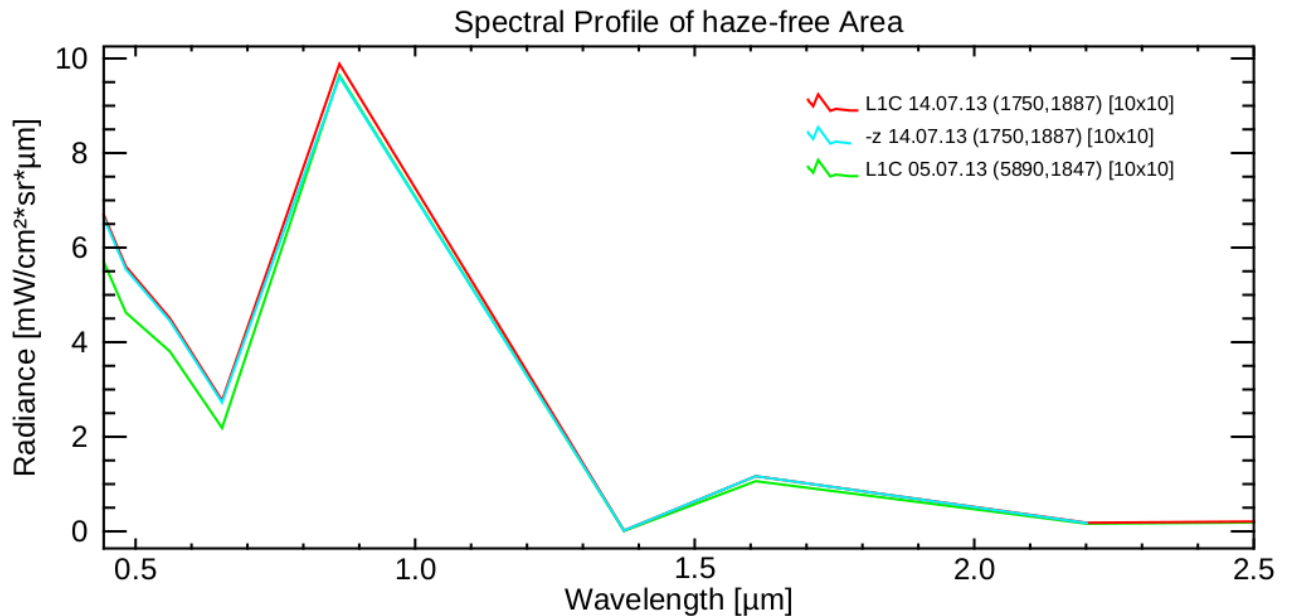

Figure 16: Spectral Profile for haze-free area in Montelimar

The BOA reflectance of the haze-free location is shown in figure 19 below. It is consistent in shape with the reflectance found in the hazy location. Since aerosol absorption and scattering effects that cause

