Application and Comparison of Inter-Calibration Methods for Satellite Microwave Humidity Sounders

Masterthesis of Marc Prange

University of Hamburg
Department Geowissenschaften

First supervisor: Prof. Dr. Stefan Bühler
Second supervisor: Dr. Martin Burgdorf

Hamburg, February 20, 2019

Masterarbeit im Fachbereich Geowissenschaften der Universität Hamburg
Thema der Arbeit: „Application and Comparison of Inter-Calibration Methods for Satellite Microwave Humidity Sounders“

Abstract

Four Inter-Calibration (I-C) methods to assess the calibration state of satellite based microwave (MW) humidity sounders (MHS and AMSU-B) are first defined and then applied to multiple operational satellite datasets. The I-C methods are designed to negate geophysically induced biases between the satellite datasets. Such biases can be caused by systematically sampling different phases of the diurnal cycle due to different sun-synchronous orbits of the satellites. The aim of this work is to assess the performance of the investigated I-C methods by comparing long timeseries of the obtained inter-satellite biases with each method against each other. The first I-C method is based on the zonally averaged satellite data between 45° N/S and filtering of the largest mountainous ranges. This data subset is found to be insignificantly affected by the diurnal cycle in the sounding channels of the MW instruments. The second I-C method is based on observation feedback from the ERA-Interim reanalysis. By using the reanalysis model background radiances, which are simulated for each assimilated MW instrument, as a reference for the observational satellite datasets (O-B), geophysically induced biases should ideally be negated. The third I-C method is based on Simultaneous All Angle Collocations (SAACs), which occur when two satellite instruments measure at the same location (5 km radius) at the same time (5 minute temporal window). Hence, no systematic geophysically induced inter-satellite biases should be present in SAACs. The fourth I-C method is designed to circumvent a known weakness of SAACs, namely that SAACs generally only occur at polar latitudes. By relaxing the temporal collocation criterion to 8 hours and introducing the geostationary SEVIRI (Spinning Enhanced Visible and Infrared Imager) dataset intermediately to find scenes that remain constant over several hours, so called Opportunistic Constant Target Matchups (OCTMs) are found.

It is found that the NCT, O-B and OCTM methods do not sufficiently negate the diurnal cycle at the surface to be applicable for the two surface channels of the MW instruments. For the NCT method, the inclusion of land data introduces a strong diurnal cycle signal into the NCT data of the surface channels. For the O-B method, the reanalysis model background is proven to overestimate the diurnal cycle at the surface. Owing to their stringent collocation criteria, only SAACs do not show diurnal cycle related inter-satellite biases in any channel. For the three humidity sounding channels of the MW instruments, all methods perform sufficiently well to catch significant inter-satellite biases. Slight deficiencies are found in the lower tropospheric sounding channel with the NCT and OCTM methods, which is caused by this channel receiving some signal from the surface. Also, slight deficincies are found in the upper tropospheric sounding channel with the O-B method, caused by an overestimation of the diurnal cycle by the model background.

It is analysed, whether BT dependencies of inter-satellite biases lead to deviating biases of the SAAC method, which is restricted to cold Brightness Temperatures (BTs). It is found that instrument noise and increased natural variabilty in one satellite dataset can induce BT correlations of the inter-satellite biases. However, no direct dependence of inter-satellite biases on the BT is identified. To avoid questioning the validity of SAACs due to their limited BT sampling in the future, it is recommended to setup a joint collocation data basis of cold SAACs and warm OCTMs. In the sounding channels SAACs and OCTMs exhibit very similar random uncertainties, already, and the OCTM methodology is believed to be easily expandable to perform well for the surface channels, too. It is concluded, that the investigated I-C methods are a solid basis for future assessments of new or reprocessed satellite datasets.

Contents

List of Figures

List of Tables

1 Introduction

Monitoring the global development of humidity in different layers of the troposphere on climatological timescales is essential to understand the climatological trend of our climate system. This is because of the many energetically important interactions water molecules have within the climate system, for example through phase changes, through the interaction with Infrared (IR) radiation emitted by the earth's surface or with aerosols to form clouds. To achieve this required monitoring of humidity, satellite measurements have become an indispensable source of information. Especially microwave (MW) sounders onboard of polar orbiting satellites have the capability to track layer averaged humidity globally and even for moderately cloudy situations (John et al., 2013a). However, to enable the use of such datasets for climate applications, quality control of the datasets in the first place is key.

Microwave Humidity Sounder (MHS) and its predecessor Advanced MW Sounding Unit-B (AMSU-B) have been operating onboard of seven polar orbiting satellites for almost two decades. They measure in three channels around the 183.31 GHz absorption line of water vapor, providing information about layer averaged humidity in three layers of the troposphere. The MW instruments also have two window channels that are sensitive against surface temperature and emissivity. For each of these MW instruments, the project "Fidelity and uncertainty in climate data records from Earth Observations" (FIDUCEO) aims to produce Fundamental Climate Data Records (FCDRs). FIDUCEO defines FCDRs as "*long, harmonised records of uncertainty-quantified sensor observations that are calibrated to physical units and located in time and space, together with all ancillary and lower-level instrument data used to calibrate and locate the observations and to estimate uncertainty.*" (FIDUCEO, 2019-01-10).

To establish such data records for the MW instruments, the various datasets need to be harmonised. Within FIDUCEO, a harmonised satellite data record is one where all the calibrations of the sensors have been done consistently relative to reference datasets which can be traced back to known reference sources, in an ideal case back to SI (Système international d'unités). The instrument calibration is given by the instrument and channel specific terms and parameters of the so called measurement equation. The measurement equation has the role of transforming raw digitized measurement counts into a radiation quantitiy, namely the Brightness Temperature (BT). For the purpose of this work, it is sufficient to understand that the harmonisation procedure encompasses two main steps: First, inter-satellite biases due to differences in instrument calibration are identified and characterised. Second, according measurement equation parameters are adjusted to reduce the identified inter-satellite biases. This also means that inter-satellite biases may still exist after the harmonisation due to systematic geophysical differences in the measured scenes of the instruments. Hence specific Inter-Calibration (I-C) methods are necessary to disentangle inter-satellite biases caused by issues in instrument calibration and geophysically caused biases.

An example of a geophysically caused inter-satellite bias between two MW instruments is the difference in Equator Crossing Time (ECT) of the different satellites. The ECT is the local time at which a satellite crosses earths equator, which is generally a fixed quantity for the satellites of interest here, since they are all placed in so called sun-synchronous orbits. Therefore, instruments on satellites with different ECTs systematically sample different phases of the diurnal cycle. This introduces a systematic geophysical bias between the satellite datasets, which needs to be disentangled from biases due to instrument calibration when harmonising the satellite datasets.

In this work, four I-C methods are first introduced and then applied to the datasets of the MW instruments. By comparing the results from these four different approaches, each I-C method is evaluated against the results of the other methods. John et al. (2013a) already applied this general idea of comparing different I-C methods. However, the set of I-C methods investigated in this work advances methods proposed in the past and introduces new approaches, that have not been applied to the MW instruments that are of interest now. Notably, this includes the I-C method chosen within FIDUCEO for the harmonisation, which is of particular interest to evaluate.

The first I-C method is based on an approach by John et al. (2013a), where tropical oceans are defined as a Natural Calibration Target (NCT). This region is assumed to be a good calibration target because the humidity diurnal cycle over tropical oceans is found to be small (Kottayil et al., 2013). In this work the method is adapted by reconsidering the geographical area and instrument viewing angles of the satellite instruments that are used. Compared to John et al. (2013a), this data filtering allows to find scan angle dependent biases and is shown to reduce diurnal cycle impacts further.

The second I-C method investigated here is based on reanalysis observation feedback from the ECMWF Reanalysis-Interim (ERA-Interim) dataset. The observation feedback is a side product when assimilating the satellite radiances into the reanalysis model. Before assimilation of a satellite measurement, a prediction for that measurement is simulated by a Radiative Transfer Model (RTM) based on the background field of the reanalysis model at the specific measurement location and time (for more detail, see section 2.2. and Dee et al. (2011)). From that, so called Observation minus Background (O-B) datasets are generated for each MW instrument, respectively. Inter-satellite biases can then be calculated as Double Differences (DDs) over the O-B datasets. The underlying assumption for this method is that the reanalysis background field is an equally robust geophysical reference across the various MW instruments. This way, geophysical causes for inter-satellite biases are in theory eliminated in the DDs. This methodology is applied to the satellite datasets of interest and by comparing the obtained biases results to those from the other methods, the underlying assumption of the O-B method is tested.

The third I-C method discussed here is based on Simultaneous All Angle Collocations (SAACs), which were first introduced by John et al. (2013b). SAACs occur when two MW instruments measure at nearly the same location and time, within confined spatial and temporal thresholds. John et al. (2012) found that the most suited values for these tresholds are 5 km in space and 5 minutes in time. It is known that SAACs mainly occur at latitudes $> 70^{\circ}$ N/S (John et al., 2012), which leads to an overrepresentation of dry atmospheres in the SAAC dataset compared to the full satellite datasets. The benefit of the method is that individual measurements are compared, making the method more traceable, whereas with the other methods only statistical quantities based on larger data subsets are used for comparing the satellite measurements. This is a reason why within FIDUCEO, SAACs have become the chosen method for the harmonisation. It is of particular interest now, how SAACs perform compared to the other three methods investigated in this work.

To assess the validity of SAACs, the measure for the inter-satellite comparison is reconsidered: Next to inter-satellite differences of monthly means, also inter-satellite differences of multiple percentiles are calculated. Compared to the bias, percentile differences give additional information about the similarity of the underlying frequency distributions. In particular, by comparing percentiles, the scene temperature correlation of the inter-satellite bias can be assessed. Scene temperature dependencies of the bias could have strong implications on the validity of SAACs, because they only represent the cold part of the full BT range in the MW sounding channels due to their high latitude restriction. Scene temperature correlations of the bias were found in the past by John et al. (2012) by looking at global Simultaneous Nadir Overpasses (SNOs) and was found indirectly by looking at the bias as a function of latitude (Shi and Bates, 2011; John et al., 2013a). However, not percentiles, but artificially binned datasets were used in those previous studies to characterise the scene temperature correlation.

The final I-C method that will be investigated here, is a newly developed method and supposed to complement the SAAC method. The goal with this new method is to find collocations in tropical and subtropical latitudes by looking for calibration targets that remain constant over several hours. This way, satellites overpassing those targets at their characteristic ECTs still look at the same scene. To find these stable calibration targets, the geostationary satellite dataset of SEVIRI (Spinning Enhanced Visible and IR Imager) is used intermediately. Since SEVIRI has water vapor sounding channels that are similiarly sensitive to water vapor in the atmosphere as the MW sounder channels, the idea is to look for MW collocations that might be several hours apart in time, but for which the SEVIRI

observations remain practically the same. It turns out that such collocations are found most often in the dry subsidence regions over subtropical oceans. We call these type of collocations Opportunistic Constant Target Matchups (OCTMs).

In section 2, a description of the technical and spectral properties of the MW instruments and the satellite orbits is given. In section 3 the four outlined I-C methods are more thoroughly explained and their data bases are defined. In section 3.5, the methodology used to apply the methods is described and an outline of how the results will be presented, is given. In section 4.1, the results of applying the different I-C methods are presented and compared. This is done for three selected satellite pairs, namely NOAA15/NOAA16, NOAA19/NOAA18 and MetOp-A/NOAA18. In section 4.1.4, the scene temperature correlation of the inter-satellite biases is investigated for cases where biases obtained from SAACs deviated from biases obtained from other methods. In section 4.2, a more concise comparison of the investigated methods is conducted by contrasting some of the methods characteristics with selected examples. Based on the results, final conclusions on the performance of each method are drawn and an outlook for future work on applying and further developing the I-C methods is given in section 5.

2 Microwave humidity sounders and satellites

2.1 Satellite orbits and implications for satellite inter-calibration

The MHS and AMSU-B instruments investigated in this work are operating onboard of different satellites that are all placed in sun-synchronous orbits. These type of orbits are low earth orbits with an altitude of about 850 km and a period of about 102 minutes, causing the satellite to orbit earth about 14 times per day (OSCAR, 2019-02-18). The orbital inclination against the equator of a sun-synchronous orbit is about 99°. On the one hand, the high orbital inclination allows the instruments onboard of the satellite to obtain measurements from around the whole globe, which makes them of particular interest for climate applications. On the other hand, the inclination is the key variable that determines another orbital property, the so called precession. The precession of an orbit describes the temporal change in orientation of the orbit. Such a motion is performed by an orbiting object, if a torque is applied perpendicular to the rotational axis of the object. For an earth orbiting satellite, such a torque is present due to earth being slightly thicker at the equator than from North Pole to South Pole, causing an asymmetry in earths gravitational field. This yields a rotation of the satellites orbital plane around earth. The angular velocity of this rotation depends on the inclination of the orbit. The closer the inclination is to 90°, e.g. an orbit directly crossing North Pole and South Pole, the weaker is the precession of the orbit. At a certain orbital inclination, the angular velocity of the orbit caused by precession is precisely the angular velocity at which earth orbits the sun. In this case, the orbital plane of the satellite orbits earth exactly once per year. This kind of orbit is referred to as sun-synchronous.

The described behaviour of a sun-synchronous orbit has some important implications regarding the sampling behaviour of the instruments onboard of the satellite. Since the orbital plane of a sun-synchronous satellite moves at the same angular velocity as the earth around the sun, the satellite crosses the equator always at the same local time. Therefore, a sun-synchronous orbit can be well characterised by its ECT. During one orbit, the satellite passes the equator two times, once during its ascending phase, from South Pole to North Pole, and once during its descending phase. Hence, a sun-synchronous satellite has two characteristic ECTs, namely the ECT of the ascending node and the ECT of the descending node, which are always 12 hours apart. Figure 1 shows the ECTs of the ascending node of the satellites regarded in this work. It becomes apparent that only MetOp-A has a constant ECT (until the year 2017), while the other satellites show drifts on long time scales. These drifts are caused by fluctuations in the gravitational field, for example due to other celestial bodies. Only MetOp-A was artifically held in a stable orbit by performing regular orbit adjustment maneuvers. The

local time at which the satellites cross other regions than the equator only deviates significantly from the ECT at high latitudes. Due to the high inclination of the orbit, the satellites cross longitudes slowly in low latitudes and quickly in high latitudes. Figure 1 shows that the different satellites, which are to be inter-calibrated, systematically sample very different phases of the diurnal cycle for the latitudes we are interested in most. In this way, the sun-synchronicity of the satellites carrying MW sounders, motivates the topic of this work.

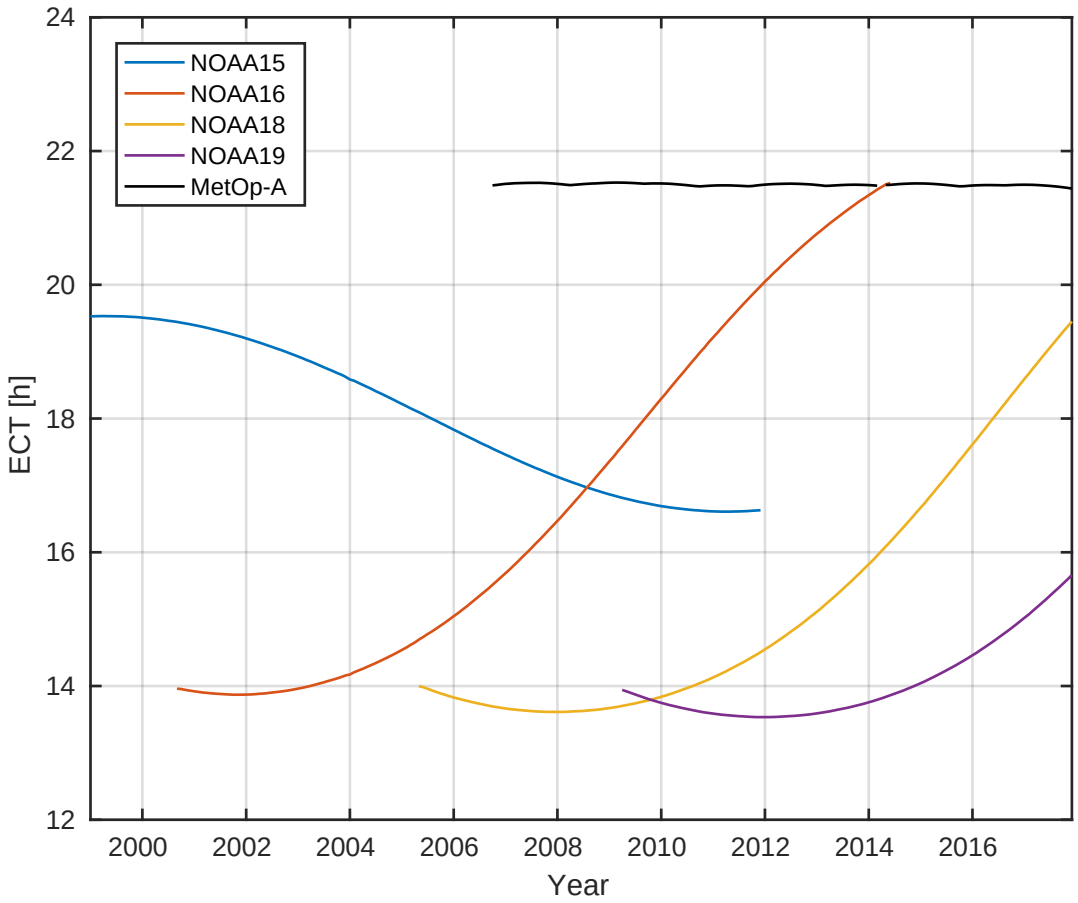


Figure 1: ECT of the ascending node of sun-synchronous satellites that carry MW sounders and are regarded in this work. Note that MetOp-B (not shown) has the same ECT as MetOp-A and started operating in January 2013.

2.2 Microwave humidity sounders

The MW sounders regarded in this work are cross track scanning radiometers, that measure the radiation that is reflected into the instrument by a rotating main mirror. One rotation of the mirror takes 8/3 seconds, during which the incoming radiation of 90 earth view pixels and two calibration targets are measured (Hans, 2018). The signal obtained from the earth views is transformed into a radiation quantity by the aid of two well characterised emitters (two point calibration): As a hot calibration target, an onboard black body with planted on thermometers is used. As a cold radiation reference, four views of the deep space are performed that detect the cosmic MW background at 2.725 K (Gawiser and Silk, 2000). The 90 earth view pixels per scanline are distributed over a 1920 km wide swath and

have a pixel diameter of 16 km at nadir. The duration of one scan takes exactly as long as it takes the satellite to move about 16 km further in its track. This way, no overlap and no gap between one scanline and the next exists.

The MW sounders are all passive radiometers, that measure incoming radiation in 5 channels. Each channel detects radiation at a distinct frequency, which carries information about specific atmospheric or surface properties. Channels 1 and 2 can generally be regarded as window channels that are only weakly sensitive to atmospheric constituents and hence mainly measure radiation emitted by the earths surface. Therefore, the radiation measured by channels 1 and 2 is mainly determined by the surface temperature and its emissivity. The main difference between channels 1 and 2 is that channel 2 is slightly more affected by water vapor, mostly from the lower troposphere, as can be seen in Figure 2 (note the logarithmic scale). The oxygen absorption always is comparatively small, but still adds a signal of the lower troposphere to the measured radiation of channels 1 and 2.

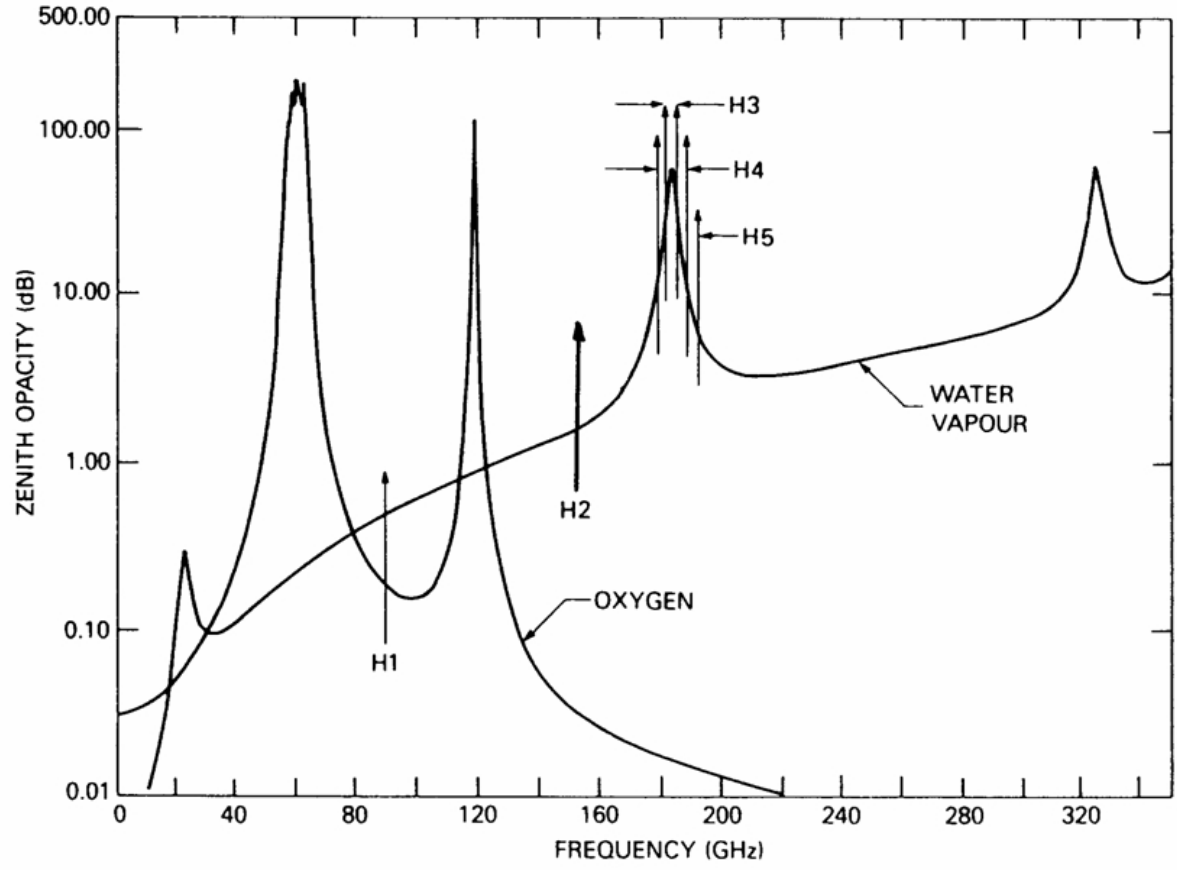


Figure 2: Location of MHS channels in water vapor and oxygen zenith opacity spectra. H1 to H5 denote channels 1 to 5, respectively. Figure by EUMETSAT (2019-02-18), viewed 09.11.2018.