haziness in images can be neglected at longer wavelengths (Makarau et al. 2014, 5898), the difference in the BOA is most likely due to a change in land cover.
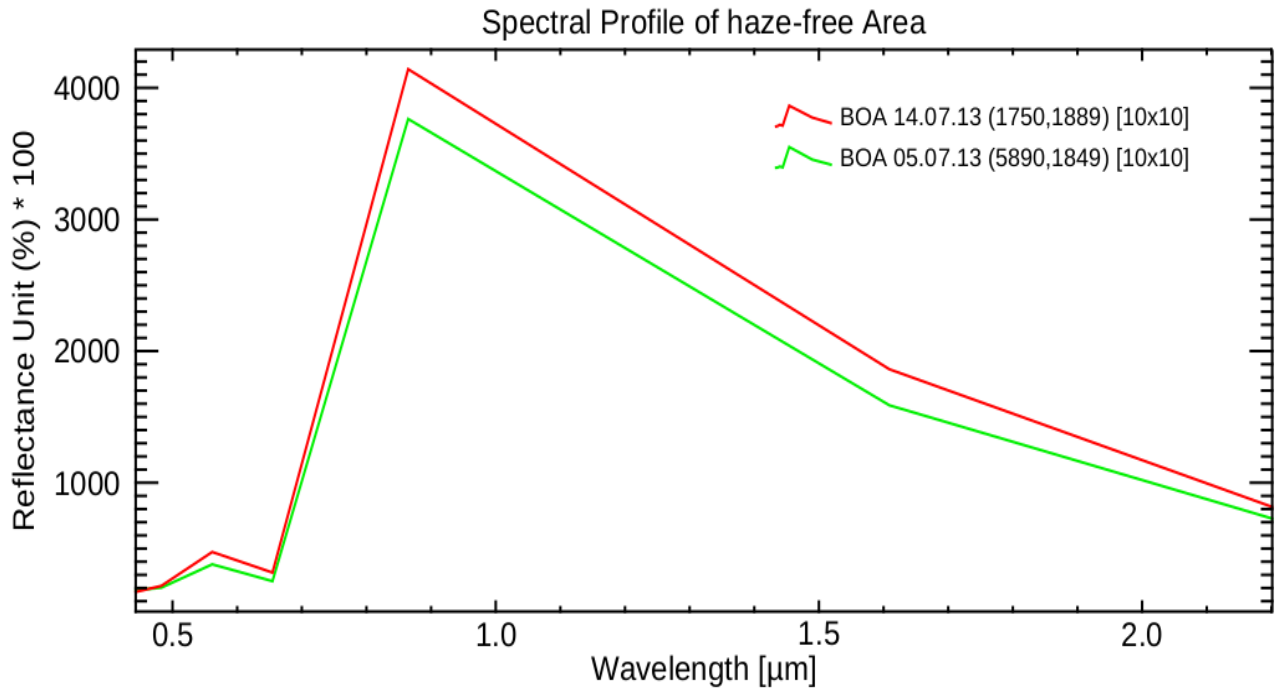

Figure 17: Spectral Profile of BOA for haze-free area in Montelimar

For Hazy Location: 625950, 4961070 (UTM 31N)	AOT (@ 550 nm)
05.07.13	0.066
14.07.13	0.337
For Haze-Free Location: 625198, 5001238 (UTM 31N)	
05.07.13	0.066
14.07.13	0.353

Table 11: AOT Values for Montelimar datasets

The AOT values presented in table 11 show an AOT more than five times higher for the 14th of July than for the reference scene on the 5th of July. A higher AOT would lead to the estimation of a lower BOA reflectance value, which is not the case in either the hazy or haze-free area.
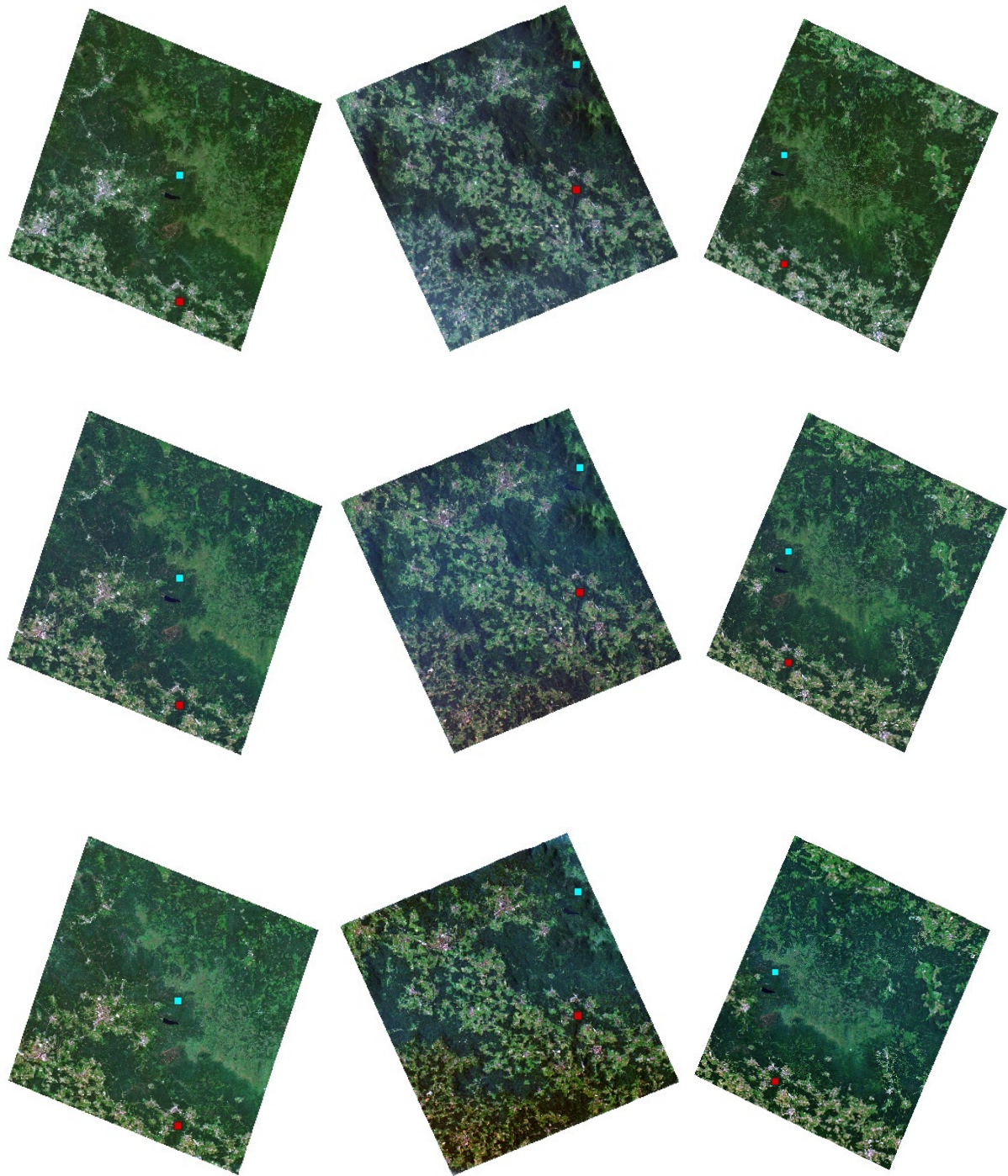
Location of Spectral Profile
- ■ Haze-Free
- ■ Hazy