Channels 3 to 5 are atmospheric water vapor sounding channels, located around the strong water vapor absorption line at 183.31 GHz (Figure 2, Table 1). By measuring the incoming radiation in 3 channels at varying distances from the 183.31 GHz line, the atmospheric water vapor profile is sounded in different layers of the troposphere. The exact sounding altitudes vary considerably with the atmospheric state, but the closer the channel generally is to the line center, the larger the contribution from higher altitudes. John et al. (2012) show in their Figure 1 how strong the zenith opacity of the atmosphere for all 5 channels depends on varying amounts of total column water vapor. They show that in cold high latitude regions even the three sounding channels can receive a significant portion of radiation from the surface due to the dryness of atmosphere. This is important later on for the SAAC I-C method, which in general only samples these high latitude regions.

Table 1: Channel names and frequencies of MHS and AMSU-B (where deviating, in brackets) and their mean sounding altitudes (OSCAR, 2019-02-18).

Channel name	Frequency [GHz]	sounding altitude
1	89.9 ± 0.9	surface
2	$157.0\,(150.0) \pm 0.9$	surface
3	183.31 ± 1.0	upper troposphere
4	183.31 ± 3.0	mid troposphere
5	$190.31\,(183.31 \pm 7.0)$	lower troposphere

3 Data and methods

The different I-C methods analysed in this work are all based on observational satellite datasets from the different MW sounders. The data used is level 1c data, which consists of BTs on a pixel by pixel basis. The data is pre-processed using the ATOVS and AVHRR Pre-processing Package (AAPP). To get an idea of the lifetime of each of the MW instruments regarded in this work, Figure 1 shows the ECT of each respective satellite until December 2017.

In this chapter, the I-C methods are introduced, which are investigated in this work. For each I-C method, a different subset of the full observational dataset is used, according to the I-C method definition. The main iterations on defining the I-C methods are presented to show why the methods are defined in the form, in which they are applied later.

3.1 Natural calibration targets

The approach with NCTs is to find stable calibration points in nature. One approach for an NCT based I-C method proposed by Burgdorf et al. (2016) is to use the moon, which occasionally appears in the deep space view of the MW instruments. In principle, the moon is assumed to be a stable source of MW radiation across the various MW instruments. Another approach for a NCT, which is more closely followed in this work, is to use certain regions on earth that are assumed to be a robust reference point across the various MW instruments. John et al. (2013a) investigated the use of Antarctica and Tropical Oceans as NCTs. It was found that over Antarctica diurnal cycle effects significantly impact the MW measurements during austral summer months, making Antarctica an unstable calibration target. However, tropical oceans imposed only a small diurnal cycle signal on the MW measurements and remained stable throughout the whole year, making tropical oceans a more robust calibration target.

In this work, a new NCT is defined, that is based on the insights obtained from the tropical oceans NCT. To define the new NCT, several changes are made to the way the tropical ocean NCT was defined previously by John et al. (2013a). They defined the tropical ocean NCT by using monthly satellite datasets over oceanic regions between 20° N/S. They also applied a cloud filter, which is based on the satellite data itself by using thresholds and channel differences. The exact procedure of the cloud filter is described by Buehler et al. (2007). Note that this cloud filter specifically aims to filter high ice clouds, while shallow not precipitating clouds are unlikely to be caught by it. The data used is also restricted to near-nadir viewing angles, e.g. the six center pixels of the MW instruments. The way John et al. (2013a) quantify the remaining diurnal cycle effect on this NCT, is to compute the difference of monthly ascending node and descending node means for the individual MW instruments. This can be viewed as an estimate for the diurnal cycle, because the ascending and descending branches of the satellite orbits are on average 12 hours apart in their local crossing time. Sticking to this method of diurnal cycle quantification, Figure 3 shows the respective monthly mean differences of ascending and

descending node for an examplary satellite and year, for accumulating changes applied to the original NCT definition of John et al. (2013a). It is found that with each additionally applied change the node differences become smaller, implying a reduction of diurnal cycle impact.

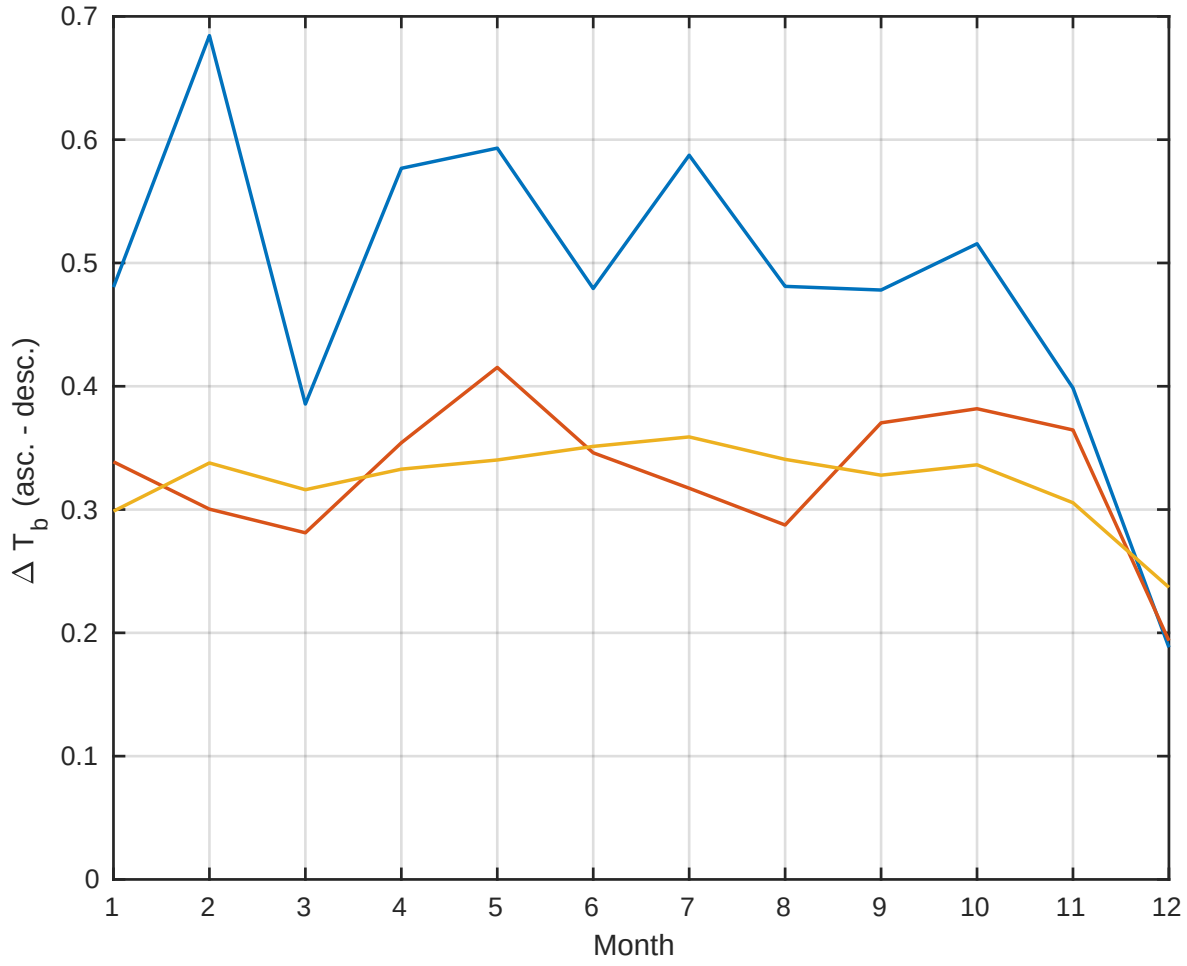


Figure 3: Difference of monthly ascending node and descending node means for the year 2010 of MHS on NOAA18 at 183.31 ± 1.0 GHz for different subsets of data. Blue: 20° N/S, land filtered out, near nadir; Orange: 45° N/S, Andes and Himalaya filtered out, near nadir; Yellow: 45° N/S, Andes and Himalaya filtered out, full swath.

The first change applied to the original tropical ocean NCT defintion is to extend the geographical region to places within 45° N/S and generally use land regions, only leaving out the tallest mountainous ranges over which the MW sensors see a strong surface signal. The extension to 45° N/S is done to include subtropical subsidence regions, which are associated with a weak diurnal cycle (Moradi et al., 2016). These subsidence regions change their extent and latitudinal position on a seasonal basis. So in order to maintain a stable databasis throughout the seasons, 45° N/S is a good estimate.

The second change applied to the original tropical ocean NCT definition is to use the data from all 90 viewing angles, opposed to only the six near nadir ones used by John et al. (2013a). On the one hand, this increases the sampling rate by a factor of 15, which minimizes the effect that strong diurnal cycle outliers can have on the computed monthly means. This can be seen in Figure 3 as the monthly mean node differences are on average smaller and less fluctuating for the full pixel range compared to just the near nadir viewing angles. On the other hand, the extension to all viewing angles introduces

two additional effects, that need to be considered: First, there are known issues of Radio Frequency Interference (RFI) for some MW sounders (Atkinson, 2001). RFI is radio frequency radiation emitted by onboard transmitters that interferes with the function of the microwave instrument and its effect strongly depends on the viewing angle of the MW instrument (for a comprehensive discussion of this effect, see Hans (2018)). Secondly, the slanted path through the atmosphere in the limb viewing angles leads to a so called limb darkening, which is the effect of systematically decreasing BTs from nadir towards the limb viewing angles. Note that this is solely a geometric effect. It is assumed that limb darkening has exactly the same impact on the different MW instruments analysed here. So when comparing sufficiently large datasets, such as monthly means, limb darkening should not significantly impact the calculated inter-satellite biases.

In summary, the newly defined NCT shows a reduction of diurnal cycle impact compared to John et al. (2013a), as is shown in Figure 3. The usage of all viewing angles brings along possible scan angle dependent biases, which will be explicitly analysed in the results in section 4.1.1. In fact, these scan angle dependent biases are useful patterns for comparing the results from the different I-C methods analysed in this work.

3.2 Observation feedback from reanalysis background

Another approach for disentangling instrumental inter-satellite biases from diurnal cycle effects is based on observation feedback from reanalysis models. Observation feedback is information about the quality of the observational data that is fed into the reanalysis model. This information is obtained by comparison of the observational datapoints with predictions of the reanalysis model. Before assimilating an observational datapoint, the reanalysis model makes a prediction that is based on other observations and the physics of the model. This model prediction is often refered to as model background, because it is independent of the particular observation that is to be assimilated. When assimilating satellite datasets, which generally consist of measured BTs, the reanalysis background is converted into BTs by a Radiative Transfer Model (RTM). This way, differences of observations and the reanalysis model background can be calculated and interpreted in terms of BT. These O-B datasets present an opportunity for satellite I-C.

The systematic difference between MW sounder datasets due to different ECTs is also, to some degree, captured by the reanalysis background. This allows to isolate the instrumental biases between the MW instruments from the effect of sampling different phases of the diurnal cycle. The procedure to achieve this is schematically depicted in Figure 4: The lefthand side shows observations and according reanalysis background predictions for three hypothetical satellites. Based on the observations alone, it would be impossible to properly estimate an inter-satellite bias, because the underlying diurnal cycle is unknown. However, when subtracting the reanalysis background from the observations (righthand side), the diurnal cycle cancels out and by Double Differencing the O-Bs of two satellites, the inter-satellite bias becomes apparent. This works under the assumption that the reanalysis background estimates the true diurnal cycle (Figure 4, dashed line) equally well throughout the day. Otherwise the inter-satellite bias would still be different depending on whether ascending or descending nodes of the two satellites are compared. This idea of using reanalysis datasets as a reference for satellite I-C has already been used in the past, for example by Kobayashi et al. (2017) to assess the calibration state of Special Sensor Microwave Water Vapor Profiler (SSM/T-2).

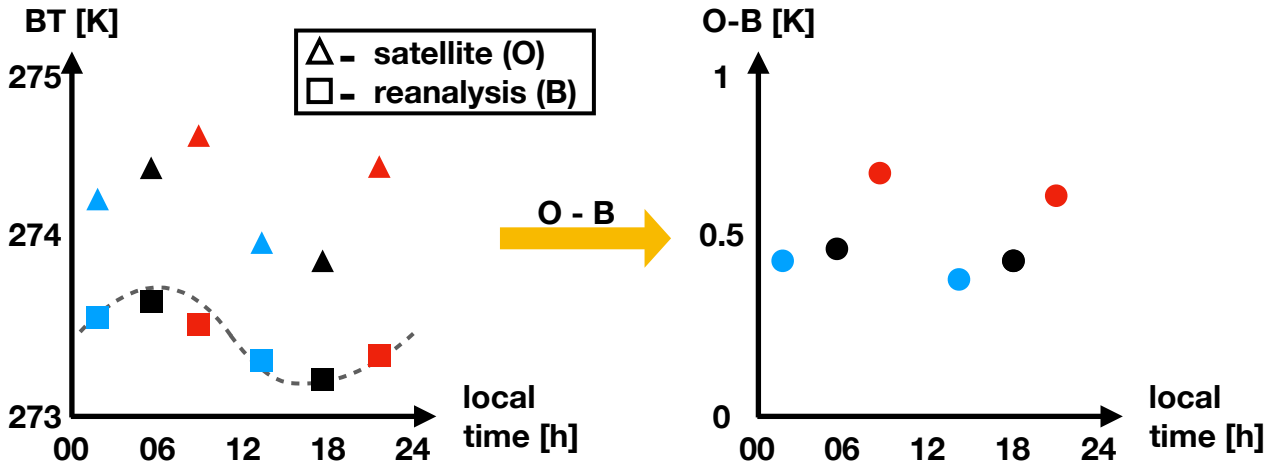


Figure 4: Schematic depiction of how O-B data is used for I-C. Each color represents a satellite that measures at two characteristic ECTs. Triangles represent observational data, squares represent reanalysis background, gray dashed line represents the true underlying diurnal cycle. Note that there is no actual data used for this schematic figure.

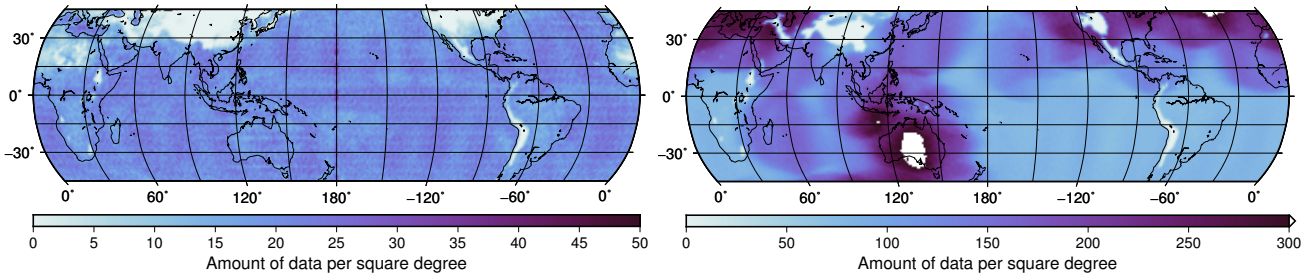
The degree to which the model background is able to capture the diurnal cycle in the quantity measured by the MW sounders depends on a number of questions regarding the reanalysis model: To which extent do the physics of the underlying Numerical Weather Prediction (NWP) model allow a diurnal cycle? Is there sufficient data assimilated into the reanalysis model to allow for a profound prediction? Is the assimilated data biased? In general, the ERA-Interim model performance is proven to be very good, as for example broadly expressed by Dee et al. (2011). Many problems of previous reanalysis datasets were able to be solved in ERA-Interim, such as a proper representation of the hydrological cycle. However, specifically for the tropical upper troposphere, the ERA-Interim background shows a moist bias against multiple observational datasets (Simmons et al., 2014). This bias would only effect the calculated inter-satellite biases, if it depended on daytime. Currently, there is no evidence in literature that biases in ERA-Interim are structured this way.

To obtain results from this analysis, which are directly comparable to the results obtained from the other investigated I-C methods, the O-B data is processed similarly as described in the NCT approach. Monthly means of satellite specific O-Bs are calculated for the region between 45° N/S. Only the tallest mountain ranges are left out due to surface contamination of the MW measurements. The ERA-Interim observation feedback dataset used here includes a bias correction variable, which is calculated by the variational bias correction system of the reanalysis model (Dee and Uppala, 2009). For this work, the bias correction variable shall not be regarded and is therefor added to the O-Bs, in order to retain all biases of the observational data. Another technical trait of the O-B dataset is that it is artificially thinned by only containing every third field of view in each scanline (e.g. pixels 3, 6, 9, ..., 90) to reduce the data size. Also, it shall be noted here that ERA-Interim observation feedback at the point of writing is only available for all satellite datasets until December 2012. The data is obtained via personal communication with Viju O. John and an extension of the dataset was unfortunately not possible.

Another consideration when using O-B data for satellite I-C is, whether only O-Bs shall be used that were assimilated into the reanalysis or to also include O-Bs that were excluded from assimilation because the bias calculated from the variational bias correction system was too large. To use only assimilated data has the advantage that cases where the observation and the model background deviate strongly, e.g. due to surface or cloud contamination, are excluded. To filter out such cases for the purpose of satellite I-C is desired, because they are likely linked to a diurnal cycle and would therefore impact the satellite datasets differently (John et al., 2013a). Additionally, it turns out that the assimilated O-Bs

are also very homogeneously distributed geographically, as shown in Figure 5a. This is advantageous, because this way the I-C dataset is not geographically biased. However, using only assimilated O-B data has a critical disadvantage. MW instruments on some satellites are subject to strong instrumental issues that yield large constant or increasing biases to the model background. Therefore, no data at all is assimilated into the reanalysis from those instruments. This is for example the case for all channels of AMSU-B on NOAA15 and channel 3 of NOAA19. However, in order to evaluate this I-C method, there needs to be some available data.

Including unassimilated O-Bs greatly increases the amount of data, but also includes cloud and surface corrupted cases again. By applying the same cloud filtering method for MW measurements by Buehler et al. (2007) as for the NCT method, such cases can greatly be reduced. Additionally, Figure 5b shows that when using unassimilated data, strong inhomogeneties in spatial sampling are introduced. It is not clear where this comes from but it needs be accounted for before using this as the data basis for this I-C method. In order to achieve spatially homogeneous weighting of the O-Bs, the monthly datasets are first put into geographical $1^\circ \times 1^\circ$ bins, which are then averaged before calculating the global monthly mean. This way, every $1^\circ \times 1^\circ$ bin has the same weighting in the global mean.



(a) Only assimilated O-Bs.

(b) Including unassimilated O-Bs and cloud filtered according to Buehler et al. (2007).

Figure 5: Monthly amount of O-Bs for two different types of data filtering. As an example, the data of January 2011 from channel 3 of MHS on NOAA18 is chosen here.

The geographical binning and averaging naturally has an impact on the frequency distribution of the BTs, that are later on the basis to calculate monthly means and percentiles. This impact is shown in Figure 6. The normalised frequency distribution of only assimilated observed BTs (red histogram) can be viewed as a reference, because it has an underlying spatially homogeneous frequency distribution (Figure 5a) and no averaging is involved. When adding the unassimilated data and manually applying the cloud filtering, strong inhomogeneties in the spatial frequency distribution are introduced (Figure 5b), which leads to an over-representation of rather low BTs as depicted by the blue histogram in Figure 6. Hence, the odd spatial sampling of unassimilated O-Bs introduces a bias into the frequency distribution of observed (and also model background, not shown) BTs. Since it is unclear where the spatial sampling pattern comes from, it cannot be assumed that it is the same for other satellites. Therefore, artificial inter-satellite biases could be introduced by this, which are simply implications of sampling. This apprehension can be eliminated by geographically averaging over $1^\circ \times 1^\circ$ bins, yielding the yellow histogram of Figure 6. Although the distribtution is more narrow due to the averaging process, it is not biased against the red reference distribution of only assimilated observations. Therefore, when inter-satellite biases are obtained with this O-B method, the data basis will be geographically binned and averaged assimilated and unassimilated O-Bs with a manually applied cloud filter.

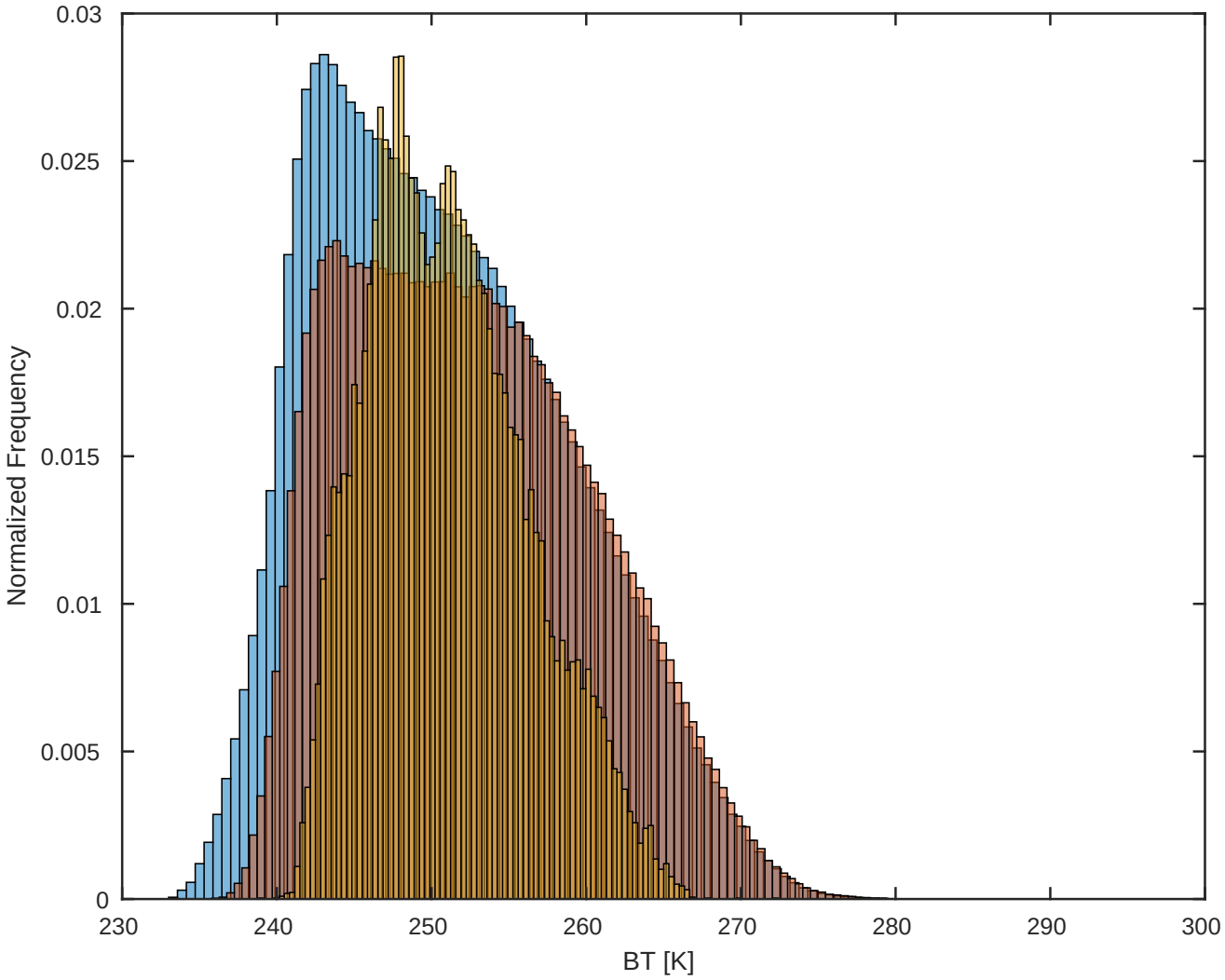


Figure 6: Normalised frequency distributions of differently filtered observations from channel 3 of MHS on NOAA18 for January 2011. Red histogram is based on only assimilated data, blue histogram also includes unassimilated data with a manually applied cloud filter. The yellow histogram uses the data of the blue histogram, but is based on geographical $1^\circ \times 1^\circ$ averages of the data.