Figure 18: Line One: L1C inputs of 26th (left), 27th (middle) and 29th (right), Line Two: Dehazed TOA of 26th (left), 27th (middle) and 29th (right), Line Three: BOA of 26th (left), 27th (middle) and 29th (right)

The Bavarian Forest data set consists of three scenes, which all apparently contain haze, though only the scene 27.06.19 shows haze directly visible to the observer by eye as seen in figure 20 for the middle column. As shown in table 12, the scene incidence angles are not consistent, they increase from 4° to 22° with time.

Scene	Scene Incidence Angle [deg]	Solar Zenith Angle [deg]
26.06.19	4.24	31.651
27.06.19	15.89	69.1
29.06.19	22.19	38.536

Table 12: Scene Incidence and Solar Zenith Angles of Bavarian Forest datasets

The spectral profile below in figure 21 shows the L1C and dehazed TOA radiances for the three scenes. With the parametrisation of haze sigma set to -0.35, all of them seem to have haze present and all of them show corrected TOA dehazed spectra that are consistent in shape.
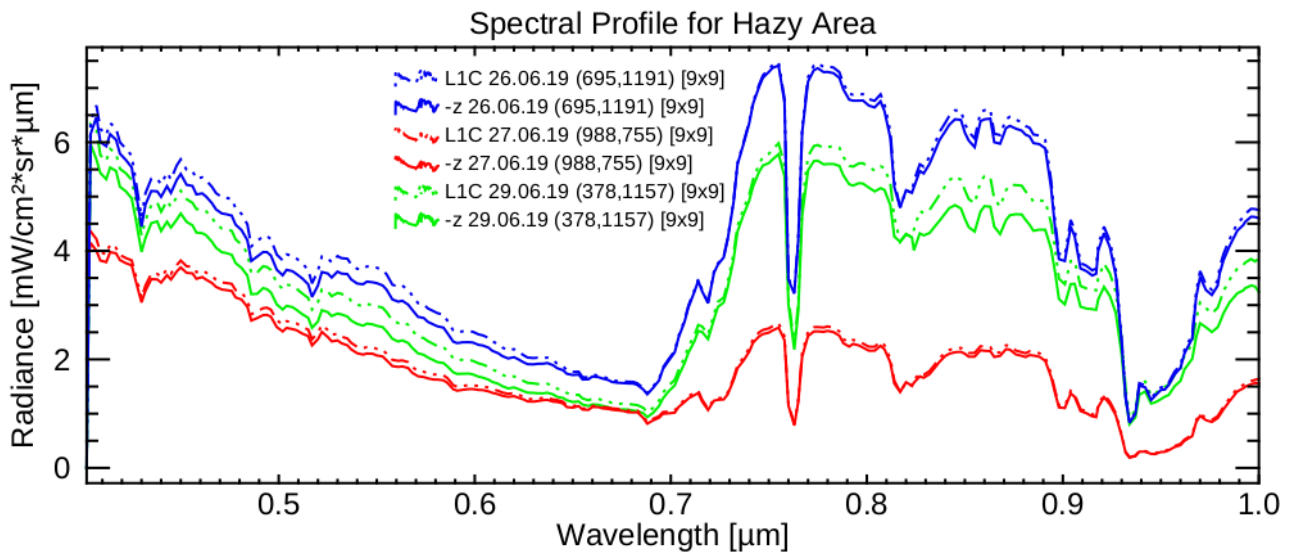