3.3 Simultaneous all angle collocations

A very intuitive and established method for satellite I-C is based on SNOs. SNOs are events, where two satellites cross approximately the same geographical location during a short timewindow, so that the instruments on board measure practically the same scene in their nadir view. The measurement pairs of the two satellites during these events are refered to as collocations or matchups. Because the timewindow within which the two satellite measurements are considered matchups is in the order of minutes, it can be assumed that by definition diurnal cycle effects cannot systematically impact SNOs.

First, Cao et al. (2004) introduced a method for finding SNOs and argued that it is important to constrict the collocation procedure to near nadir to eliminate differences in atmospheric path and viewing geometry. However, John et al. (2013b) extended this method to all viewing angles to allow the explicit analysis of viewing angle dependent biases. They applied the restriction that matched scan positions have to be equal, for example scan position 1 of one satellite is only matched with scan position 1 of another satellite. This negates the possibilty of limb effect impacts in the matchup data. John et al. (2013b) refer to their methodology as SAACs.

The procedure of SAACs is also followed in this work, with a relaxation in viewing angle restriction, in order to obtain as many matchups as possible. Here, also opposing scan positions that still have the same viewing angles are allowed for collocations, for example scan position 1 of one satellite is allowed to match with scan position 90 of the other satellite. Additionally, any combinations of the near nadir scan positions, implying the nadir $\pm$ 5 pixels, are allowed for collocations. This is reasonable because the limb effect is negligible that close to nadir, as shown for example in Figure 2 of Moradi et al. (2016).

The exact spatial and temporal collocation criteria used in this work follow the established best estimates by John et al. (2012): Spatially, the two satellite measurements are allowed to be up to 5 km apart and they have to occur within a 5 minute temporal window. John et al. (2012) find these thresholds to be suited, by looking at the sensitivity of the standard deviation of inter-satellite differences from matchups with varying spatio-temporal collocation criteria. Based on that analysis, 5 km in space and 5 minutes in time can be considered stringent collocation criteria, which satisfy the assumption that both satellites are practically looking at the same scene. Note that cloud contaminated scenes are not filtered out in this method. This is because the cloud filter used by Buehler et al. (2007) also filters out basically all measurements over land for the dry polar atmospheres, dramatically reducing the amount of matchups obtained. Also, it was assessed by John et al. (2013a), whether inter-satellite biases of NOAA16 against NOAA15 change with or without the application of a cloud filter and no significant differences were found.

A main consideration with this method is the frequency and regularity with which matchups occur. Since all MW instruments of interest in this work are located on polar orbiting satellites in sun-synchronous orbits, a few general statements about the spatio-temporal collocation patterns can be made: SAACs will only regularly occur over polar regions, because it is the only region where all polar orbits intersect. SAACs in lower latitudes can only occur due to orbital drift, which is the slow drift in ECT of a satellite in orbit (see Figure 1). This can result in a short period in the order of weeks during which two satellites have a very similar ECT. During such a period of orbital overlap, SAACs can occur globally and much more frequently than usual. Figure 7 shows the frequency of matchups per month for three satellite pairs regarded in this work. NOAA19 and MetOp-A have very different ECTs and therefore show very different matchup patterns. NOAA19 shows by far its most collocations with NOAA18 during their period of orbital overlap around November 2009. John et al. (2012) investigated such events of orbital overlap for MW sounders onboard of multiple satellite pairs. They found that inter-satellite biases can vary with BT in inconclusive patterns and with varying magnitude. It remains unclear why inter-satellite biases show such variations with BT, but this raises the question, whether SAACs from polar regions are a sufficient data basis for fully capturing inter-satellite biases. By comparison of calculated inter-satellite biases from SAACs to the results from other methods investigated in this work, this question will be addressed in chapter 4.1.3.

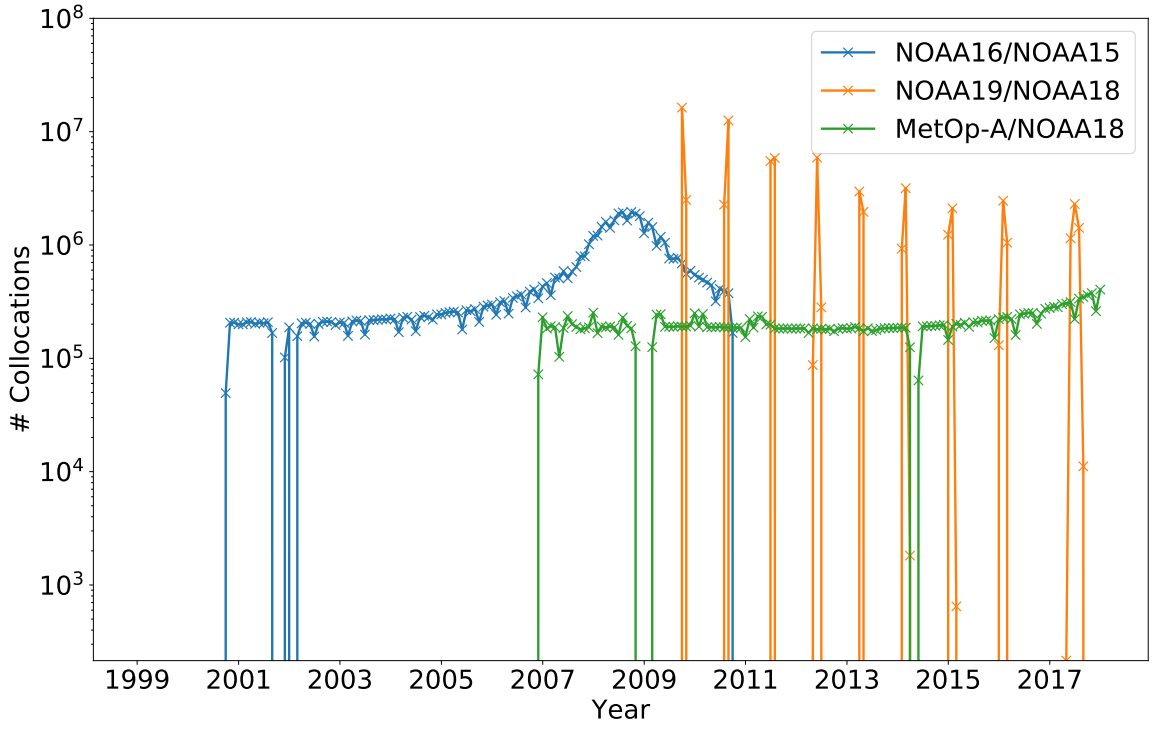
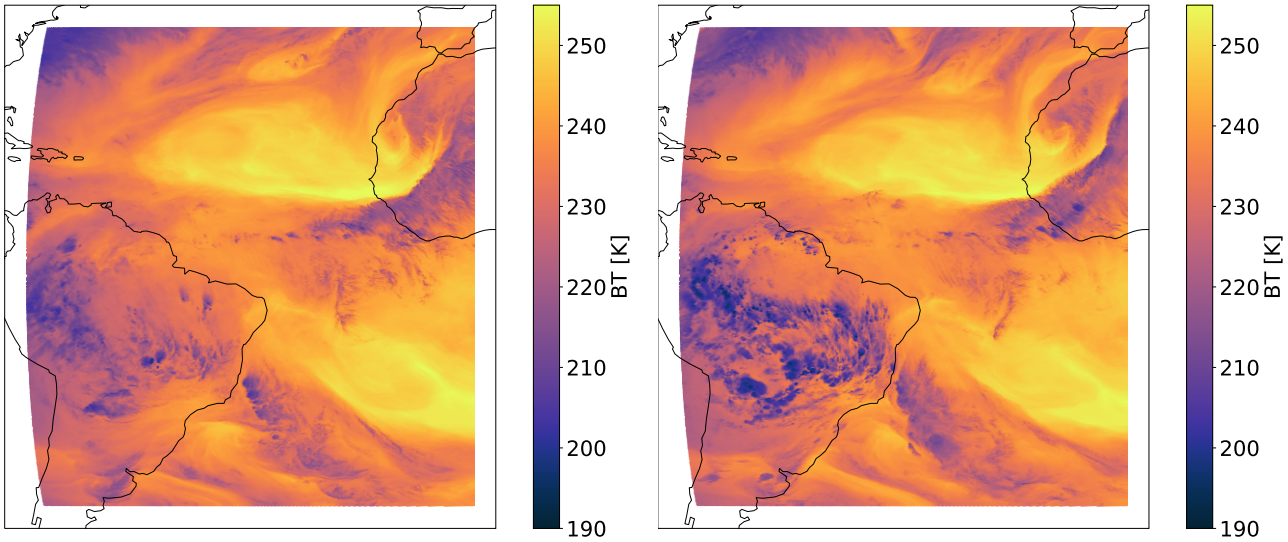


Figure 7: Frequency of SAACs per month for NOAA16 with reference to NOAA15 and NOAA19 and MetOp-A with reference to NOAA18, when applying the described collocation criteria.

3.4 Opportunistic constant target matchups

The major drawback of SAACs being that they almost exclusively occur over polar regions motivated the investigation of a complementary approach for finding collocations. To find collocations over lower latitudes, the rather stable subsidence regions in the sub-tropics appear as potential calibration targets. Figure 8 shows subsequent geostationary satellite images of SEVIRI onboard of Meteosat-10 at 6.2 μm, which is a water vapor channel, receiving most of its signal from the upper troposphere. A quasi-stability of the dry subsidence regions (marked by low BTs) is found, meaning that while the subsidence regions may show slight shifts in extent and intensity, they generally retain their dryness and large scale position over several hours. In contrast, major local disturbences probably due to quickly developing deep convective events can be found over the Atlantic Intertropical Convergence Zone (ITCZ) near the equator and also over land. Next to their stability, the dry regions are additionally suited because they are associated with relatively large BTs, which are missing in the SAACs.



(a) SEVIRI image, 15.12.2013, 12 pm

(b) SEVIRI image, 15.12.2013, 8 pm

Figure 8: SEVIRI images at 6.2 μm that are 8 hours apart in time.

The quasi-stability of subsidence regions allows to greatly ease the temporal collocation criterion and replace it with the condition that the measured SEVIRI BT at 6.2 μm has to remain roughly constant between the overpasses of two MW instruments, which may be several hours apart. By this procedure, collocations are found that may be several hours apart in time, but the new SEVIRI condition assures that only those events are selected, for which the measured scene practically does not change. Figure 9 shows that such OCTMs mainly occur over the southern subsidence region and surprisingly no partiuclar abundance maximum is found over the north Atlantic subsidence region.

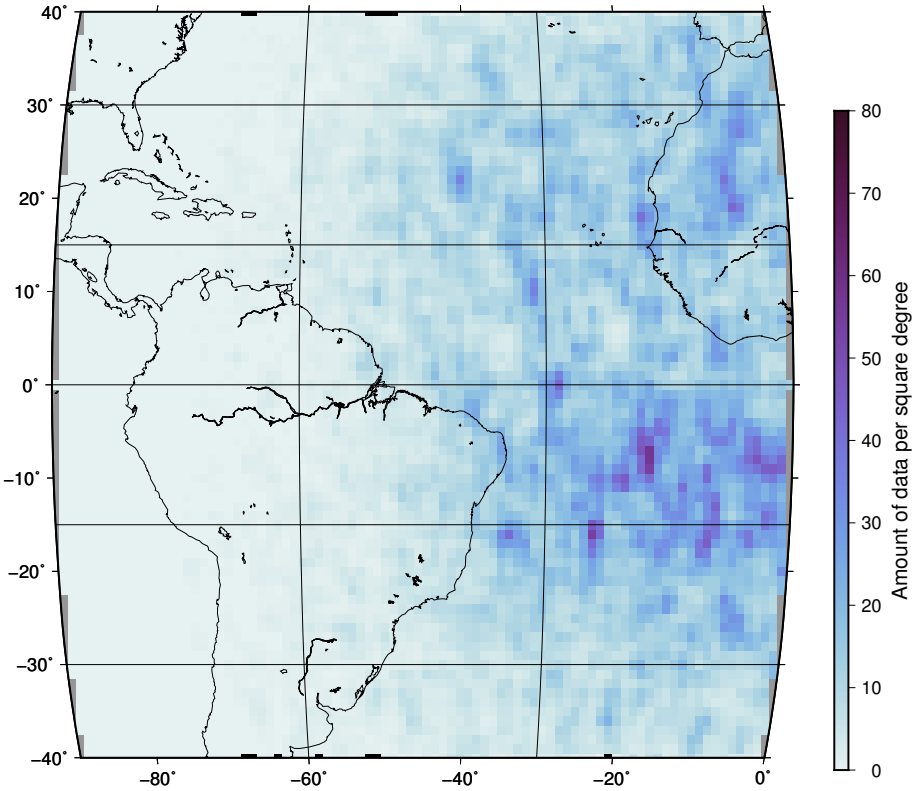


Figure 9: Geographical distribution of frequency of occurence of OCTMs between MetOp-A and NOAA18 in the year 2013, when applying all collocation criteria as described in section 3.4.

The workflow of creating OCTMs is schematically depicted in Figure 10, together with the exact collocation criteria applied in each of the collocation steps. Starting with 3 input datasets (2 MHS datasets and the SEVIRI dataset), the first computational step (gray boxes) is to collocate SEVIRI with each of the two MHS input datasets. The two obtained collocation datasets are then used as input for the second collocation step. For the second step, the temporal criterion is significantly relaxed to 8 hours in order to get matchups of the two MHS datasets. The required temporal threshold to obtain collocations at low latitudes can generally be estimated by the difference in ECT of the two collocated MHS satellites. Here, 8 hours were selected because it is a suited value for all satellite pairs regarded in this work. For example, the ECTs of the ascending node of NOAA18 and MetOp-A are up to 8 hours apart, before the strong drift of NOAA18 starts around 2012 (Figure 1). Additionally for this collocation step, a threshold is put on how much the two SEVIRI measurements are allowed to deviate between the two MHS overpasses. This threshold has to be sufficiently small to reduce the impact of systematic effects of the atmosphere on the matched MHS observations as much as possible, for example the mean diurnal cycle. However, the smaller the treshold, the less collocations are obtained. Considering this, $\Delta SEVIRI_{6.2\,\mu\mathrm{m}} = 0.8\,K$ is found to be a suited value. Additional insight on finding this threshold and more detail on the methodology itself is aimed to be published in a seperate publication following this thesis. The output of this collocation step are the seeked after OCTMs, e.g. matchups of two MHS instruments at low latitudes. The performance of this methodology will be evaluated by comparing its results to the other methods investigated in this work. In particular, the complementarity of SAACs and OCTMs will be discussed later on.

Technically, SEVIRI allows for this procedure to work because of its high temporal and spatial resolution: SEVIRI makes an image of the full earth disk, as seen from the zero meridian over the equator every 15 minutes at about 3 km spatial resolution. SEVIRI data is available since 2002 from three consecutive Meteosat Second Generation (MSG) satellites, namely Meteosat-8, -9 and -10. The collocations for this work were created in collaboration with Viju O. John, using native SEVIRI data and by using the Collocation Software created by John Mrziglod that is publically available as part of the Typhon Toolbox (Typhon, 2019-02-18).

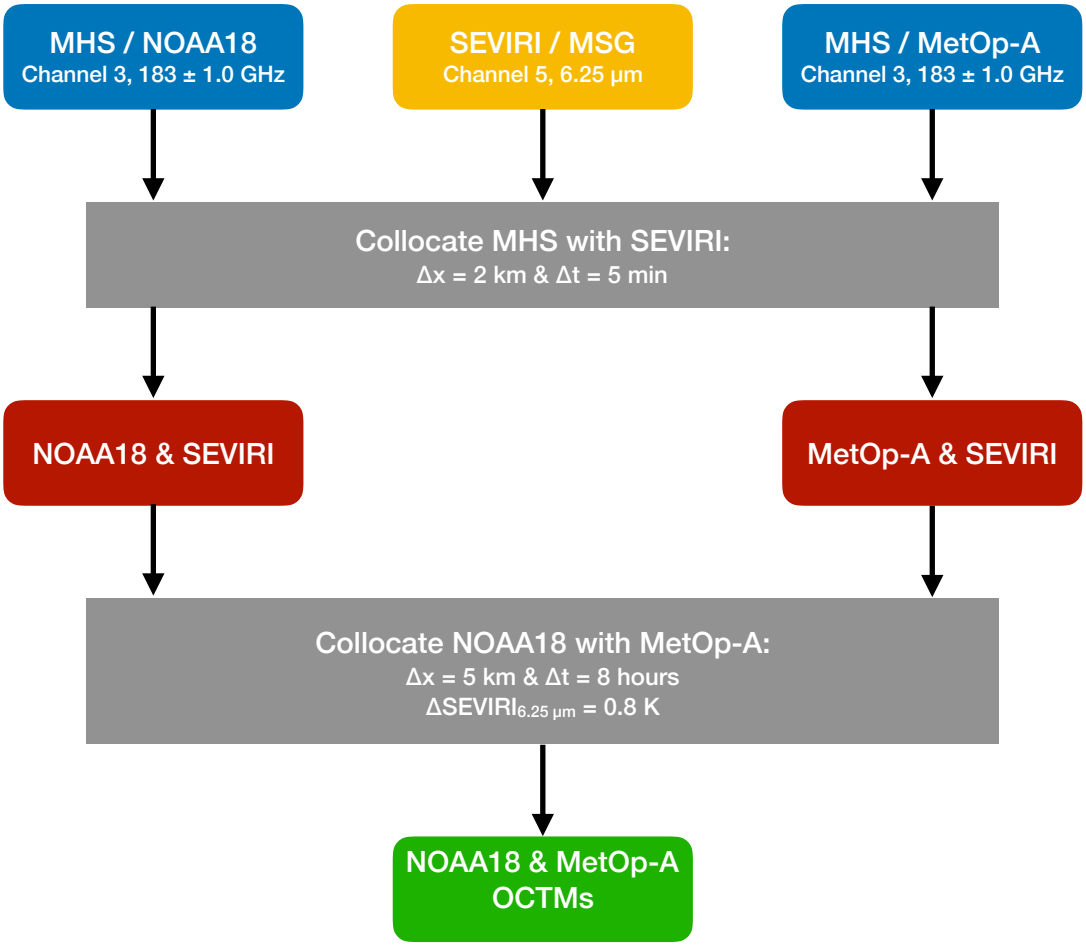


Figure 10: Workflow of creating OCTMs for the example of collocating NOAA18 and MetOp-A (Figure adapted from John Mrziglod, used with permission).

3.5 Method for comparison of inter-calibration methods

Here, the concept that will be followed in the next chapter to analyse and compare the different I-C methods is outlined. The general approach is similar to that of John et al. (2013a), as the basic quantity to obtain from the different I-C methods are monthly inter-satellite biases. Long timeseries of monthly inter-satellite biases are computed for each of the four introduced I-C methods. These timeseries are presented and comprehensively discussed in chapter 4.1. The timseries are also the basis for the I-C method comparison. In chapter 4.2, a more direct and concise comparison of the I-C methods is conducted by looking at specific examples from the previous chapter more closely. Possible deviations in inter-satellite biases between the methods are discussed and tried to be attributed to method specific merits and demerits. Prior to having seen the I-C method results, the following considerations can already be made and expected to be addressed in the next chapter.

The basic assumptions that are required to hold true for each of the I-C methods to work need to be checked. This is important because the obtained inter-satellite biases could be erroneously affected by assumptions not holding true. As discussed in more detail in the previous chapter, such assumptions are:

- The diurnal cycle is negligibly small over NCTs so that it does not impact the resulting inter-satellite biases.

- The reanalysis background field has no daytime dependent bias so that it is an equally robust reference for the satellites with different ECTs.
- The set of collocation criteria defined for SAACs or OCTMs assures that no systematic scene difference exists for a given satellite pair.

When discussing the obtained inter-satellite biases in chapter 4.1., these assumptions will also be addressed. Next to the performance of the underlying assumptions, it can be expected to receive differing inter-satellite biases from the different methods due to their very own sampling of the satellite datasets. In the following a few examples are given of what implications come along with the way the I-C method datasets are structured:

- For the SAAC and OCTM methods, only cases of both satellites having either near nadir or the same viewing angles are allowed for collocations. This yields more matchups at near nadir viewing angles compared to off nadir. This is not the case for the NCT and O-B methods, which are not subject to such constraints. Therefore, viewing angle dependent biases may show up differently between those methods.
- The SAAC method may show different biases compared to the NCT and OCTM methods due to the former sampling exclusively polar regions while the two latter sample exclusively the tropics and sub-tropics. If the inter-satellite bias dependends on the scene temperature, such a sampling effect would yield deviations in the resulting inter-satellite biases.
- The O-B data only contains of every third viewing angle of each scanline, possibly capturing viewing angle dependent biases differently compared to the other methods, where every viewing angle is included. This effect is only expected to be significant if the viewing angle dependence of the bias is very strong.

Viewing angle dependent biases will be discussed in chapter 4.1, when looking at the timeseries of inter-satellite biases of each method. Also, other discrepancies in the results of the different I-C methods that become apparent during the discussion, will be pointed out there. Correlations of the inter-satellite biases to the scene temperature will be analyzed in chapter 4.1.4 for cases where such correlations are suspected to exist. Particularly, it is assessed here whether the latitudinally constrained SAAC method is affected by such correlations. The final goal of this work is to thoroughly characterise each I-C method in terms of how well it is able to capture instrumental biases caused by deficiencies in calibration.

4 Results and discussion

4.1 Application of inter-calibration methods

This chapter places the focus on the presentation and interpretation of the obtained inter-satellite biases from the different I-C methods. For each method, long timeseries of inter-satellite biases are shown and discussed. This is done for three satellite pairs: NOAA15/NOAA16, MetOp-A/NOAA18 and NOAA19/NOAA18. The pair NOAA15/NOAA16 is picked because it also was subject of previous work (John et al., 2012, 2013a). Therefore, comparability of the methods presented here, to what was done in the past offers an interesting perspective. The other two satellite pairs are picked, because they have been operating the longest.

4.1.1 Natural calibration targets

Figures 11, 13 and 14 show timeseries of monthly mean BTs for MW sounder channel 1 (top) to 5 (bottom) in the left and middle panels. The black and red lines show means for ascending and descending nodes, respectively. The blue line shows the difference of monthly ascending and descending node means, which will be referred to as the node bias. The node bias is linked to the right hand side y-axis of the panels. The panels on the right hand side depict the monthly inter-satellite biases of the two satellites regarded in each figure. The black and red colors represent inter-satellite biases based on only the ascending or descending node data, respectively, and green denotes the bias based on the full orbits. These type of plots are inspired by figures 2 and 3 of John et al. (2013a). It is helpful to stick to these type of plots here to allow comparability to the previous study. The data basis used for plots presented in this chapter is described in chapter 3.1.

The timeseries depicted in figures 11, 13 and 14 are subject to some common features which shall be examined first. Channels 1 and 2 can both be described as surface sensitive channels. However, as described in chapter 2 they are differently sensitive to water vapor and oxygen in the lower troposphere. Channel 2 receives a smaller contribution of surface radiation, due to its stronger sensitivity to water vapor compared to channel 1. Therefore, channel 2 is less sensitive to temperature variations of the surface compared to channel 1 and receives more radiation from the lower tropospheric water vapor. Because the surface emissivity, in particular of the oceans is quite small for the MW domain, channel 1 generally shows smaller BTs than channel 2 (John et al., 2012).

The monthly mean BTs for ascending and descending nodes of channels 1 and 2 show pronounced seasonal patterns. The maximum is found during boreal summer, which can be explained by the larger proportion of land in the northern hemisphere. Since land generally has a smaller heat capacity than the oceans, its surface and the lower tropospheric temperature adjust quicker to a change in incident radiation from the sun connected to seasons. Therefore, a distinct seasonal pattern emerges, although an equal amount of measurements is used from both hemispheres. The three sounding channels 3 to 5 show a much less pronounced seasonal pattern. This is owed to the fact that when sampling only tropical and subtropical atmopsheres, as done with this NCT method, these channels are mainly sensitive to changes in the humidity field of the troposphere and the mean humidity between 40 $^\circ$ N/S is not subject to a strong seasonal cycle. This is indirectly supported by the results shown in Figure 2 of John et al. (2013a) using Antarctica as a NCT, where it becomes apparrent that the dryness over Antarctica practically makes the 3 sounding channels surface channels, which in return show a distinct seasonal pattern.

Another systematic difference between the surface and the sounding channels is the magnitude of the node bias, e.g. the blue line in the left and middle panels. While this difference is generally very close to 0 for the sounding channels, it is in the order of 1 to 2 K for the two surface channels. As explained in chapter 2, the important difference between ascending and descening node data is the phase of the diurnal cycle being sampled. The two nodes are generally 12 hours apart in daytime and therefore a diurnal cycle in the measured quantity yields a difference between the node means. Consequently, the results indicate that channels 1 and 2 are subject to diurnal cycle effects while channels 3 to 5 are - generally speaking - insignificantly influenced by a diurnal cycle. Evidently, this does not hold true for the full timeseries of all instruments, as can for example be seen after 2007 for NOAA15 and after 2010 for NOAA16, when clear time dependencies of the node biases are found. This will be investigated in the following paragraphs. However, it can be concluded already that the used subset of data as defined for this NCT method, seems to not be subject to significant diurnal cycle effects in the sounding channels. This allows to confidently assume, that the inter-satellite biases obtained for the sounding channels are barely affected by diurnal cycle effects due to differences in ECT of the different satellites. Hence, existing inter-satellite biases shown for the sounding channels are a consequence of instrumental calibration issues. The inter-satellite biases will be more specifically examined in the

following paragraphs, respectively for each satellite pair.

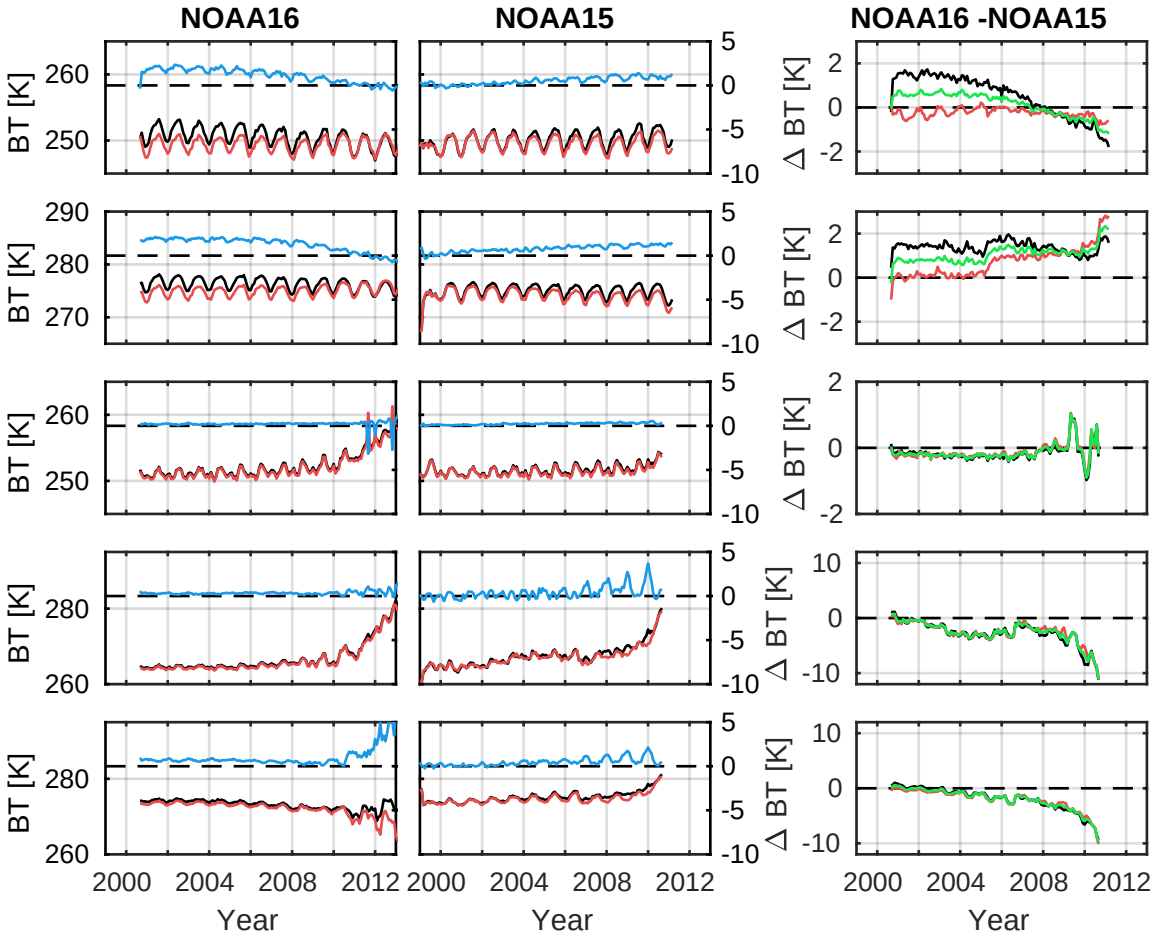


Figure 11: Long timeseries of monthly means over NCT, as defined in section 3.1, for AMSU-B on NOAA15 and NOAA16. Shown are means over ascending node (black), descending node (red) and node differences (ascending - descending, blue line, right y-axis) for channel 1 (top) to 5 (bottom). Right panels show inter-satellite biases for ascending nodes (black), descending nodes (red) and full orbits (green). Figure inspired by John et al. (2013).

Figure 11 depicts long timeseries of the satellite pair NOAA15 and NOAA16. Some of these timeseries include trends or jumps that hint towards instabilities connected to the instruments or the satellites. One trend on a long timescale of several years is present in channels 1 and 2 of both NOAA15 and NOAA16. For NOAA15 it starts roughly in the year 2001, from when on a positive trend can be observed in the node bias. The node bias increases to about 1.5 K for channel 1 and 2 in 2011. For NOAA16 a negative trend on a similar timescale can be identified starting around 2006. The node bias of about 2 K present in channels 1 and 2 prior to that trend decreases to 0 K in 2012. Consequently, a clear trend can be identified in the inter-satellite bias for channels 1 and 2. For channel 1, this trend in inter-satellite bias is only present in the ascending part of the orbit, while for channel 2 it is negative for the ascending node and positive for the descending node.

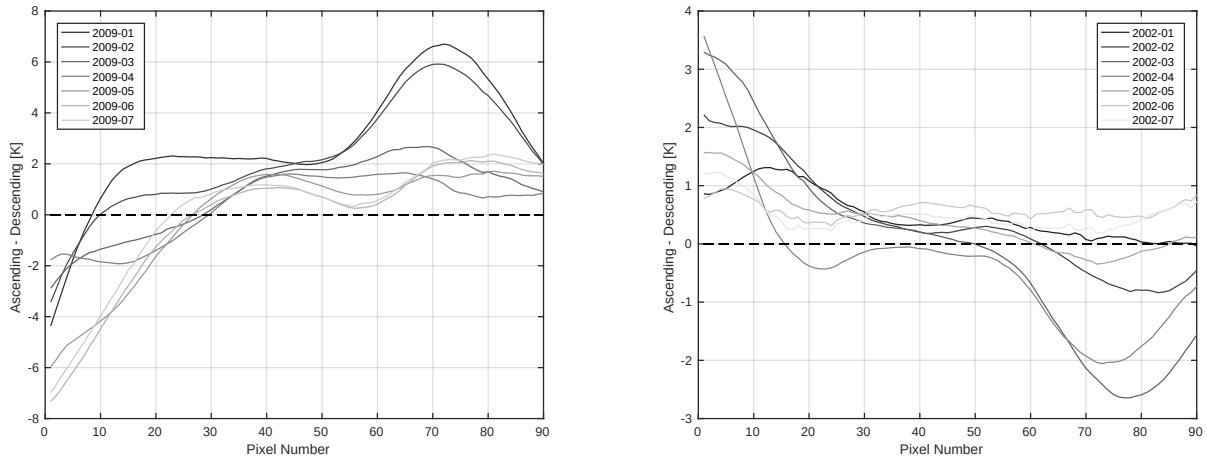
Comparing the timescale of the described trends in channel 1 and 2 with the timescales of shift in ECT seen in Figure 1 for NOAA15 and NOAA16, it becomes apparent that the trends are likely the signal of the diurnal cycle. The diurnal cycles of humidity and temperature in the lower troposphere and at

the surface are expected to be stronger than in higher altitudes (Moradi et al., 2016). This explains why particularly in channels 1 and 2 such trends are found. Due to their very different ECTs, NOAA15 and NOAA16 show opposing trends in their node bias. However, it is rather unexpected that the inter-satellite bias in channel 1 only has a trend in the ascending node. This either means that the diurnal cycle is negligible when comparing the daytimes covered by the shifts in ECT of the descending nodes, while it is very pronounced when comparing the ascending nodes. Or it hints towards an instrumental issue, which is only present during the ascending node. As will be seen later on when examining the results of the SAAC method, which by defintion cannot be subject to diurnal cycle effects, the former is the case. For channel 2, opposing trends in ascending and descending nodes are found, which would for example be the expected behaviour assuming a sinusoidal diurnal cycle. It is interesting that channel 2 shows such different behavior in the inter-satellite bias for the descending node compared to channel 1, but this might be due to channel 2 being more sensitive to the atmospheric diurnal cycle of water vapor, whereas channel 1 is more sensitive to surface temperature and emissivity. In general, the trends found in timeseries of channels 1 and 2 for NOAA15 and NOAA16 are likely all linked to diurnal cycle effects via the drift in ECT of the satellites. This preliminary result will be further evaluated by the results of the other I-C methods.

The sounding channels 3 to 5 also show trends that require further characterisation. NOAA16 channels 3 and 4 and NOAA15 channels 4 and 5 all show a positive trend in BT. It is unlikely that these trends are also caused by the drifts of ECT, because the trends start past the start of the strong ECT drifts. John et al. (2013a) do not find these trends, hence the trends must be connected to differences in the monthly subsets of data used in this work compared to theirs. One significant difference is that John et al. (2013a) constrain the used viewing angles to near nadir, whereas here the full range of viewing angles is used. John et al. (2013a) purposefully exclude off nadir viewing angles due to known issues with RFI (Atkinson, 2001) leading to viewing angle dependent biases. In this work, these viewing angles are purposefully included to enable the search for possible viewing angle dependent biases and to quantify them. Hans (2018) found that the impact of the known RFI can change with time due to changes in the channel's gain. Hans (2018) shows that both AMSU-B on NOAA15 and NOAA16 suffer from a steady gain degregation, which consequently increases the relative contribution the RFI has on the Earth-Atmopshere signal. The decrease of gain shown by Hans (2018) correlates well with the trends seen in all sounding channels for NOAA15 and NOAA16 shown in Figure 11 and can therefore in combination with RFI savely be assumed to be the cause for the trends. Although not a clear trend is visible in channel 3 of NOAA15, this channel is also subject to RFI as shown by Hans (2018). However, in the case of this channel, the viewing angle dependence almost averages out over all viewing angles.

Next to trends in both ascending and descending node, also node biases (blue lines) emerge over time in the sounding channels. All sounding channels show a temporal increase in the node bias, although different in magnitude. In general, these trends have to be caused by an effect that depends on the phase of the satellite orbit, e.g. the position of the satellite in its orbit. The node bias arising in NOAA15 channels 4 and 5 additionally features a seasonal signal. The node bias of NOAA15 channel 4 emerging in 2006 is looked into here more thoroughly as an example. To further characterise the seen behaviour of the node bias, Figure 12 shows the monthly averaged node bias of this channel as a function of the viewing angle. The monthly means for January to July in 2009, when the node bias becomes particularly large, are shown in Figure 12a. Figure 12b depicts the same for the year 2002, which can be regarded as a reference period, because the net node bias did not show a seasonal pattern as it does in 2009. However, it becomes apparent that in both years distinct viewing angle dependencies of the node bias are present. In the year 2002, the months February, March and April stick out with a significant local minimum in node bias emerging from month to month at pixel numbers around 75. On the contrary, in the year 2009 a significant local maximum around Pixel Number 75 is present in January and February. This monthly change in the bias pattern yields the seasonal dependence seen in the node bias of Figure 11. While in the year 2002, the monthly averaged node biases almost completely average out over the whole scanline for each month, in 2009 a net bias remains, particularly during

January and February.



(a) (b)

Figure 12: Viewing angle specific monthly biases between ascending and descending node of MHS channel 4 on NOAA15. Year and month are given in legends.

As a first guess RFI is suspected as the cause for these viewing angle dependent biases. In order to explain the observed orbital phase dependence of the bias, the impact of RFI would have to depend on orbital phase. This would be possible if the channel's gain depended on orbital phase (e.g. be different for ascending and descending phase of the orbit). This was checked and is proven to not be the case (not shown). Another conceivable cause for this bias behaviour is a combination of a change in the temperature environment of the spacecraft due to changes in solar insolation, with erroneous non-linearity estimation in the calibration equation, that transforms the raw counts to BTs. Zou and Wang (2011) show that the warm calibration target temperature of AMSU-A and the solar-β-angle, as a measure for incident solar radiation, are significantly correlated for NOAA15. The thereby induced increase in warm calibration point temperature can cause the transformation to radiances to be prone to errors, if non-linearity is not accounted for correctly. Zou and Wang (2011) showed that this is the case for several channels of AMSU-A on NOAA15 and it is the cause for seasonal and longterm drifts in inter-satellite biases. Hence, inter-satellite biases on similar timescales, including the seasonal signal in the node bias, depicted in Figure 11 for AMSU-B might be of similar nature. However, a more thorough investigation of this guess would neglect the aims of this work.

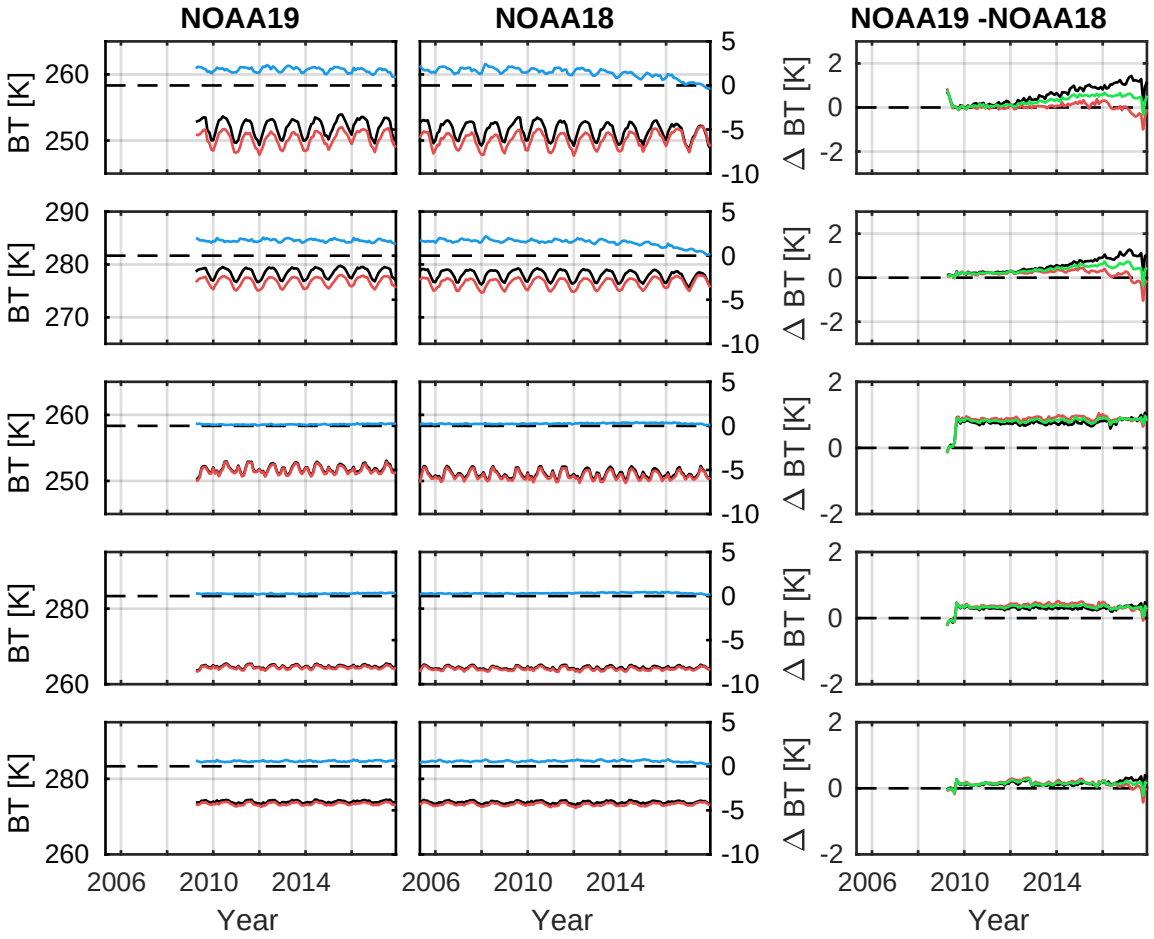


Figure 13: Same as Figure 11, but for MHS on satellite pair NOAA18 and NOAA19.

Figure 13 shows timeseries for MHS onboard of NOAA19 compared to NOAA18. Channels 1 and 2 of NOAA18 show a trend in the node bias starting around January 2016. This is similar to what was seen in channels 1 and 2 of NOAA15 and NOAA16 with the underlying reason being the strong drift in ECT and therefore sampling different phases of the diurnal cycle. Consequently, also an inter-satellite bias emerges for channels 1 and 2 between NOAA19 and NOAA18.

The sounding channels 3 to 5 for NOAA19 and NOAA18 do not show as many trends or jumps as were seen for NOAA16 and NOAA15. However, channels 3 and 4 show constant inter-satellite biases, that emerges right after when NOAA19 becomes operational in April 2009. Since the node bias for both instruments is found to be negligibile and also does not show a jump right after NOAA19 becomes operational, this inter-satellite bias is unlikely to be caused by diurnal cycle effects. In fact, Hans (2018) shows that channel 3 of NOAA19 is also suspected to be subject to RFI. The bias caused by RFI emerges in the year 2009 and then remains constant with time because the channel's gain has one strong drop in 2009 and is constantly low afterwards as shown by Hans (2018). The gain of channel 4 shows a very similar behaviour but does not drop as much as channel 3. It is therefore a save assumption that also channel 4 of NOAA19 is affected by RFI.

Another feature in the inter-satellite bias timeseries that sticks out for all channels is the strong drop of 0.6 K seen for all channels in October 2017. Generally, BTs decrease in all channels in this season from month to month. However, in this case, NOAA19 drops by about 0.6 K one month ahead of NOAA18, leading to the found bias in October 2017. It is currently unclear why this behaviour is

present. Unfortunately, this feature cannot be assessed by the other methods, presented later, due to lack of temporal coverage with the underlying datasets of the other methods.

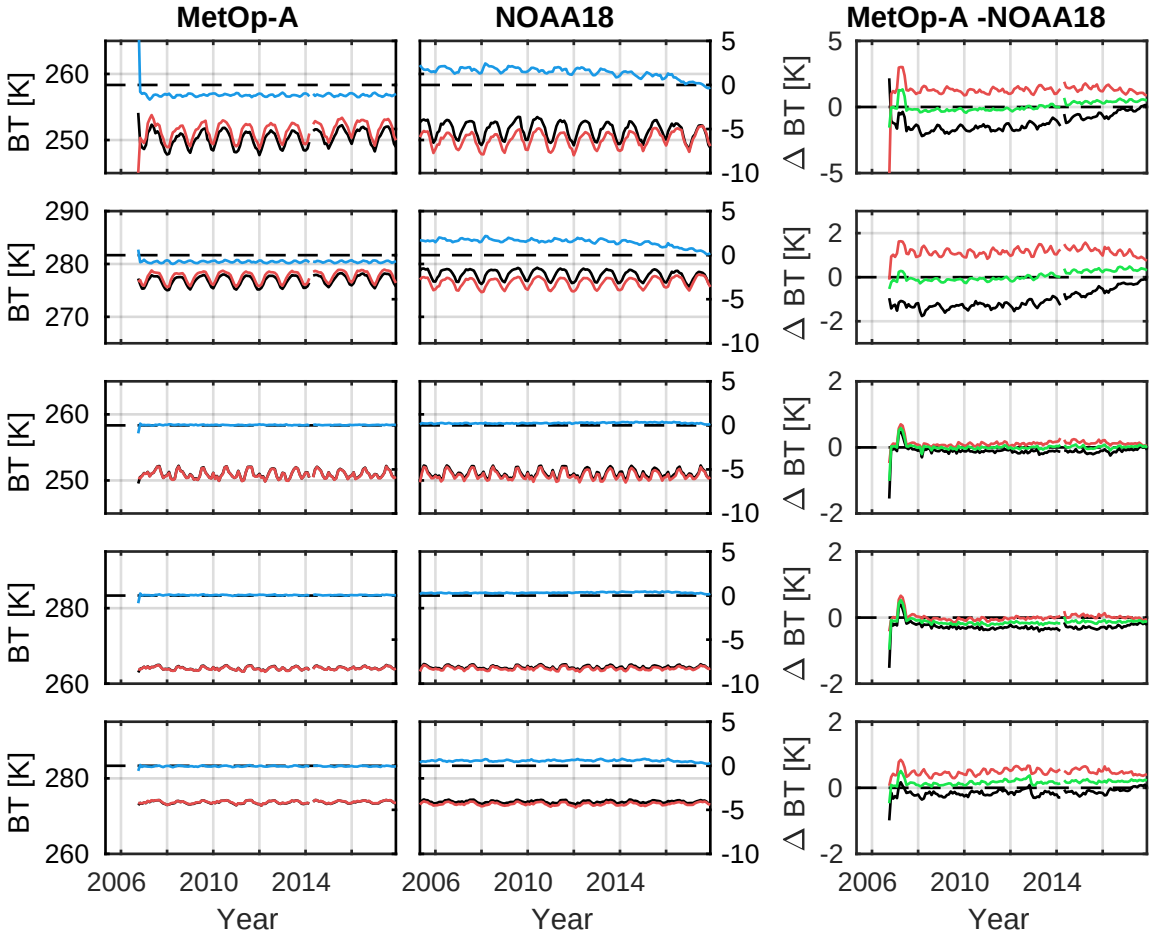


Figure 14: Same as Figure 11, but for satellite pair NOAA18 and MetOp-A.