Figure 19: Spectral Profile for Hazy Area in Bavarian Forest

As shown in figure 22, the BOAs retrieved by PACO are also diverging with an increasing wavelength. This could be due to the different scene incidence angles and BRDF effects that change the apparent surface reflection. If that were the case, one would expect a direct correlation between the increasing incidence angles and the increasing BOA reflectance. The scene of the 26th shows the highest reflectance values with the lowest incidence angle respectively, the 27th has the lowest surface reflectance while regarding the scene incidence angle, it ranges in between the 26th and 29th. The latter has the highest incidence angle, while reflectance-wise showing a lower reflectance curve than the 26th's. Another possibility would be a change in surface coverage, but due to the high temporal resolution and the date of acquisition during summer, this would not be expected.
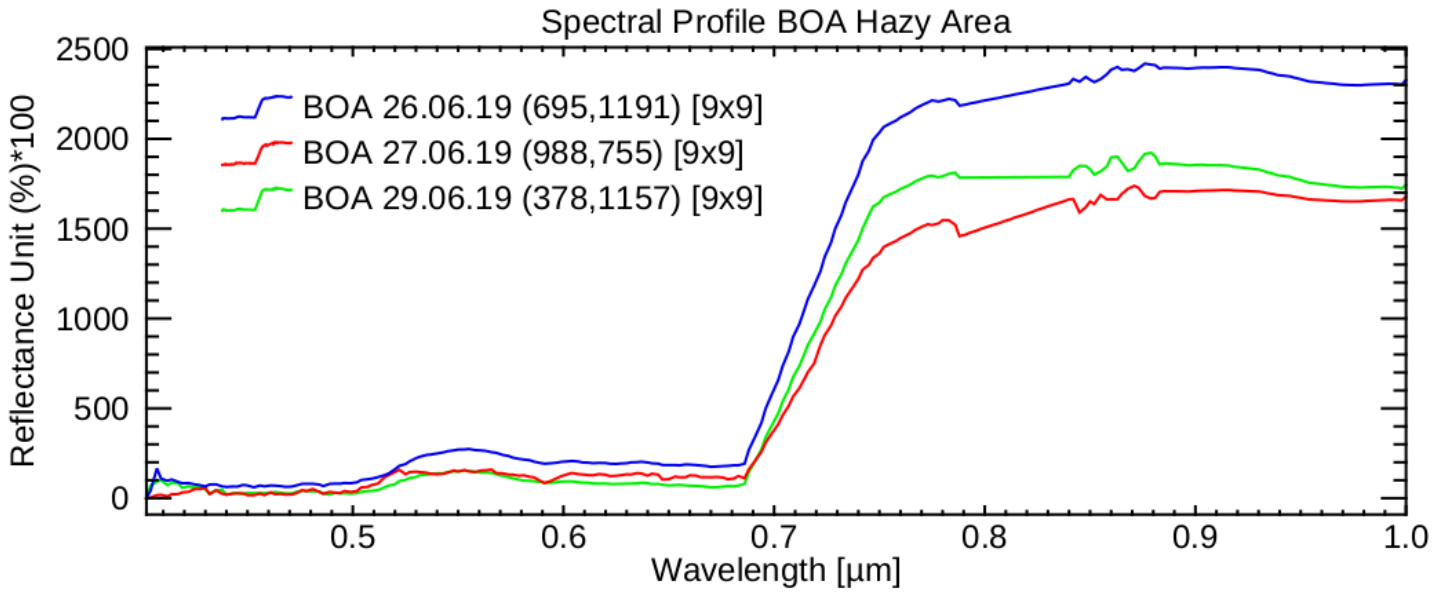
Figure 20: Spectral Profile of BOA for Hazy Area in Bavarian Forest

For the haze-free location, no difference in L1C and dehazed TOA is expected, which is consistent with the data shown in figure 23. The BOAs plotted below in figure 24 only marginally diverge from the ones in the hazy area. But since they are obviously not the same for each day, although a relative stability of reflectance would be expected, the overall conclusiveness for this validation dataset is questionable. Neither the incidence angle nor the solar zenith angle differences explain the difference in BOA reflectance.
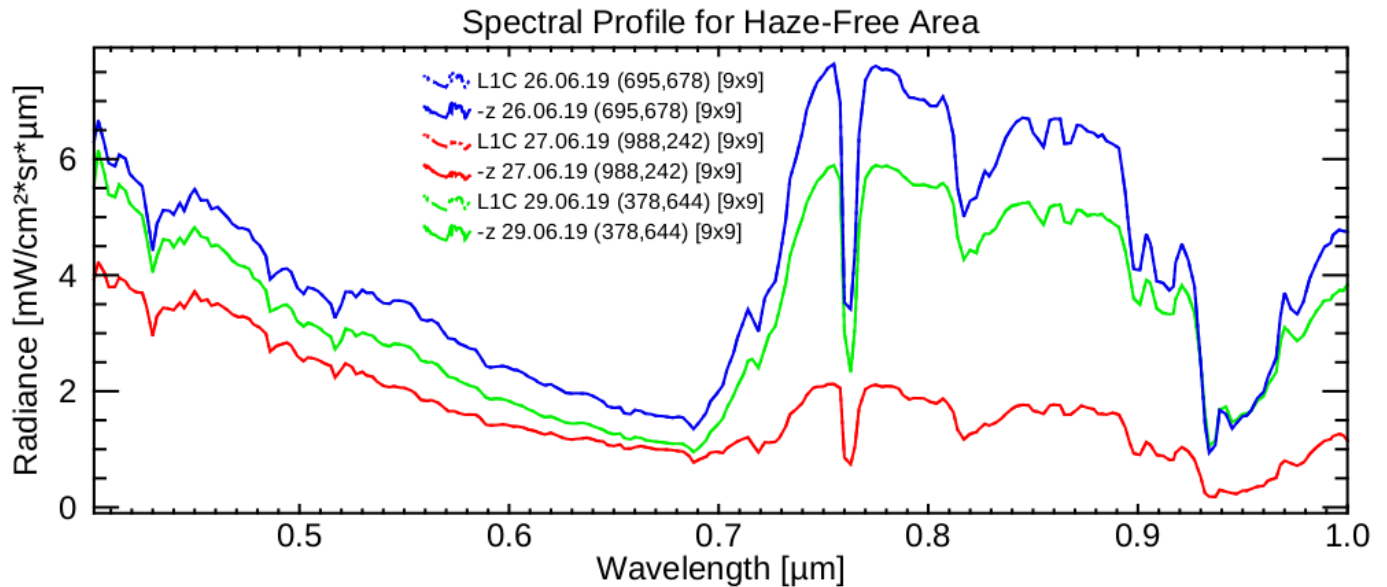

Figure 21: Spectral Profile for Haze-Free Area in Bavarian Forest

For Hazy Location: 379995, 5417265 (UTM 33N)	AOT (@ 550 nm)	WV [cm]
26.06.19	0.147	2.031
27.06.19	0.211	2.110
29.06.19	0.060	0.665
For Haze-Free Location: 379995, 5432655 (UTM 33N)		
26.06.2019	0.137	2.030
27.06.2019	0.201	2.357
29.06.2019	0.057	1.268