Figure 14 shows timeseries for MHS onboard of MetOp-A and NOAA18. For channels 1 and 2, only NOAA18 shows the trend in both ascending and descending node and also in the node bias, which is related to the drift in ECT and was discussed previously. MetOp-A does not show any trends, because this satellite is held in a stable orbit (constant ECT, see Figure 1) by frequent orbit adjustment maneuvers performed by EUMETSAT. It is interesting to note that MetOp-A and NOAA18 show opposite signs in the monthly node biases in channels 1, 2 and 5, e.g. the node bias is positive for NOAA18 and negative for MetOp-A. This is owed to the very different ECTs both satellites sample the diurnal cycle at. This yields a relatively large dependence of the inter-satellite biases on which node is used for the comparison in the order of 0.5 K to 1 K for the sounding channels and 3 K for the surface channels. This reveals that inter-satellite biases obtained from the NCT method can be significantly affected by the diurnal cycle in channel 5, if the ECTs of the two satellites deviate sufficiently. Channels 3 and 4 are less affected by the diurnal cycle.

There are no other trends found for these two satellites that would be related to causes other than the orbital drift of NOAA18. This can at least be reasoned within the limit of remaining diurnal cycle impact on the two datasets. Therefore, these two MHS instruments can be viewed as very stable, well performing instruments, that do not appear to be subject to significant degregation effects as seen for

other instruments previously. It is now interesting to see, whether the same conclusions can be drawn when looking at the results from the other I-C methods, or whether those might be able to reveal more subtle instrumental trends or jumps.

4.1.2 Observation feedback from reanalysis background

In this section, long timeseries of monthly averaged satellite specific O-Bs and DDs are presented, which can be interpreted in terms of inter-satellite biases as outlined in chapter 3.2. Figures 15, 17 and 18 depict these O-Bs and DDs in a similar way as was done previously for the NCT results. This way, the results of the two methods can intuitively be compared, although there are certain limits in how strictly one can expect the plots of the two methods to agree. This is because the fundamental quantity are monthly mean O-Bs, which introduce the model background as a new variable that can potentially cause jumps, trends or other features in the presented timeseries, additionally to the observation variable. Therefore, deviations of O-Bs from zero are either caused by instrumental issues impacting the observations, by the model background not being capable of fully capturing the geophysical states of the atmosphere or by problems in the used RTM to calculate the model background BTs.

Before looking into the specific timeseries of each satellite pair, their common features shall be examined first. Looking at the timeseries of O-Bs of channels 1 and 2, a clear seasonal signal is present, which is lower in amplitude than what was seen in the results of the NCT method in the previous chapter. This was previously attributed to the assymetric distribution of land surface between the northern and southern hemispheres. The fact that this signal remains in the O-Bs at a lower amplitude goes to show that the model background is not fully capturing the seasonal cycle as measured by the satellite. In particular, this might be due to changes in surface emissivity over land with seasons, which is particularly challenging to model, but which these channels are sensitive against. In fact no data of these channels is assimilated into the reanalysis, because the departure statistics between model background and observations show that the reanalysis model cannot properly simulate these measurements (Dee et al., 2011).

Furthermore, the O-Bs in channels 1 and 2 show a systematic node bias (blue lines). Looking at the node biases of NOAA15 and NOAA16 (Figure 15), the O-Bs of the descending node are generally larger than those of the ascending node. In Figure 11, the opposite behaviour was seen, where the observed BTs in the ascending node were larger than those of the descending node. Therefore, the following inequalities hold true for NOAA15 and NOAA16:

$$\begin{aligned} & \mathrm{O_{asc}} > \mathrm{O_{desc}} \ , \quad \text{from NCT method (Figure 11).} \\ & \mathrm{O_{asc}} - \mathrm{B_{asc}} < \mathrm{O_{desc}} - \mathrm{B_{desc}} \ , \quad \text{from O-B method (Figure 15).} \\ \Longrightarrow \quad & \mathrm{B_{asc}} - \mathrm{B_{desc}} > \mathrm{O_{asc}} - \mathrm{O_{desc}} > 0 \end{aligned} \tag{1}$$

Here, the subscripts refer to the ascending (asc) and descending (desc) node portion of the data. This shows that for channels 1 and 2 the model background overestimates the node bias, which is a proxy for the diurnal cycle. This is also seen in channels 1 and 2 of the other satellite pairs shown in figures 17 and 18. The overestimation of the diurnal cycle by the model background also affects the trends seen in the O-Bs of channel 1 and channel 2 of NOAA15 and NOAA16. It makes the diurnal cycle induced trends caused by orbital drift generally go into the opposite direction compared to what was seen in the NCT method, e.g. the ascending node of NOAA16 has a positive trend in O-Bs, while it is negative in the NCT method. This is caused by the model background being subject to even stronger trends with the same sign as the observations. Thus, the model background being the subtracted quantity in the O-Bs swaps the sign of the net trend. Such opposing trends between the NCT and O-B methods are not found in any of the sounding channels for any satellite pair.

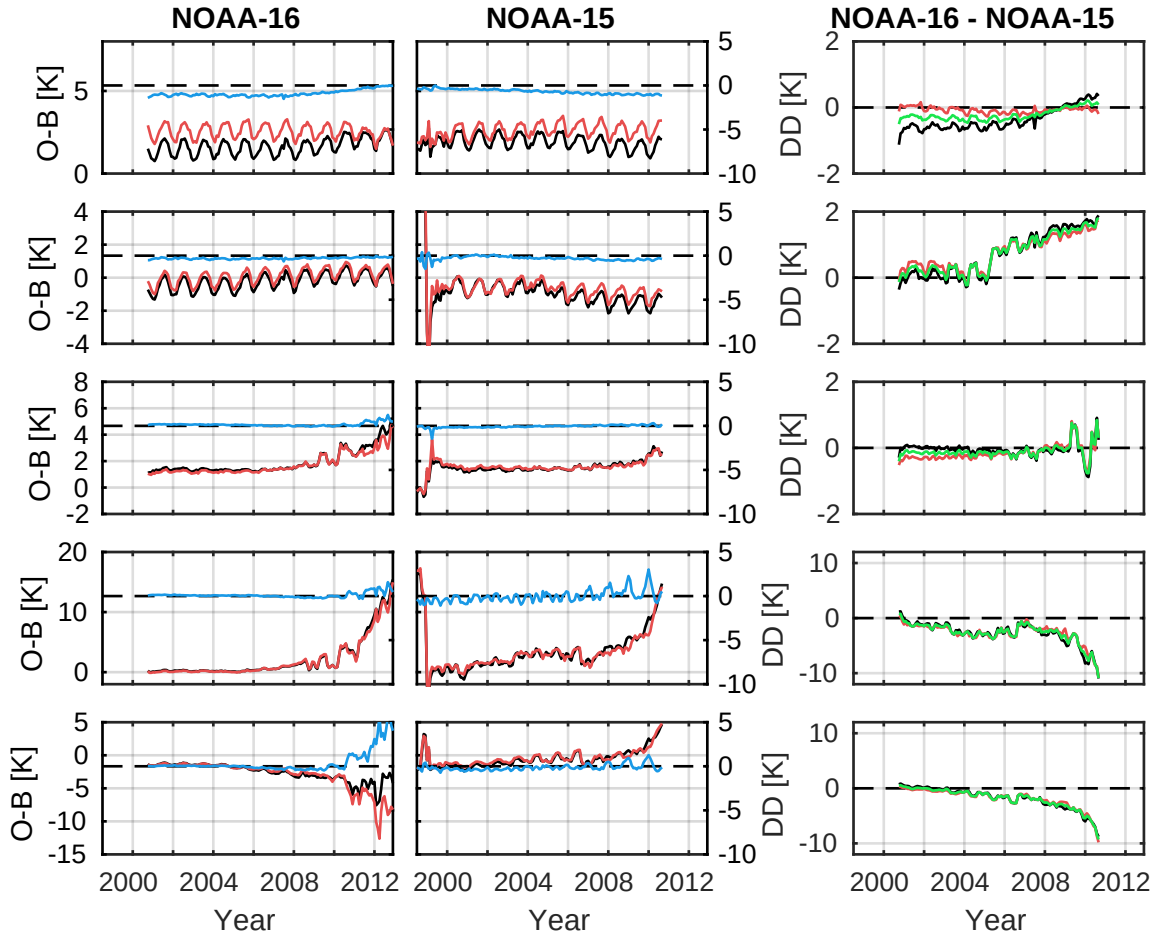


Figure 15: Long timeseries of monthly mean O-Bs between 45° N/S for AMSU-B on NOAA15 and NOAA16. Shown are means over ascending node (black), descending node (red) and node differences (ascending - descendingblue, right y-axis) for channel 1 (top) to 5 (bottom). Right panels show DDs of O-Bs for the two respective satellites for the ascending node (black), the descending node (red) and full orbits (green). Figure inspired by John et al. (2013).

The inter-satellite biases obtained from the O-Bs for NOAA16/NOAA15 are depicted in the rightmost column of Figure 15 in terms of DDs. For channel 1, a negative trend is found from the O-Bs, which is the opposite of the trend found by the NCT method in Figure 11. This is explained by the trends of the underlying O-Bs, which are opposite in sign compared to the NCT method, as explained in the previous paragraph. It is clear that this behaviour is caused by the overestimation of the diurnal cycle by the model background. The same is the case for channel 2: While a negative trend in inter-satellite bias is found in the ascending node after 2005 in the NCT method (Figure 11), a positive trend is found in the DDs of the O-B method. Hence, the O-B method has severe deficiencies for channels 1 and 2 that impact the obtained inter-satellite biases. Therefore, it can be concluded that for channels 1 and 2, the model background appears as an unsuited reference and the results should not be used for assessing inter-satellite biases.

The sounding channels 3 to 5 of NOAA15 and NOAA16 are subject to some very strong trends which generally tend to increase deviations of observation and model background (Figure 15). This is the case for channels 3 to 5 of NOAA16 and channels 4 and 5 of NOAA15, for which such trends were also seen in the NCT method. It can be inferred that the cause for these trends lies in the observations

rather than the model background. The node biases of O-Bs tend to be very small for the sounding channels, indicating that diurnal cycle effects are represented correctly by the model background in these channels. Only as RFI issues arise in channels 3 to 5 of NOAA16 and NOAA15, as described in chapter 4.1.1, do node biases in the O-Bs arise, which are very similar to the node biases seen in the NCT method. This for example includes the seasonal signal in the node bias that arises in channels 4 and 5 of NOAA15, which was addressed in Figure 12.

The inter-satellite biases of the sounding channels of NOAA16 and NOAA15 in terms of DDs shown in Figure 15 look very similar to the inter-satellite biases of the NCT method shown in Figure 11. Only for channel 3 a slight deviation between the two methods can be identified. Here, slightly different inter-satellite bias trends are found from the O-Bs from 2001 to 2008 between the ascending and descending node. This difference was not present in the results from the NCT method for this channel, indicating that the model background introduces a daytime dependend bias between the two satellites in channel 3. Although for this case the node dependence of the bias averages out well over both nodes, this issue should be treated with caution, because it could in principle lead to erroneous trends of the overall inter-satellite bias.

Another consideration mentioned in chapter 3.5 was that the data thinning conducted on the O-B dataset by including only every third viewing angle, might change the way biases that are strongly viewing angle dependend are perceived in the O-Bs compared to the other methods. In particular, this is a concern for the RFI induced biases between NOAA16 and NOAA15 in the sounding channels, because the RFI signal can be strongly viewing angle dependent (Hans, 2018). As an example, this is shown in Figure 16, where the difference of pixel specific monthly means of NOAA15 channel 3 between January 2009 and January 2003 is shown. In 2003, RFI had no significant impact because the channel's gain was at a sufficiently high level. However, as the channel's gain steadily decreases over the years (see Figure 4.5 of Hans (2018)), the impact of RFI on the measured signal becomes stronger and the pattern seen in Figure 16 emerges. From this pattern it can be concluded that the data thinning of the O-Bs does not significantly impact how the bias is perceived, because the RFI signal changes roughly speaking only every 20 viewing angles, e.g. it is around 1.4 K between pixels 1 to 20, around 1 K between pixels 20 and 40 and around 0.5 K between pixels 70 and 90. Hence, it is sufficient to only sample every third viewing angle to capture the main bias pattern. This is reflected by the fact, that inter-satellite biases between NOAA16 and NOAA15 in channels 3 to 5 do not deviate significantly between the NCT and O-B methods.

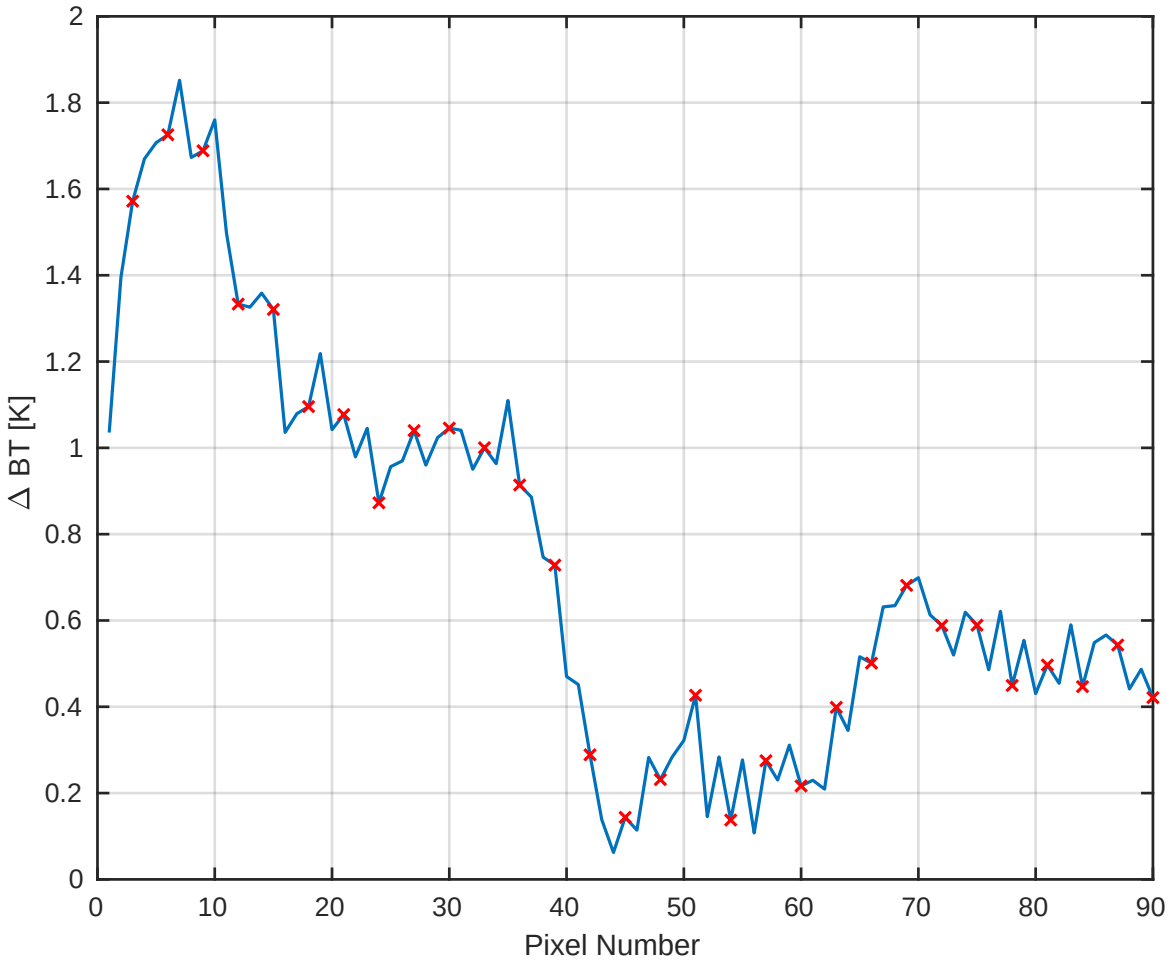


Figure 16: Pixel specific monthly mean difference of NOAA15 channel 3 observed BTs between January 2009 and January 2003 (reference). Data is based on NCT method, defined in chapter 3.1. Emerging pattern is the signal of RFI. Red crosses denote pixels that are sampled by the O-B data, which is artificially thinned beforehand.

Concluding, the fact that NOAA15 and NOAA16 show numereous instrumental issues in the sounding channels, which are caught by the NCT and O-B method in the same way, gives confidence that both methods are valid I-C methods for these channels. Surprisingly, inter-satellite biases in channel 3 showed longterm trends associated with the diurnal cycle of the model background. This is specifically looked for in the other satellite pairs, too. The instrumental issues causing jumps and trends in the shown timeseries turn out to be valuable for the purpose of I-C method evaluation. For channels 1 and 2, the reanalysis background is not a suited reference because it introduces artificial trends in the inter-satellite biases due to the joint effect of an overestimation of the diurnal cycle with the satellite drift of ECT. It is still interesting to see, whether the more stable satellite pairs presented in figures 17 and 18 reveal some more subtle methodological instabilities of the O-B method, particularly of the sounding channels.

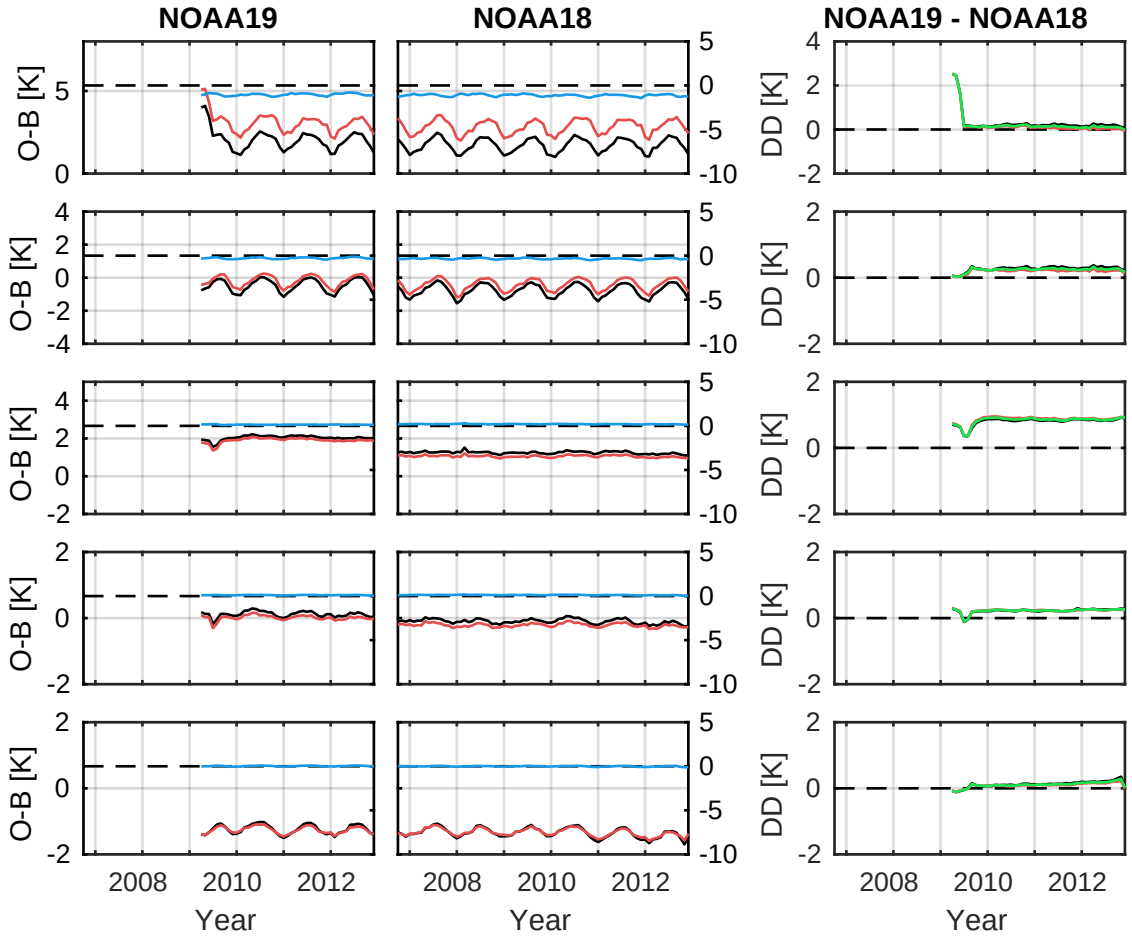


Figure 17: Same as Figure 15, but for MHS on satellite pair NOAA18 and NOAA19.

Figure 17 depicts timeseries of O-Bs for NOAA19 and NOAA18. Besides the overestimation of diurnal cycle effects by the reanalysis background, which was examined at the start of this chapter, channels 1 and 2 only show one distinct jump for NOAA19 at the start of operation. For channel 1, this jump is also caught by the NCT method in Figure 13, with a lower magnitude. Particularly in the inter-satellite biases calculated from the two methods, the difference in magnitude of this jump becomes apparent. Thus, the cause for the jump in channel 1 can only partly be attributed to the observations. However, it is unclear why this jump would be amplified by the model background.

The O-B timeseries of channels 3 to 5 for NOAA19 and NOAA18 are rather stable, when compared to NOAA15 and NOAA16. This stability reveals a seasonal cycle in the O-Bs of channels 4 and 5 which for NOAA15 and NOAA16 was most likely concealed by their strong instrumental trends. For the surface channels it was discussed previously, that the underestimation of a seasonal cycle by the reanalysis background is causing a net seasonal cycle in the O-Bs. The lowest peaking sounding channel 5 is more affected by this than channel 4, because it has more surface contribution in its measured radiation. Because the effect appears in the same way for the two compared instruments, no seasonal cycle is visible in the DDs. Hence, this effect does not impact the validity of the O-B method for I-C purposes.