Table 13: AOT and WV values for Bavarian Forest scenes

Looking at the AOT and WV column height summarized in table 13 below, no direct or inverse proportionality can be detected. Those values show a small variance from the hazy to the clear area of the scene tested but stay consistent in relative amounts to each other except for the 29th of June, where the water vapor column is around twice as high for the haze-free area than for the hazy one.
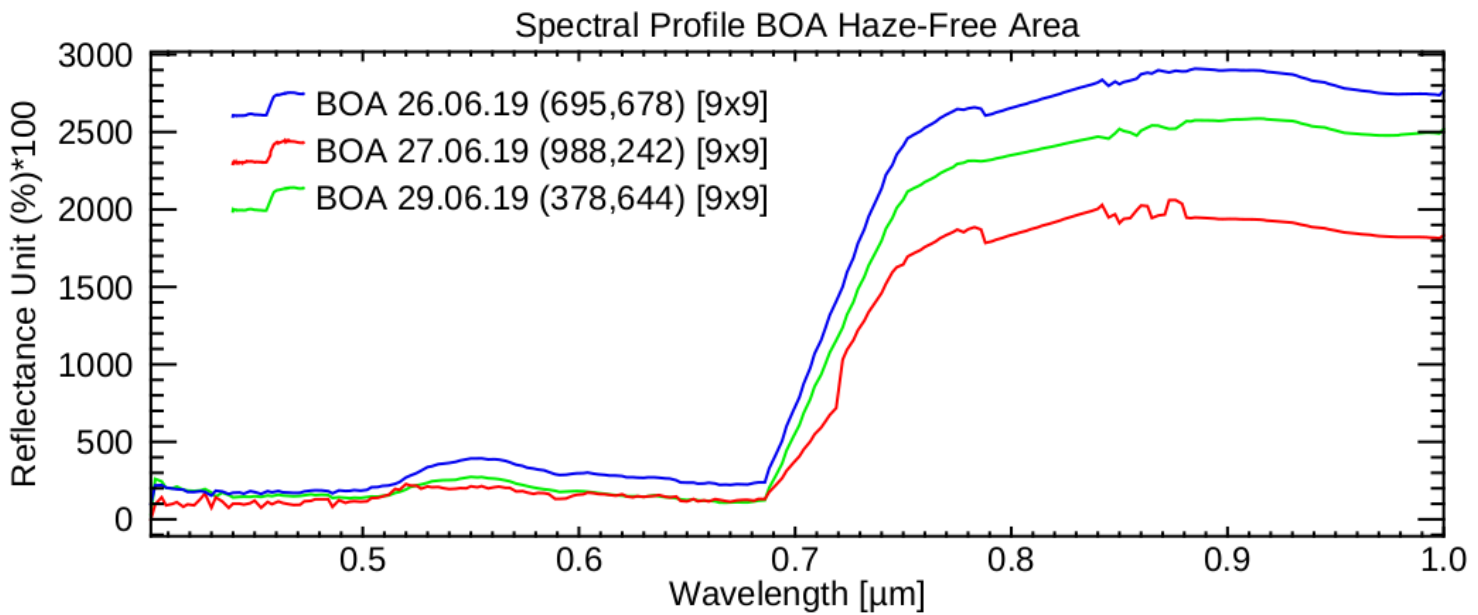

Figure 22: Spectral Profile of BOA for Haze-Free Area in Bavarian Forest scenes

4. Discussion

4.1 Advantages and disadvantages of the presented method

The presented DDV-Patch Approach consists in a relatively easy to implement method, since it requires no external measurements be conducted for validation. The validation can be carried out without ground measurements if a clear scene counterpart is available. The problematic of a clear scene counterpart is obvious though, since it cannot be guaranteed that there is no haze present from a visual assessment as well as identical or at least consistent ground reflectance cannot be assured if no spectra measurements are done.

Several design improvements should be conducted to improve the method. Although the patches have the same minimum size conditions, the size of each patch should be included as a weighting factor in the calculation of the ratio. Furthermore, to test more scenes, the analysis-scripts should be automatized regarding the scenes. At the moment, each scene has its individual jupyter-script to retrieve the patches and the statistical values which increases the time needed to analyse.

A general disadvantage of this approach is its limitation to scenes with an availability of DDV pixels. This limits the scene pool to certain land covers. It was found to be necessary to have at least 5% of DDVs to get a representative number of valid patches. Also, the spatial distribution dependency of the haze regarding the patch availability is limiting the usability of this method. For a scene to be providing enough patches, as described in chapter 3.1 (ref. figure 8), haze should be distributed equally along the DDV patches. Since haze is not always evenly distributed and even in a scene with enough DDVs, can be found not to match the DDV locations, the criteria for valid patches can be seen as limitations regarding a diverse population in terms of land cover.