The DDs show a stable 1 K bias of NOAA19 against NOAA18 in channel 3, which emerges in the first few months of the mission. In contrast to the NOAA15/NOAA16 satellite pair, the inter-satellite bias in channel 3 does not deviate significantly between ascending and descending node data. Hence, a

daytime dependence of the model background performance is not indicated by the NOAA19/NOAA18 satellite pair. Channel 4 shows an emerging inter-satellite bias in the DDs of about 0.5 K. The same behaviour of channels 3 and 4 was found previously in Figure 13 and is attributed to RFI in these channels.

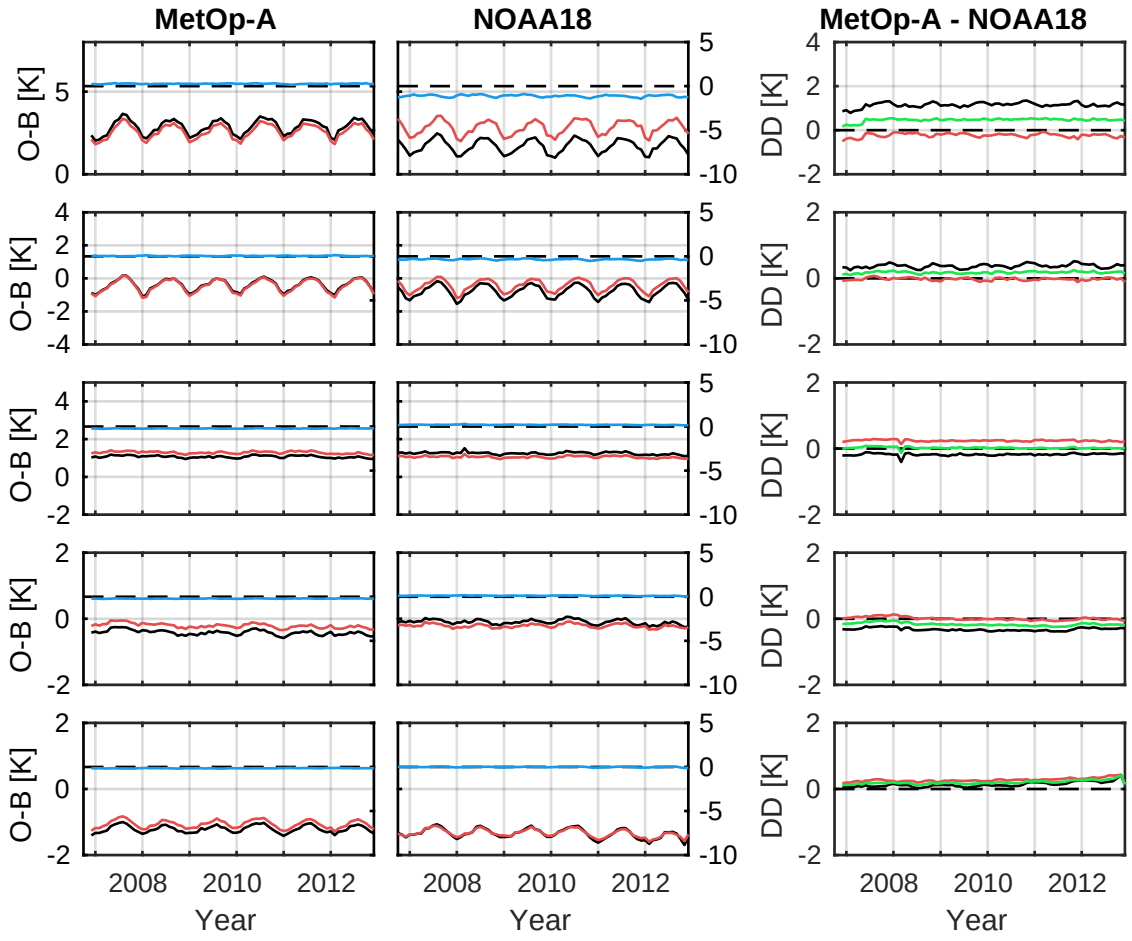


Figure 18: Same as Figure 15, but for MHS on satellite pair NOAA18 and MetOp-A.

Figure 18 depicts O-B timeseries for MetOp-A with reference to NOAA18. MetOp-A shows the opposite sign in its node bias in every channel compared to NOAA18, which was also found in Figure 14 from the NCT method and can be explained by the different ECTs of the satellites. The fact that this effect is also present in the O-Bs shows that this I-C method, too, is not capable of completely removing diurnal cycle effects for this satellite pair. Channel 1 shows very different inter-satellite biases depending on which node is compared and the bias calculated from both nodes (green line) is about 1 K larger than what was seen in Figure 14 from the NCT method. This is similar for channel 2, but to a smaller degree. This is caused by the overestimation of diurnal cycle effects in channels 1 and 2 by the reanalysis model, as was concluded already for the previous satellite pairs. The results of the SAAC method presented in the next chapter will give additional insight on what the inter-satellite bias looks like without diurnal cycle effects, which are excluded in SAACs.

The sounding channels do not show significant deviations in the overall inter-satellite biases (green lines) to the NCT method. However, the inter-satellite biases vary considerabely depending on ascending or descending node, compared to the other satellite pairs. Particularly channel 3 sticks out, showing an about 0.6 K difference between ascending and descending node inter-satellite biases, whereas for channel

5 the difference is comparatively small. The results of the NCT method showed the opposite behaviour, with a smaller difference in channel 3 compared to channel 5. Since the result of the NCT method can be physically explained, e.g. that the lower peaking channel 5 is more affected by diurnal effects than channel 3, the behaviour in the DDs is most likely caused by an overestimation of the diurnal cycle in channel 3 by the model background. A diurnal cycle related issue of the model background in channel 3 was also found in the NOAA15/NOAA16 satellite pair, already. The geographical pattern seen in Figure 19 of DD node differences for channel 3 reinforces that the issue is related to the diurnal cycle. This Figure can be understood as the geographical distribution of subtracting the red line from the black line of channel 3 DDs in Figure 18. The fact that over convectively more active regions, such as the Amazonian Basin or over South Africa, differences of ascending and descending node DDs are generally larger, shows that over those regions the model background has more trouble simulating the observations from different sampled daytimes. It is very unlikely that inter-satellite biases caused by instrumental deficiencies are geographically structured this way.

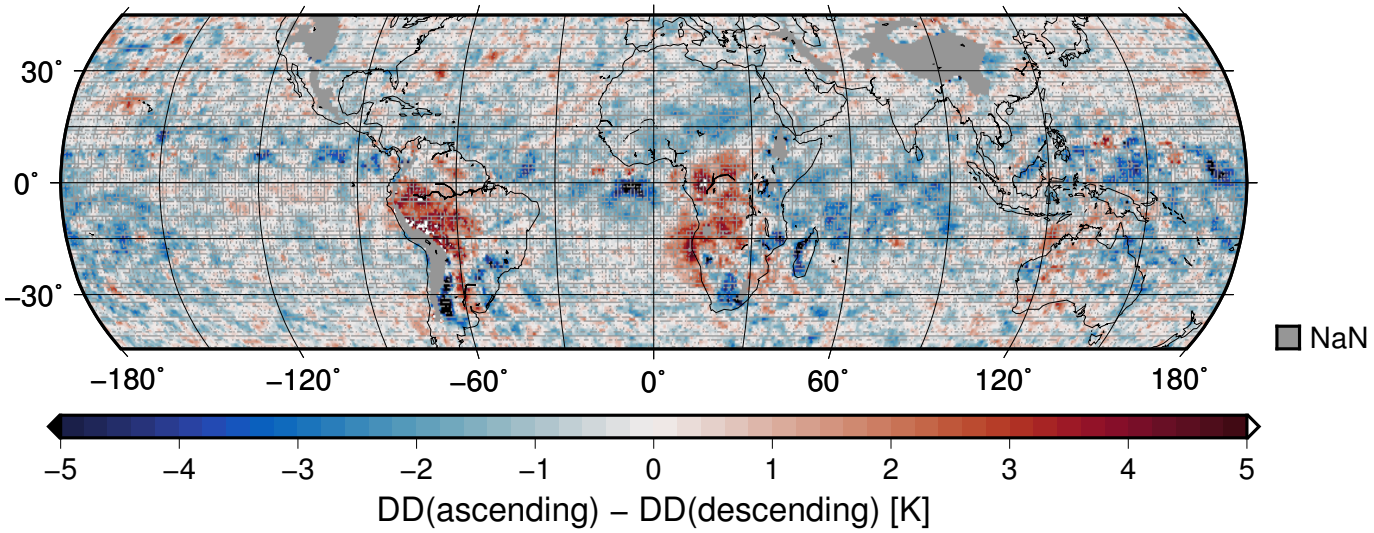


Figure 19: Map of channel 3 NOAA18/MetOp-A DD node differences, e.g. ascending minus descending node, for January 2011. Large mountainous ranges are filtered out (NaNs).

It can be concluded that O-Bs are able to reveal major instrumental issues, such as the long-time trends or jumps caused by RFI in NOAA15 and NOAA16. However, the reanalysis background is not completely capable of removing seasonal and diurnal cycle effects, introducing patterns in the O-B timeseries that could only be fully characterised by comparing the results to those of the NCT method. Particularly the results of the window channels 1 and 2 deviate strongly between the two methods, which is caused by deficincies of the reanalysis background to simulate the diurnal cycle for these channels. Although also the sounding channels still show some leftover diurnal cycle impact in the O-Bs, this impact is only in the order of 0.1 to 0.6 K, but surprisingly larger for channel 3 than for channels 4 and 5. Therefore, significant inter-satellite biases in the sounding channels in the order of several Kelvin are confidently caught by the O-Bs, but more subtle instrumental issues might be conceiled by diurnal cycle effects, especially in channel 3.

4.1.3 Simultaneous all angle collocations

In this chapter, inter-satellite biases calculated from SAACs are presented. Figure 20 shows long time-series of monthly inter-satellite biases obtained from SAACs of the three chosen satellite pairs. Also, the monthly standard deviations of the BT differences between the matchups are shown as gray error-bars. The standard deviations are not divided by the square root of the amount of data, because the

pure random uncertainty associated with each individual matchup is supposed to be reflected. Since the amount of matchups is generally in the order of 10^5 per month or higher, the random errors of the monthly means are always small anyway. Note that with this method, no distinction is made between ascending and descending node data because the diurnal cycle can not impact SAACs by definition, as described in chapter 3.3. Still, the inter-satellite biases calculated from this method can be directly compared to the overall inter-satellite biases depicted by the green lines in the previous two chapters.

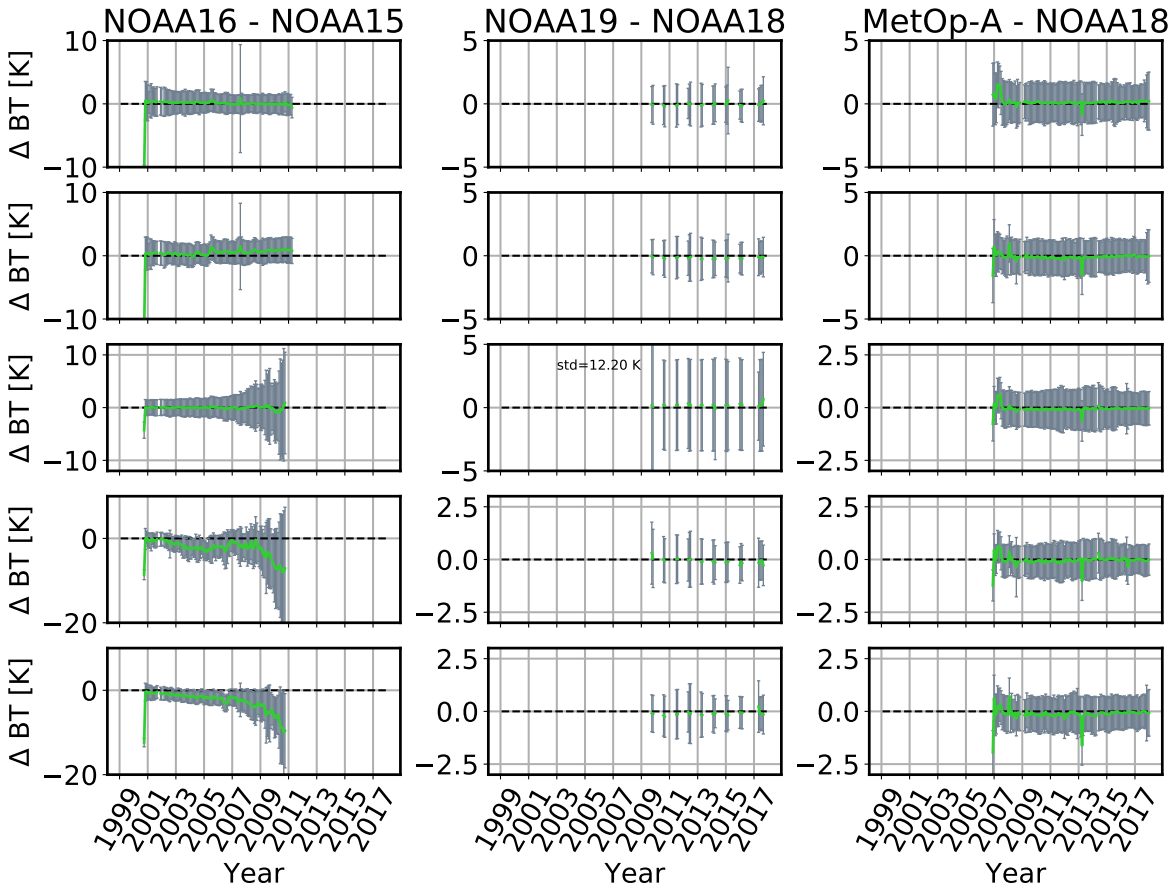


Figure 20: Long timeseries of monthly inter-satellite biases obtained from SAACs for channels 1 (top) to 5 (bottom) for three satellite pairs. Gray errorbars show standard deviations of monthly matchup BT deviations. In channel 3 of the NOAA19/NOAA18 satellite pair, the text in the figure denotes the standard deviation during the first two months of the timeseries, which is cut off on the y-axis.

The inter-satellite biases of NOAA16 and NOAA15 depicted in the left column of Figure 20 show some similar trends as already seen in figures 11 and 15 from the NCT and the O-B method. Some common trends are also found in channels 1 and 2, which is surprising, because the trends that were previously found in these channels were attributed to the combination of orbital drift and diurnal cycle effects. For channel 1, a slight negative trend is found in the SAACs, which was also seen in the NCT method with a much greater magnitude. It can be infered that the dominating effect causing the trend in channel 1 of Figure 11 is in fact the diurnal cycle, but a small contribution to the trend might also be due to an instrumental issue, as seen in Figure 20. For channel 2, a jump of about 1 K is found in 2005, which was also found by the NCT and O-B methods. After that jump a positive trend is found in the SAACs. Here, the NCT method rather showed a negative trend (Figure 11) due to diurnal cycle effects. It is interesting that the O-Bs show a positive trend in the inter-satellite bias after 2005 (Figure 15), but as concluded earlier, this is caused by an erroneous representation of the diurnal cycle by the reanalysis

model. This shows that an agreement of two independent I-C methods in terms of inter-satellite biases can be purely coincidental and one method, in this case the O-B method, gets it right for the wrong reason. The inter-satellite biases of the sounding channels 3 to 5 of NOAA15 and NOAA16 in Figure 20 show very similar trends as the O-B and NCT methods did, which are caused by RFI.

The standard deviations shown for the NOAA16/NOAA15 satellite pair are in the order of 2 to 4 K at the start of the timeseries, being larger in the surface channels than in the sounding channels. These uncertainties are caused by instrument noise and atmospheric or surface variability between matched observations. Hans et al. (2017) show that at the start of both instruments lifetime, the Noise Equivalent Differential Temperatures (NEΔTs, a measure for the sensitivity of the MW radiometer) are below 1 K for all channels, except channel 3 of NOAA15. Hence the variability seen in terms of the standard deviations at the start of the NOAA15/NOAA16 timeseries can mostly be attributed to atmospheric and surface variability between the matched observations. Standard deviations in channel 1 and 2 decrease from about 4 K in 2001 to about 2 K in 2007. Since NEΔTs of channels 1 and 2 do not change significantly over time, this is likely linked to changes in how far apart in space and time the collocations actually occur, which can change with the difference in ECTs of the two satellites. As the two satellites drift towards a common ECT (see Figure 1), the standard deviations, particularly in the surface affected channels decrease, in which spatial and temporal scales of 5 km and 5 minutes are more significant than in the mid or upper troposphere. A strikingly large standard deviation of almost 10 K is found in 2007 in channels 1, 2 and to a lesser degree in channel 5. The fact that this occurs in the two surface channels and the sounding channel most affected by the surface makes it seem like there is a physical reason behind this outlier. It could have to do with the fact that SAACs start occuring globally during that time period. However, it is unclear why only one particular month is affected. Otherwise, the increasing standard deviations in the sounding channels can be directly explained by the increasing NEΔTs of those channels, as shown by Hans et al. (2017).

The inter-satellite biases of NOAA19 and NOAA18 depicted in the middle column of Figure 20 show many large gaps, because SAACs do not occur every month for this satellite pair (see Figure 7). Still, longtime trends or biases can be identified. For channel 1, no bias or trend between the two satellites is found. The NCT method only showed a slight net trend in channel 1 due to diurnal cycle effects. For channel 2, a negative bias of about -0.3 K suddenly emerges in 2010 (Figure 20), with a slight positive trend afterwards. Interestingly, such a bias is not found in the results of the NCT or O-B methods, which rather showed slightly positive biases. The fact that this bias is only found in the SAACs indicates that the bias is only present in data that is sampled by the SAAC method and the bias is not present in the data sampled by the NCT or O-B method. The main difference of sampling is that the SAACs only cover polar latitudes while the NCT or O-B methods are constrained to -45° S to 45° N. Therefore, the bias observed in channel 2 might be correlated to the scene temperature. This will be comprehensively investigated in the next section.

In channel 3 of the NOAA19/NOAA18 comparison, a constant inter-satellite bias of about 0.3 K is found in Figure 20, which is about 0.7 K lower than in the NCT and O-B methods. Channels 4 and 5 barely show significant inter-satellite biases, although for channel 4 a positive bias of about 0.5 K would be expected, according to the results of the other I-C methods. It is known that the NOAA19/NOAA18 biases in channels 3 and 4 are caused by RFI, which is a viewing angle dependent effect. Therefore, the lower sampling of off-nadir viewing angles compared to near nadir viewing angles in the SAAC method (see chapter 3.3) appears as a plausible cause for the comparatively low magnitudes of inter-satellite biases in channels 3 and 4. However, as will be seen in the next section, also OCTMs show greater inter-satellite biases in channels 3 and 4 between NOAA19 and NOAA18, although OCTMs have the same viewing angle sampling as SAACs. Hence, the lower magnitudes of inter-satellite biases found in channels 3 and 4 between NOAA18 and NOAA19 by SAACs compared to the three other methods might be linked to a BT dependence of the bias, which will be investigated in the next section.

The standard deviations of NOAA19/NOAA18 remain below 3 K throughout the full timeseries, except

for channel 3, which shows constantly large standard deviations of about 4 K and up to 12.20 K during the first few months in 2009. These issues in channel 3 are related to the noise of NOAA19, which is strikingly large during the first few months when NOAA19 became operational ($NE\Delta T > 5\,K$) and remains at a high level afterwards ($NE\Delta T \approx 3\,K$) (Hans et al., 2017).

MetOp-A/NOAA18 has shown to be a very stable satellite pair in the NCT and O-B methods in all channels. The very different ECTs of the two satellites still led to trends and biases (Figure 14 and 18) caused by the inability of the previous methods to completely remove diurnal cycle effects. Using SAACs, Figure 20 shows no significant trends or constant biases in any channel, indicating that those trends and biases seen previously were in fact caused by diurnal cycle effects, which are eliminated in SAACs. However, for all channels the SAACs show spiky patterns of about 0.5 K to 1 K jumps in the first months of the timeseries and also at the start of 2013. Negating a first guess, these patterns do not correlate with the amount of SAACs available each month shown in Figure 7. An initial drop in inter-satellite bias at the start of 2007 was also seen in the NCT, but the subsequent irregularities are specific to the SAACs. Since the behaviour is specific to only this method, it is assumed that rather a methodological flaw, such as sampling behaviour during some months, than an instrumental issue is causing the spikes.

4.1.4 Scene temperature correlation of inter-satellite biases

In the previous section, the sampling limitation to low BTs of the SAAC method was suspected to lead to deviations in inter-satellite biases compared to the other I-C methods in two cases:

1. The bias in channel 2 between NOAA19 and NOAA18 suddenly drops to -0.3 K in 2010 in the SAACs, which does not occur in any of the other I-C methods.
2. The biases in channels 3 and 4 between NOAA19 and NOAA18 are about 0.7 K and 0.5 K lower in the SAACs compared to any of the other methods.

In the following, it is investigated whether these deviations are consequences of a BT dependence of the inter-satellite biases. A BT dependence of the bias could for example be caused by erroneously measuring the warm calibration target temperature in one of the instruments, giving the two point linear calibration equation an erroneous slope. However, note that a direct dependence of the inter-satellite bias on BT cannot be attested by simply plotting the inter-satellite bias as a function of BT, as done with BT bins by (John et al., 2012) or with percentiles in Figure 21. Such illustrations only reveal the *correlation* of inter-satellite bias and BT. One cannot directly infer the underlying cause for the found correlation of BT and bias, such as the example given above of an erroneous warm calibration target temperature. Hence, one should be careful with the word *dependence*, which usually implies a causal connection.

The following analysis is conducted by looking at monthly differences of various percentiles (Figure 21) instead of monthly differences of means (e.g. monthly biases). The data basis for this analysis is chosen to be the same as for the NCT method, but without a cloud filter, to cover also the low BTs sampled by SAACs and still use data that is independent of the SAAC data itself. The mean BT of SAACs in channels 2 and 3 generally corresponds to the BTs of the 5th percentile of the NCT data. For channel 4 it is the 1st percentile. Hence, if one of the two listed deviations in bias patterns was solely caused by a BT dependence of the bias, then the bias pattern would also be present in the differences of those percentiles of the NCT data. The fact that SAAC data and NCT data are generally independent increases confidence in this procedure, because for example SAAC specific sampling effects would not show through in the NCT data. The split into ascending and descending node data again helps distinguishing between diurnal cycle and calibration induced effects.