4.2 Evaluation of plausibility

The most important factor to be discussed here, is the significance of the results of the DDV-Approach. All differences found are in a very small order of magnitude. The mean weighted ratios over all scenes range from 0.987 to 1.029 for different haze sigma parametrisations, an average 2% difference in the results. Together with the rather low number of tested scenes, the significance of the resulting ratio differences should not be overestimated. Although the total number of patches from all scenes considered in the statistics is an average of >4000 individual patches per run, the deviation of the total count of valid patches in different runs (ref. chapter 3.1) is up to ± 50% of the mean valid patch number.

The initial recommendation from the first part of the study to change the static haze sigma value in PACO code to -0.35 cannot be completely verified. Though presenting some consistent results in each part of the validation, residual effects leading to inaccuracies in the dehazed TOAs and BOAs were not fully identified. Further validation methods should be evaluated, as described in chapter 5.1.

5. Conclusion

An initial, experimental study part to fine-tune the haze detection threshold was carried out, using the processors own products, partly the DDV mask. A new approach to test dehazing accuracy was developed and tested with 29 satellite images. The initially anticipated scene-dependency of the haze threshold could not be identified within the first part of the study, which is the main reason, why no dynamic threshold could be implemented. Instead, a new static haze sigma value for the threshold was chosen for PACO with the requirement to show improvement compared with the old static haze sigma. A validation of the new haze sigma was added to the study, containing four datasets. First, an inter-sensor validation with DESIS and Sentinel-2 data was done, afterwards two temporal validations were made, one with DESIS and one with Landsat-8 data. The validation results do not affirm a general dehazing improvement due to the change of haze sigma only. A full identification of the residual effects found to cause the differences in BOA reflectance was not achieved. Partly, the new static haze sigma showed good performance but due to the uncertainty regarding the residual spectral differences, it is recommended to leave the static haze sigma at the current value of zero.

5.1 Recommendations for future studies

Further evaluation should dig deeper into the influence of the scene incidence angle. In this study, a scene dependency regarding the solar zenith or incidence angle could not be detected but there might be other scene-related parameters to evaluate, such as land cover, presence of water bodies or different aerosol types. To evaluate the effects of a dynamic haze sigma inside the dehaze module, an adjusted output analysis method should be carried out while at the same time, intermediate results from the programs code itself should be included to assess the performance of the dehazing module. Therefore, other kinds of land cover should be included, which requires at least one different method in addition to the DDV-Approach to validate the dehazing results. One could use the AERONET Stations network as done by the PACO developers' group (de los Reyes et al. 2020, 8ff) and measure the accuracy of dehazed results with different parametrisations of haze sigma with external measurements, including ground measurements of BOA spectra for a viable assessment of the atmospheric conditions.

6. Literature

(1) Alonso, K.; Bachmann, M; Burch, K.; Burch, K.; Carmona, E.; Cerra, D.; de los Reyes, R.; Dietrich, D.; Heiden, U.; Hölderlin, A.; Ickes, J.; Knodt, U.; Krutz, D.; Lester, H.; Müller, R.; Pagnutti, M.; Reinartz, P.; Richter, R.; Ryan, R.; Sebastian, I.; Tegler, M. (2019): *Data Products, Quality and Validation of the DLR Earth Sensing Imaging Spectrometer (DESIS).* In: Sensors 2019, 19, 4471.

(2) Ancuti, C.O.; Ancuti, C. (2013): Single Image Dehazing by Multi-Scale Fusion. In: IEEE Transactions on Image Processing, 22 (8), 3271-3282.

(3) Berk, A.; Conforti, P.; Kennett, R.; Perkins, T.; Hawes, F.; van de Bosch, J. (2014): MODTRAN6: A major upgrade of the MODTRAN radiative transfer code. In: SPIE Proceedings, 9088, Algorithms and Technologies for Multispectral, Hyperspectral and Ultraspectral.