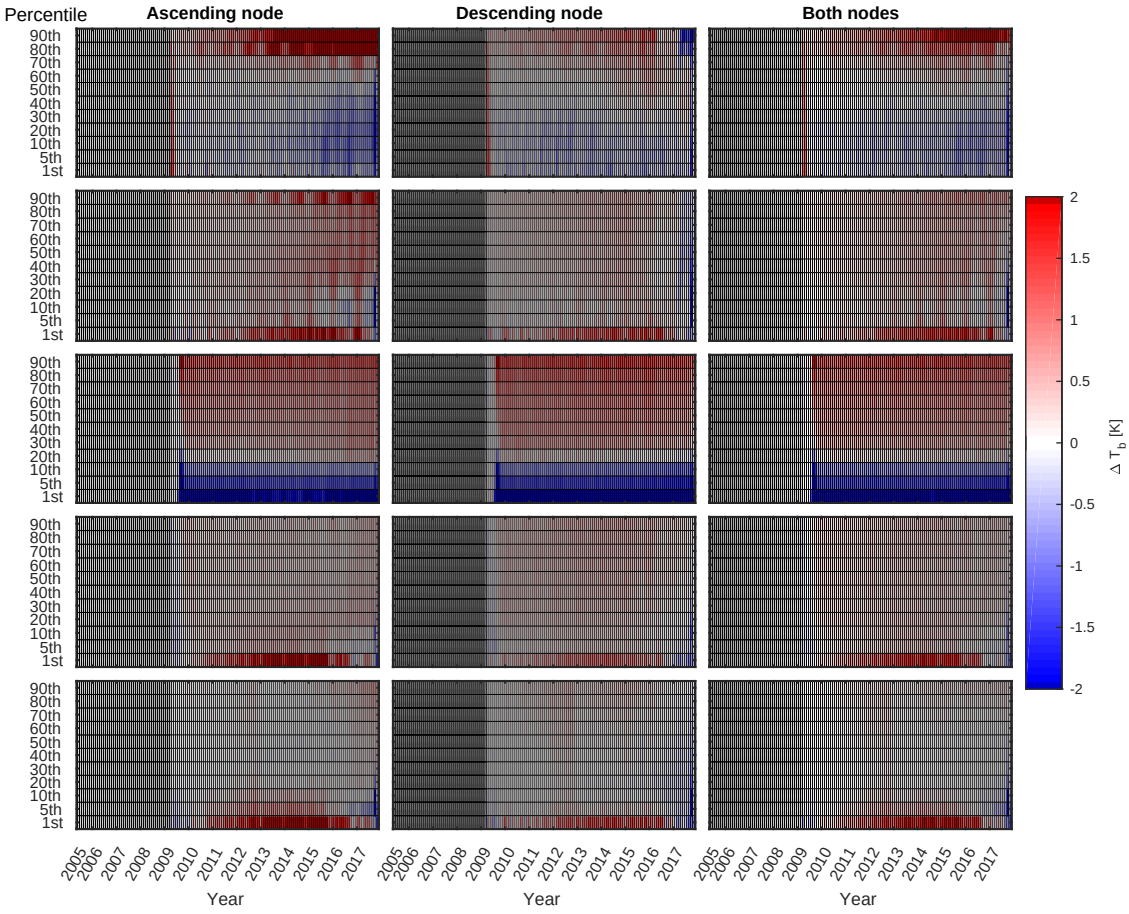


Figure 21: Differences of 10th to 90th percentiles between NOAA19 with reference to NOAA18, based on monthly NCT data, without cloud filter applied, to also obtain low BTs that correspond to SAAC data. Left, middle and right column are respectively based on the ascending node, descending node and all data of the two satellites. Rows denote channels 1 (top) to 5 (bottom). Gray areas denote months without available data.

The inter-satellite bias drop to -0.3 K listed as case 1. above is not found in the percentile differences shown in Figure 21. The inter-satellite percentile differences in channel 2 are positive in all percentiles (averaged over both nodes) and in 2010 no particular negative jump in the lower percentiles is found, which would be necessary if the jump was caused by a BT dependence of the bias. Therefore, it remains unclear what the cause for the negative bias in channel 2 from SAACs is. Nonetheless, figure 21 still shows BT correlations of the bias in channel 1. Starting in 2015, the 5th to 50th percentile differences start to decrease, which also occurs in channel 2. In channel 1, a simultaneous increase in percentile differences occurs in the higher percentiles. These trends are even more pronounced when just looking at the ascending node data (left column), indicating that the underlying cause is the diurnal cycle. Hence, this pattern just shows that the NCT methodology is not able to remove diurnal cycle effects for channel 1, which was already concluded earlier. The way this pattern emerges could be that during the daytimes sampled by NOAA19, the atmospheric states are more diverse than at the daytimes sampled by NOAA18. This would make the BT distribution of NOAA19 wider and cause the negative to positive differences from low percentiles to high percentiles as seen for channel 1.

The inter-satellite biases in channels 3 and 4 between NOAA19 and NOAA18 are caused by RFI, but are about 0.7 K and 0.5 K lower in SAACs compared to the other methods. This appears to be reflected

by the negative 5th percentile differences of channel 3 in Figure 21. However, the percentile difference pattern in channel 3 is not triggered by a scene temperature dependence of the bias, but rather by the strong noise of NOAA19, making the BT distribution wider than the one of NOAA18. This is an important realization because noise is in general not capable of changing the mean of a distribution. Therefore, the strong BT correlation of the inter-satellite bias found for channel 3 in Figure 21 is not responsible for the deviation of the bias obtained from SAACs compared to the other methods. For channel 4, no BT correlation of the bias is found, because the noise of NOAA19 is only about 0.2 K larger than that of NOAA18 (Hans et al., 2017).

It can be concluded that the few deviations of SAAC based inter-satellite biases to the biases obtained from the other I-C methods are not linked to a BT dependence of the bias. By BT dependence a true misscalibration of the instrument in the sense of an erroneous calibration point or a lacking non-linearity correction is meant. However, the inter-satellite biases are still in some cases correlated to the BT, which is caused by differences in the random uncertainty of the two compared instruments. The random uncertainty can be larger for one instrument due to natural variability or instrument noise. Because differences in random uncertainty do not impact the inter-satellite biases, the underlying cause for the deviating biases of SAACs to the other methods remains unclear. It shall be noted here, though, that these deviations actually disappear when applying a cloud filter to the SAAC datasets (see Appendix A). Due to time constraints, this cannot be assessed in much more detail. It shall only be noted, that also the application of a cloud filter on SAACs has its drawbacks, because it effectively cuts off the very cold end of the BTs and can in conjunction with large noise of one instrument introduce artificial inter-satellite biases. This issue is much more significant for SAACs than for the other methods because the BT distribution of SAACs is translocated closer to the cloud filter threshold than the other methods (see figure 23). Still, the cloud filter in conjunction with large noise is expected to artificially increase the biases obtained from each method, other than SAACs. Further investigations are necessary to quantify this issue. However, this might explain, why inter-satellite biases in the presence of large instrument noise are larger with each method in which a cloud filter is applied.

Since no clear indications of BT dependent biases were found, SAACs appear as a suitable I-C method. However, if BT dependencies of the biases existed, SAACs by themselves would not be able to capture those and therefore lead to erroneous conclusions. Complementing SAACs with OCTMs seems like the most promising approach to obtain a comprehensive matchup data basis. Inter-satellite biases obtained from OCTMs alone are investigated in the next section.

4.1.5 Opportunistic constant target matchups

Next to SAACs, a new way of directly matching individual satellite measurements from different polar orbiting satellites was introduced in section 3.4. While SAACs in general only cover rather cold BTs due to their polar restriction, OCTMs particularly sample the warm end of BTs at tropical latitudes. In Figure 22, long timeseries of OCTM based inter-satellite biases and their standard deviations are shown. As for SAACs, the standard deviations are also here not divided by the squareroot of the number of matchups.

Unfortunately, OCTMs for the NOAA16/NOAA15 satellite pair are not ready at the moment of writing, but the other two chosen satellite pairs are presented. Also biases between MetOp-A and MetOp-B are shown, which can be viewed as a reference with regard to the uncertainty associated with diurnal cycle effects, since these two satellites have exactly the same ECT. Note that OCTMs for these two satellites are only found for off nadir viewing angles due to the fact that they always cross the same location about 50 minutes apart in time. This fixed temporal deviation of simultaneous overpasses also explains why SAACs do not exist between MetOp-A and Metop-B.

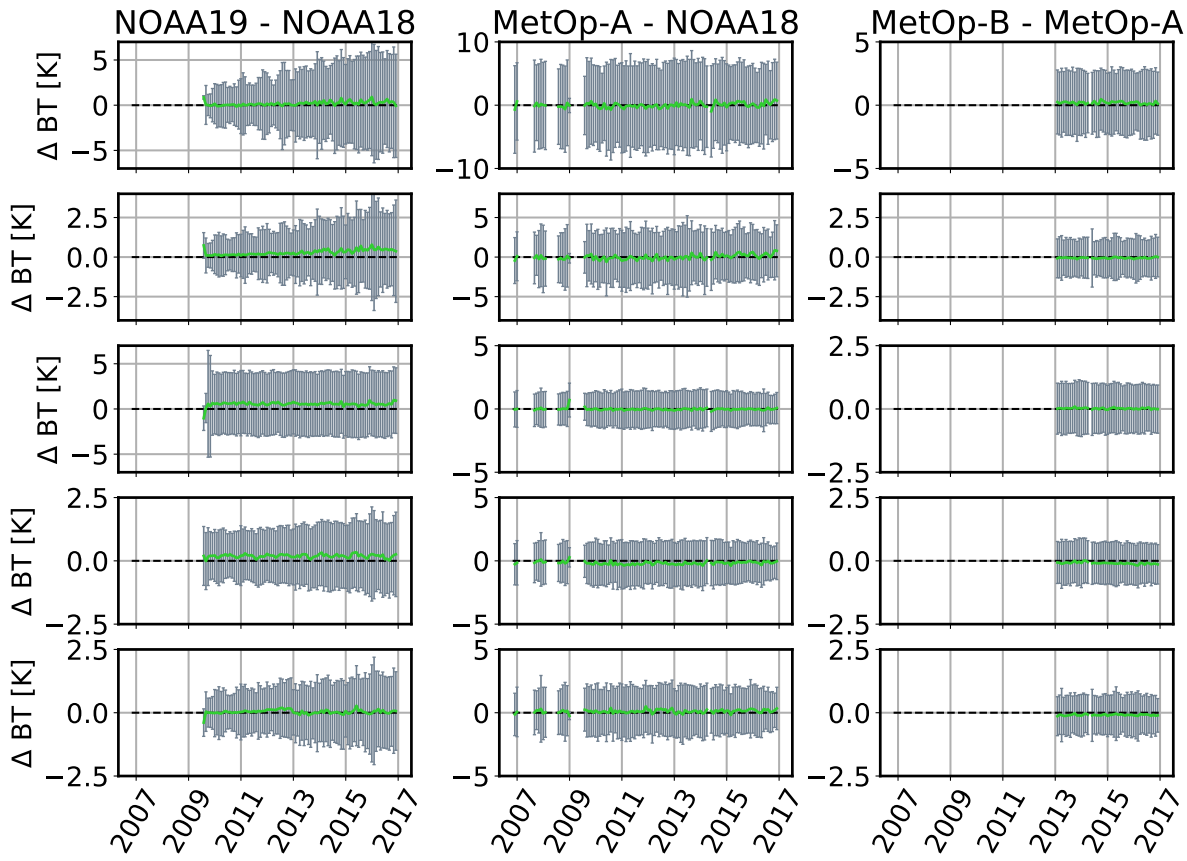


Figure 22: Long timeseries of monthly inter-satellite biases obtained from OCTMs for channels 1 (top) to 5 (bottom) for three satellite pairs. Gray bars show standard deviation of monthly matchup BT deviations. Note that the satellite pair NOAA15/NOAA16 is missing, but MetOp-A/MetOp-B is shown.

Looking at the timeseries of inter-satellite biases between NOAA19 and NOAA18 in Figure 22, it sticks out that OCTMs exist for every month, whereas SAACs only occur in about three months per year (Figure 20). OCTMs occur more regularly (not more often though) due to the relaxation of the temporal collocation criterion from 5 minutes to 8 hours, which allows satellites with very similar ECTs to match regularly, such as NOAA19 and NOAA18 or MetOp-B and MetOp-A. Regarding the inter-satellite biases presented in Figure 22, it needs to be noted that the results of OCTMs for channels 1, 2 and to a lesser extent also channel 5 need to be considered with caution, because the SEVIRI channel used to find constant BT targets is mostly representative of the upper troposphere. However, for channels 3 and 4, the methodology is expected to perform as intended.

Channels 1 and 2 of the NOAA19/NOAA18 satellite pair show slight positive shifts after 2014 with a subsequent increase in monthly variability. A slight positive trend in the results of the other methods for these satellites was only seen in the biases calculated with the NCT method in Figure 13. This was attributed to the satellite drift in conjunction with a diurnal cycle, which means that most likely OCTMs in channels 1 and 2 are also affected by diurnal cycle effects. A positive trend in channels 1 and 2 is also found in the monthly standard deviations and to a lesser extent also in channels 4 and 5. A plausible cause for this would be an increase in instrument noise. However Hans et al. (2017) did not find noise patterns in NOAA18 or NOAA19 that are correlated with the positive trend in standard deviation found here. Therefore, the trend in standard deviation found for NOAA18/NOAA19 OCTMs is most likely caused by an increase in atmospheric variability between the matched MHS overpasses due to the increasing deviation of the ECTs of the two satellites. This is indirectly supported by the

fact that in the respective SAAC timeseries (Figure 20), which are not affected by shifts in ECT, no such trends in standard deviations are found. Channels 3 and 4 show constant positive inter-satellite biases emerging directly after NOAA19 becomes operational in April 2009. In channel 3 the bias is about 0.6 K and in channel 4 it is about 0.3 K. These biases are respectively about 0.4 K and 0.2 K lower than from the NCT and the O-B methods. This is most likely explained the cloud filtering issue outlined at the end of the previous section and in Appendix A.

The inter-satellite biases between MetOp-A and NOAA18 are close to zero in all channels. Only channels 1 and 2 show small positive trends due to the orbital drift of NOAA18. What sticks out about this satellite pair, however, are the severe differences in standard deviations to the NOAA19-NOAA18 satellite pair. In channels 1 and 2, the standard deviations are about twice as large compared to NOAA19/NOAA18 with values of up to almost 10 K. Also in channel 5 standard deviations are comparatively large. The reason for these large standard deviations is the same as for the temporal increase in standard deviations for channels 1 and 2 of NOAA19/NOAA18, namely the large difference in ECT of the two satellites. MetOp-A and NOAA18 OCTMs are between about 5 and 8 hours apart in local time, depending on whether ascending with ascending node, or ascending with descending node data was matched. The criterion that the upper tropospheric water vapor channel of SEVIRI remains constant between the overpasses barely constrains which lower tropospheric or surface states can be matched. Therefore, these channels show large variability between the matched observations. On the contrary, the channel 3 standard deviations between MetOp-A and NOAA18 are surprisingly low, compared to NOAA19/NOAA18. This is caused by the constant large noise that channel 3 of NOAA19 is subject to. Hans et al. (2017) found NEΔTs of about 3 K for this channel. The strong differences in standard deviations between the channels of the MetOp-A/NOAA18 satellite pair show that the OCTM methodology is working as intended only for channels 3, 4 and arguably channel 5, but not for the surface channels.

Standard deviations in all channels of the MetOp-B/MetOp-A satellite pair are about half or less as large as for the the MetOp-A/NOAA18 satellite pair. This is because MetOp-A and MetOp-B have exactly the same ECTs, making OCTMs between these satellites always only about 50 minutes apart in daytime. This reveals that even for channels 3 to 5, a large contribution to the uncertainty of satellite pairs with deviating ECTs still comes from diurnal variability of the atmosphere. However, it is important to note that this variability is random in its nature and hence does not introduce artificial biases. This is also supported by the fact that the timeseries of inter-satellite biases do not deviate significantly compared to the other investigated methods. Therefore, inter-satellite biases can be calculated to a high degree of certainty by averaging over sufficient amounts of OCTMs.

It is interesting to consider how the presented standard deviations between the SAAC and OCTM methods compare. The random uncertainty in both methods is driven by their collocation criteria, except when instrument noise becomes anomalously high. For SAACs, the spatial and temporal variability mainly of the polar surface on scales of 5 km and 5 minutes is decisive for the introduced random uncertainty. Considering all collocation steps performed to obtain OCTMs (see Figure 10), matched MHS collocations may be up to 5 km apart, under the condition that two upper tropospheric SEVIRI measurements, that may be up to 9 km and 8 hours apart, agree to a degree of 0.8 K in BT. The associated random uncertainty with this collocation procedure is less intuitive to understand than for SAACs, because each of the three collocation steps contributes a part to the uncertainty. However, the general substitution made in terms of collocation criteria from SAACs to OCTMs is the replacement of the temporal criterion with the condition that SEVIRI observations have to agree within 0.8 K. It turns out that standard deviations of channels 1 and 2 are comparatively low in the SAACs, with values around 3 K, and varying a lot in OCTMs, with values between 3 to 10 K, depending on how much the ECTs of the compared satellites deviate. It goes to show that the collocation criterion based on the upper tropospheric SEVIRI channel does not carry over to the surface channels. In the sounding channels, however, standard deviations of SAACs and OCTMs are generally below 2.5 K, not considering

instrument noise driven effects. Hence, SAACs and OCTMs appear as similarly precise methods to obtain matchups for the sounding channels of the MW instruments. It is particularly surprising to see that OCTMs reach as low uncertainties in channels 4 and 5 as SAACs, since the SEVIRI condition was expected to only be completely valid for channel 3. This indicates that a tropical or subtropical constant upper troposphere implies that middle and lower tropospheric layers remain relatively stable, too. However, standard deviations of SAACs can be roughly halved in each channel by applying a cloud filter, which was briefly investigated and is shown in Appendix A. Due to time constraints this cannot be discussed in much detail here. But this shows, that by not applying a cloud filter, the random uncertainty is increased in the SAACs because the MW radiation emitted into space can vary on very small spatial and temporal scales in the presence of clouds. However, by averaging over sufficient amounts of collocations, these random uncertainties become negligible anyway.

4.2 Synthesis of inter-calibration method comparison

In this chapter, a concise summary of the comprehensively discussed implications of the different I-C methods is given. This is done by providing summarizing figures that feature direct comparisons of the I-C methods based on selected examples. Such examples are certain satellite pairs or time periods that demonstrate some of the main conclusions on the I-C methods. In the first section, the frequency distributions of the respective I-C methods are compared, which reflect some of the main implications of the different I-C method definitions. By implications the abilities of the I-C methods to capture variously structured biases are meant. In the second section, the respective timeseries of inter-satellite biases for the NOAA19/NOAA18 satellite pair are directly compared for each channel. This satellite pair is chosen, because it could be investigated with all four methods and showed some interesting inter-satellite biases that were already discussed in more detail in section 4.1.4. With the help of this example, some of the main conclusions from previous chapters on the performance of the different methods are reiterated. Although no new data is presented in this chapter, some subtle insights on the different methods can still be gained by the more direct way of comparing the different methods.

4.2.1 Impact of methodological definitions on the methods data structure

The different definitions of the I-C methods set the data bases upon which the satellite datasets are compared. These data bases are differently structured depending on the method. For example, SAACs are generally only found at high latitudes, while the NCT data base is constrained to latitudes between 45° N/S. Other example would be that OCTMs generally sample the subtropical subsidence regions more often than the moist convective regions or that OCTMs and SAACs sample the near nadir viewing angles more frequently than other viewing angles. An inter-satellite bias can be structured in these ways, too, which would make it show through differently between the I-C methods. The histograms in Figure 23 highlight some of the main structural differences between the methods, which shall now be summarized and interpreted in terms of their implications on the inter-satellite biases that were presented earlier.

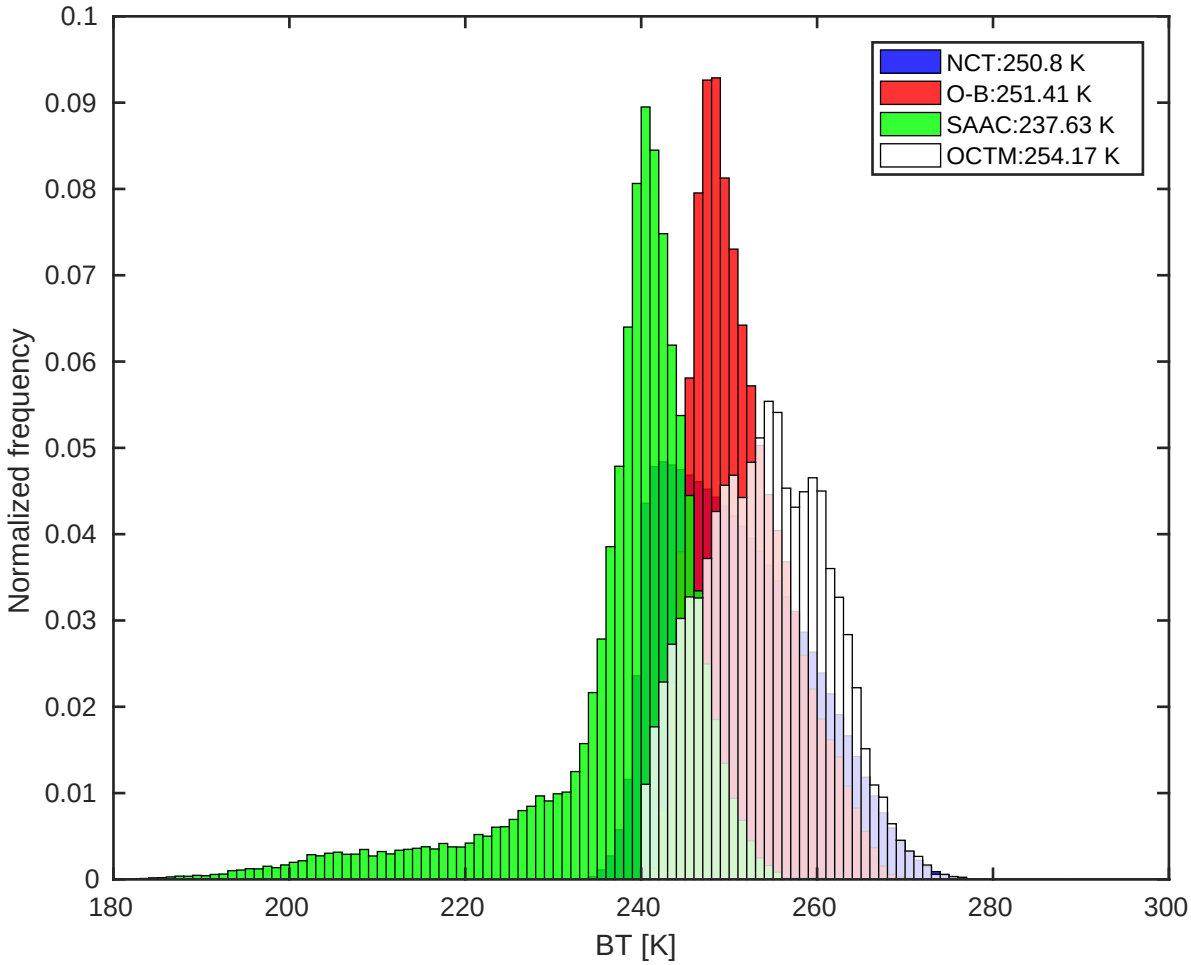


Figure 23: Histograms of channel 3 BTs of different I-C methods for one exemplary year (2009) of NOAA18 data. O-B BTs are observed values (not model background). For SAACs and OCTMs, the NOAA18 BTs from matchups with MetOp-A are taken. Temperatures in legend denote BT means of the distributions.