(4) Chavez, P. (1988): *An Improved Dark-Object-Subtraction Technique For Atmospherical Scattering Correction of Multispectral Data.* In: Remote Sensing Environment, 24/3, 459-479.

(5) de los Reyes, R.; Langheinrich, M.; Schwind, P.; Richter, R.; Pflug, B.; Bachmann, M.; Müller, R.; Carmona, E.; Zekoll, V.; Reinartz, P. (2020): *PACO: Python-Based Atmospheric Correction.* In: Sensors 2020, 20, 1428.

(6) Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; Meygret, A.; Spoto, F., Sy, O.; Marchese, F.; Bargellini, P. (2012*): Sentinel-2: ESA's Optical High Resolution Mission for GMES Operational Services.* In: Remote Sensing of Environment, 120, 25-36.

(7) ESA (European Space Agency): Sentinel-2 Data Products. URL: https://sentinel.esa.int/web/sentinel/missions/sentinel-2/data-products (obtained: 24.09.2020).

(8) Gavilán, V.; Lillo-Saavedra, M; Holzapfel, E.; Rivera, D; García-Pedrero, A. (2019): *Seasonal Crop Water Balance Using Harmonized Landsat-8 and Sentinel-2 Time Series Data.* In: Water, 11, 2236.

(9) Guanter, L.; Kaufmann, H.; Foerster, S.; Brosinsky, A.; Wulf, H.; Bochow, M., Boesche, N.; Brell, M.; Buddenbaum, H.; Chabrillat, S.; Hank, T.; Heiden, U.; Heim, B.; Heldens, W.; Hill, J.; Hollstein, A.; Hostert, P.; Krasemann, H.; Leitão, P.J.; van der Linden, S.; Mauser, W.; Mielke, C.; Müller, A.; Oppelt, N.; Roessner, S.; Röttgers, R.; Schneiderhan, T.; Staenz, K.; Segl, K. (2016): EnMAP Technical Reports: EnMAP Science Plan.

(10) Ientilucci, E.J.; Adler-Golden, S. (2019): *Atmospheric Compensation of Hyperspectral Data: An Overview and Review of In-Scene and Physics-Based Approaches.* In: IEEE Geosci. Remote Sens. Mag, 7, 31-50.

(11) Makarau, A.; Richter, R.; Schläpfer, D.; Reinartz, P.; (2016): *Combined Haze and Cirrus Removal for Multispectral Imagery.* In: IEEE Geoscience and Remote Sensing Letters, 13/3.

(12) Makarau, A.; Richter, R.; Müller, R.; Reinartz, P. (2014): *Haze Detection and Removal in Remotely Sensed Multispectral Imagery.* In: IEEE Transactions Geoscience and Remote Sensing, 52/9.

(13) Richter, R.; Schläpfer, D. (2019): *ATCOR Theoretical Background, Version 1.0.* ReSe Applications, Switzerland.

(14) Roy, D.P.; Li, J.; Zhang, H.K.; Huang, H.; Li, Z. (2017): *Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance*. In: Remote Sensing of Environment, 199, 25-38.

(15) Roy, D.P; Wulder, M.A; Loveland, T.R.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Helder, D.; Irons, J.R.; Johnson, D.M.; Kennedy, R.; Scambos, T.A.; Schaaf, C.B., Schott, J.R.; Sheng, Y.; Vermote, E.F.; Belward, A.S.; Bindschadler, R.; Cohen, W.B.; Gao, F.;

Hipple, J.D.; Hostert, P.; Huntington, J.; Justice, C.O.; Kilic, A.; Kovalskyy, V.; Lee, Z.P.; Lymburner, L.; Masek, J.G.; McCorkel, J.; Shuai, Y.; Trezza, R.; Vogelmann, J.; Wynne, R.H.; Zhu, Z. (2014): *Landsat-8: Science and product vision for terrestrial global change research.* In: Remote Sensing of Environment, 145, 154-172.

(16) Taylor, R.J. (1988): Fehleranalyse: eine Einführung in die Untersuchung von Unsicherheiten in physikalischen Messungen. VCH.

(17) Walpole, R.E.; Myers, R.H.; Myers, S.L.; Ye, K. (2012): Probability & Statistics for Engineers and Scientists – 9th Edition. Pearson.

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources /resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Date, Place   …………………………………………………………………………………..

Signature   ………………………………………………………………………………….