Figure 23 shows the normalized frequency distributions of channel 3 BTs for one year of NOAA18 data, based on the four different I-C methods investigated in this work. A systematic difference between SAACs and the three other methods is that only for SAACs no cloud filter was applied. This explains the long tail on the cold end of the SAAC distribution, while the other methods cut off around 240 K. On the other side, the SAAC distribution is limited towards the warm end due to the exclusive sampling of polar regions. As long as inter-satellite biases are not dependent on the BT, these limitations should not yield varying biases between the methods. As concluded in chapter 4.1.4, no severe BT dependencies of the biases could be found for the investigated satellite pairs. However, when assessing other satellite pairs or instruments, one should handle the limited BT ranges of the different I-C methods with caution and consider following a similar methodology as in chapter 4.1.4 to assess whether BT dependent biases are relevant.

It also goes to show how well SAACs and OCTMs turn out to complement each other, since jointly, both methods cover the largest BT range of all methods. This is because on the one hand SAACs occur in the cold polar regions and do not need to be cloud filtered and on the other hand because OCTMs are prone to sample rather high BTs in the stable dry subsidence regions. Therefore, if a matchup based I-C method is the desired way for the I-C of polar orbiting satellites, by combining the matchups from these two methods, one can in advance avoid running into issues of BT dependencies

of the bias. This way, the different structures of two methods can be used to gain an advantage. However, note that Figure 23 depicts the normalized frequency distributions. When setting up a joint matchup data basis of SAACs and OCTMs, it should be assured that similar amounts of cold SAACs and warm OCTMs are included. The example depicted in Figure 23 is based on about 50,000 SAACs and only about 24,000 OCTMs between MetOp-A and NOAA18 in 2009. Hence, some data thinning of SAACs would be suggested to obtain roughly an even sampling across BTs.

The NCT and O-B methods are based on the same geographical constraints, but still show deviations in the means and widths of their BT distributions. The difference in width is caused by the $1^\circ \times 1^\circ$ geographical averaging conducted on the O-B data, as described in chapter 3.2. The slightly higher mean of observed BTs from the O-Bs is explained by the combination of data thinning conducted on the O-Bs (e.g. including only every third viewing angle) and the limb effect. Due to the limb effect, the BT decreases with an increasing rate from nadir towards the scan edges. By an evenly spaced data thinning, the low BTs at the swath edges are undersampled relative to the high BTs near nadir, which slightly increases the mean of the O-B distribution. The differences in mean and width between NCT and O-B distribution by themselves do not directly impact the inter-satellite biases. It was also found that the data thinning conducted on the O-Bs does not impact the way RFI induced biases are perceived, which are strongly dependent on the viewing angle (see Figure 16). Still, for the purpose of I-C, it would be desirable that the data thinning was conducted on varying viewing angles from scanline to scanline (e.g. shifting the included pixels by 1 each scanline), to be completely certain that biases structured with viewing angle are caught reliably.

It can be concluded that although the I-C methods are structured differently in a lot of ways, no cases were encountered in which severe biases in the order of several Kelvin were detected by one method and missed by another. This could be very different, for example in the presence of severe BT dependent biases, which could yield significant deviations between biases from SAACs and the other methods. But the fact that no severe disagreements between the methods occured, is also owed to the fact that when defining the methods, it was attempted to make them as comparable as possible, although the ideas behind them are very different. For example, instead of considering SNOs, SAACs were a chosen I-C method, because the other methods also included all viewing angles. Also, the O-B data was purposefully constrained to the same geographical region as the NCT method to better isolate where deviating inter-satellite biases from the two methods might come from. Still, methodological flaws became apparent throughout this work, that introduced misleading trends or jumps in the inter-satellite biases. These flaws could be identified by comparing the different I-C methods. The next section summarizes the I-C method comparison by looking at previously shown long timeseries of inter-satellite biases obtained from the different methods in a comprising figure.

4.2.2 Direct comparison of biases obtained from different inter-calibration methods

The inter-satellite biases obtained from the different I-C methods were extensively discussed and comparisons between the methods were drawn throughout chapter 4.1. To provide a more concise summary of how the calibration methods compare, the inter-satellite bias timeseries of the NOAA19/NOAA18 satellite pair from the different methods are compressed into Figure 24. This helps to more directly point out some of the main conclusions regarding each methods performance and even reveals some additional subtle, but systematic bias behaviours.

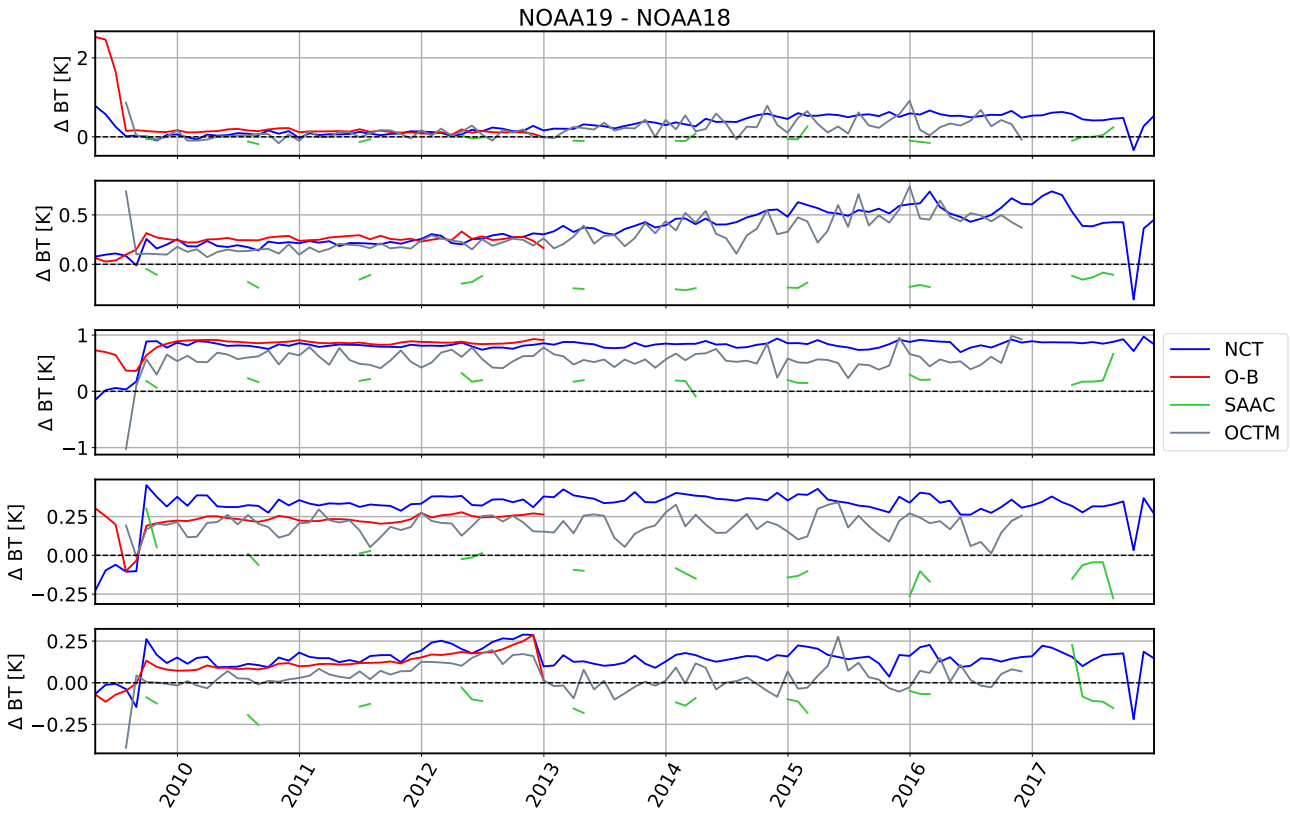


Figure 24: Timeseries of monthly inter-satellite biases between NOAA19 and NOAA18 for all four I-C methods investigated in this work. NCT and O-B data is not seperated into ascending and descending data here, but data of the full orbit is used. Panels denote channel 1 (top) to 5 (bottom).

What generally sticks out in Figure 24 is that O-B data is only available until 2013, which is purely owed to personal data availability limitations and should not be viewed as a methodological issue. On the other hand, the fact that SAACs for this satellite pair are only available for about three months per year can be viewed as a methodological issue. The sampling of SAACs strongly depends on the ECTs of the two satellites (see Figure 7), which can either conceil temporal calibration issues or oversample them, which is undesired for I-C purposes.

Looking at the actual biases calculated by the different methods, a few characterstics of the methods for channels 1 and 2 stick out. The NCT and OCTM methods both show positive trends of about 0.5 K per decade, starting in 2012, which are not found in the SAAC timeseries. These trends are caused by the insufficient ability of the NCT and OCTM methods to filter out diurnal cycle effects that impact the datasets due to orbital drifts of the satellites. For OCTMs, it was concluded earlier that the methodology cannot be expected to work for channels 1 and 2, because of the upper tropospherical SEVIRI constraint that is applied. The O-B methodology does not show trends in channels 1 and 2 for this particular satellite pair. However, trends found previously for the NOAA16/NOAA15 satellite pair by O-Bs were opposite in sign compared to the NCT method because the reanalysis model background overestimates the diurnal cycle relevant for the surface channels. This makes the O-B methodology impractical for channels 1 and 2. Regarding the performance of SAACs in channels 1 and 2, Figure 24 suggests that in channel 1 the two instruments are completely stable, which is in agreement with the other methods that only show the diurnal cycle induced signal. However, in channel 2, a negative bias of about -0.2 K is found by the SAACs, which could not be attributed to a scene temperature dependence of the bias in chapter 4.1.4 and remains inconclusive. Although this is unclear, SAACs appear as the most suited I-C method for channels 1 and 2, because it is the only method that properly excludes diurnal cycle induced effects in these channels.

In the sounding channels 3 and 4, biases are induced by the increased instrumental noise of NOAA19 in conjunction with RFI. Also, in channel 5 a bias of about 0.20 K to 0.25 K is detected by the NCT, O-B and OCTM methods, which shows a sudden drop at the start of 2013. This bias was not discussed up to now, but can by a closer look also be found in the according previous figures (Figures 13, 17 and 22). The temporal bias behaviour correlates well with the the noise evolution of NOAA18 in channel 5, as shown by Hans et al. (2017). This indicates that channel 5 of NOAA18 is also subject to a bias that depends on instrument noise, similar to RFI. However, one has to take into account that NOAA18 is the subtracted satellite, hence NOAA18 channel 5 BTs are decreasing as the noise increases. RFI related biases up to now were always linked to increasing BTs as noise also increased (e.g. NOAA15 bias shown in Figure 16). So either RFI can also occur as a destructive interference effect or a different instrumental issue causes the bias in channel 5 between NOAA18 and NOAA19. Further investigation is necessary here, which has to be postponed to future work.

It is prominent, but unclear up to now, why biases in channels 3 to 5 obtained from OCTMs are generally lower than from the NCT and O-B methods and why biases from SAACs are generally even lower. There are two known effects that contribute to this behaviour, which are not quantified yet and there might be channel specific additional effects causing these systematic deviations between the methods. The first reason is the oversampling of near nadir viewing angles compared to off nadir viewing angles caused by the pixel matching criterion applied for SAACs and OCTMs (see chapter 3.4). If the inter-satellite bias is lower at near nadir viewing angles, this can lower the obtained bias of SAACs and OCTMs. A simple solution and test for this would be to make the pixel matching criterion near nadir as strict as for the off nadir viewing angles, to obtain the same sampling of all viewing angles. The second effect that effectively causes biases from NCT and O-Bs be erroneously large, is the cut-off of low BTs by the cloud filter applied to these methods, in conjunction with differences in instrument noise of the two compared satellites. In such cases, the noise driven outliers towards the cold end are effectively removed by the cloud filter, while the warm outliers remain. This introduces an artifical bias between the two instruments, which is positive if the subtracted satellite is the one with lower noise, as for example the case in channel 3 and 4 in Figure 24. The magnitude of this effect depends on how often the cloud filter kicks out measurements that undercut the cloud filter threshold simply due to noise. For the NCT and O-B method, this is expected to occur more often than for OCTMs, because their BT distributions (depicted in Figure 23) are translocated towards lower BTs. And since such cases do not occur at all for SAACs, because no cloud filter is applied, SAACs generally show the lowest biases in the presence of a noise difference between the instruments. An exception to this explanation, however, is already found with the just described bias in channel 5 between NOAA19 and NOAA18. NOAA18 is subject to larger noise than NOAA19 from 2009 to 2013 and is the subtracted satellite, which should cause the methods with an applied cloud filter to obtain lower biases than SAACs, according to the just given explanation. However, biases from SAACs are still lowest. Hence, additional effects appear to come into play, which need to be investigated in future work to fully understand the deviations between the I-C methods.

The direct comparison of inter-satellite biases obtained from different I-C methods depicted in Figure 24 appears as a valuable tool to assess systematic deviations between the methods. Based on the previously conducted more indepth investigations in this work, some of these deviations could be attributed to their causes, such as the diurnal cycle induced effects in channels 1 and 2 or undesired interactions of instrument noise and the applied cloud filter. Also, the exclusion of a scene temperature dependence of inter-satellite biases to explain deviations between biases from SAACs and the other methods was important, because such an effect first appeared as the most likely explanation. As a next step, it should be considered how to avoid the undesired interaction of the cloud filter and instrument noise. It should be taken into account that in these cases the observational data is corrupted by the strong noise, which can not be corrected for in hindsight. Therefor, very noisy datasets might not be of value for climate applications anyway and should simply be disregarded.

5 Conclusion and outlook

In this work four I-C methods for MW humidity sounders onboard of sun-synchronous satellites were defined and their utility for capturing instrument calibration driven inter-satellite biases was assessed. These four methods are respectively based on monthly means of zonal averages between 45° N/S as a Natural Calibration Target (NCT), on Observation-Background (O-B) data from the ERA-Interim reanalysis, on Simultaneous All Angle Collocations (SAACs) of satellite pairs and on Opportunistic Constant Target Matchups (OCTMs) using SEVIRI data as an intermediate reference dataset. Such I-C methods are key for producing harmonised satellite data records that are used for climate applications. The four I-C methods investigated in this work are based on fundamentally different concepts of how to disentangle the instrument calibration induced biases from geophysically caused biases, such as the diurnal cycle. The performance of these different I-C methods was assessed by applying the methods to multiple satellite datasets and comparing the obtained inter-satellite biases between the methods.

Specific expectations and research questions were expressed in chapter 3.5. On the one hand, it had to be assessed whether the different underlying concepts to negate diurnal cycle induced biases work as intended. This assessment was realized by applying a methodology used by John et al. (2013a), in which ascending and descending passes of the satellites are treated seperately. On the other hand, it was questioned, whether differently structured biases are caught differently by the I-C methods due to their specific sampling of the satellite data. For example, viewing angle dependent biases and possible scene temperature dependencies of the inter-satellite biases were investigated.

It was found for the two surface channels that no method except SAACs is capable of removing diurnal cycle induced biases to a sufficient degree. Longterm trends caused by sampling varying phases of the diurnal cycle due to orbital drift are the predominent inter-satellite biases found in these channels, which only SAACs negate as intended. For the NCT method, this issue can be reduced if only oceanic regions were used, since the land surface is subject to a much stronger diurnal cycle than the sea surface. This is shown by results of John et al. (2013a), who used tropical oceans as a NCT. However, for the sounding channels it was shown that the NCT defined in this work yields a lower diurnal cycle impact than by using tropical oceans. Only when the ECTs of the compared satellites deviate by about 8 hours, as is the case for MetOp-A/NOAA18, do channel 5 inter-satellite biases deviate significantly by up to 1 K between ascending and descending node. The O-B method showed significant diurnal cycle related trends in channels 1 and 2 because the model background is found to overestimate the diurnal cycle in these channels. Also, channel 3 showed up to 0.6 K differences in inter-satellite biases between ascending and descending node, which are negligible in channels 4 and 5. This indicates that the model background overestimates the diurnal cycle in the upper troposphere. The diurnal cycle impact on OCTMs is found to be similar to that of the NCT method. Significant diurnal cycle induced trends in inter-satellite biases are only found in channels 1 and 2 with OCTMs, which was also expected because no collocation criterion constrains the surface variability between matched observations. However, this is something that can be expended on in the future by considering surface sensitive SEVIRI channels.

While SAACs do a good job at negating the diurnal cycle, it was suggested from earlier work that SAACs might not be able to capture inter-satellite biases that depend on the BT due to their exclusive sampling of cold BTs (John et al., 2012; Shi and Bates, 2011). Therefore, it was investigated whether inter-satellite biases from SAACs deviated from the other methods due to BT dependencies of inter-satellite biases. For NOAA19/NOAA18, only correlations of the BT and the inter-satellite bias were found, that could be attributed to differences in noise of the compared instruments. No direct dependencies between BT and inter-satellite bias were found that could explain deviations of the SAAC biases to the other methods. The most plausible explanation for deviating biases between SAACs and the other methods, is an issue with the cloud filter in the presence of instrument noise. By excluding cloud-free scenes that happen to undercut the cloud filter threshold, an artificial warm bias is introduced to noisy datasets. The significance of this effect was shown for SAACs, by comparing the obtained inter-satellite

biases with and without an applied cloud filter, which yielded bias differences of up to 1 K in channel 3 of NOAA19/NOAA18. However, the magnitude of this effect on the other methods is not quantified yet.

It can be concluded that for the sounding channels all investigated I-C methods generally perform as intended, with slight diurnal cycle issues in channel 5 with the NCT and OCTM methods and in channel 3 with the O-B method. For the surface channels, only SAACs are found to be a suitable I-C method, but OCTMs can most likely be adapted to also work for these channels by utilizing different SEVIRI channels, which presents a future task. Although no evidence for BT dependent biases was found, such dependencies might still exist for other instruments. Therefore, merging the SAAC and OCTM datasets appears most reasonable for future applications of these I-C methods. An issue with the cloud filtering method was identified, which introduces a warm bias to noisy satellite datasets. To circumvent this issue in the future, the most consequent step would be to reject such noisy datasets from further analysis, in general, because the individual measurements are irreversibly corrupted by the noise anyway. Alternative considerations of finding a cloud filter that would circumvent this issue do not appear worthwhile, knowing that the measurements are in bad condition anyway.

For the purpose of future I-C studies, for example of additional or reprocessed satellite datasets, the I-C methods defined in this work provide a solid toolbox. Only O-Bs are not easily produced for additional datasets, since they would require the full reanalysis system. However, especially NCTs offer a quick way of calculating inter-satellite biases due to their simple data filters. Also SAAC and OCTM datasets can be computed with relative ease with the help of a suited collocation toolkit (one is provided with Typhon (2019-02-18)). Possible future refinements to the methods were pointed out, particularly to increase their performance in the surface channels and further adjustments may come up in the future.

References

Nigel C. Atkinson. Calibration, monitoring and validation of amsu-b. *Elsevier Science*, 28(1):117–126, 2001.

S. A. Buehler, M. Kuvatov, T. R. Sreerekha, V. O. John, B. Rydberg, P. Eriksson, and J. Notholt. A cloud filtering method for microwave upper tropospheric humidity measurements. *Atmospheric Chemistry and Physics*, 7(21):5531–5542, nov 2007. doi: 10.5194/acp-7-5531-2007.

Martin Burgdorf, Stefan A. Buehler, Theresa Lang, Simon Michel, and Imke Hans. The moon as a photometric calibration standard for microwave sensors. *Atmospheric Measurement Techniques*, 9 (8):3467–3475, aug 2016. doi: 10.5194/amt-9-3467-2016.

Changyong Cao, Michael Weinreb, and Hui Xu. Predicting simultaneous nadir overpasses among polar-orbiting meteorological satellites for the intersatellite calibration of radiometers. *Journal of Atmospheric and Oceanic Technology*, 21(4):537–542, apr 2004. doi: 10.1175/1520-0426(2004)021<0537:psnoap>2.0.co;2.

D. P. Dee and S. Uppala. Variational bias correction of satellite radiance data in the ERA-interim reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 135(644):1830–1841, oct 2009. doi: 10.1002/qj.493.

D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart. The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597, apr 2011. doi: 10.1002/qj.828.

EUMETSAT. What is mhs?, 2019-02-18. URL `https://www.eumetsat.int/website/home/Satellites/CurrentSatellites/Metop/MetopDesign/MHS/index.html`.

FIDUCEO. Vocabulary - fcdr, 2019-01-10. URL `http://www.fiduceo.eu/`.

Eric Gawiser and Joseph Silk. The cosmic microwave background radiation. *Physics Reports*, 333-334: 245–267, aug 2000. doi: 10.1016/s0370-1573(00)00025-9.

Imke Hans. *Towards a new fundamental climate data record for microwave humidity sounders based on metrological best practice*. PhD thesis, Universität Hamburg, Von-Melle-Park 3, 20146 Hamburg, 2018. URL `http://ediss.sub.uni-hamburg.de/volltexte/2018/9454`.

Imke Hans, Martin Burgdorf, Viju O. John, Jonathan Mittaz, and Stefan A. Buehler. Noise performance of microwave humidity sounders over their lifetime. *Atmospheric Measurement Techniques*, 10(12): 4927–4945, dec 2017. doi: 10.5194/amt-10-4927-2017.

Viju O. John, Gerrit Holl, Stefan A. Buehler, Brett Candy, Roger W. Saunders, and David E. Parker. Understanding intersatellite biases of microwave humidity sounders using global simultaneous nadir overpasses. *Journal of Geophysical Research: Atmospheres*, 117(D2):n/a–n/a, jan 2012. doi: 10.1029/2011jd016349.

Viju O. John, Richard P. Allan, William Bell, Stefan A. Buehler, and Ajil Kottayil. Assessment of intercalibration methods for satellite microwave humidity sounders. *Journal of Geophysical Research: Atmospheres*, 118(10):4906–4918, may 2013a. doi: 10.1002/jgrd.50358.

Viju O. John, Gerrit Holl, Nigel Atkinson, and Stefan A. Buehler. Monitoring scan asymmetry of microwave humidity sounding channels using simultaneous all angle collocations (SAACs). *Journal of Geophysical Research: Atmospheres*, 118(3):1536–1545, feb 2013b. doi: 10.1002/jgrd.50154.

Shinya Kobayashi, Paul Poli, and Viju O. John. Characterisation of special sensor microwave water vapor profiler (SSM/t-2) radiances using radiative transfer simulations from global atmospheric reanalyses. *Advances in Space Research*, 59(4):917–935, feb 2017. doi: 10.1016/j.asr.2016.11.017.

Isaac Moradi, Philip Arkin, Ralph Ferraro, Patrick Eriksson, and Eric Fetzer. Diurnal variation of tropospheric relative humidity in tropical regions. *Atmospheric Chemistry and Physics*, 16(11):6913–6929, jun 2016. doi: 10.5194/acp-16-6913-2016.

OSCAR. Noaa-18, 2019-02-18. URL `https://www.wmo-sat.info/oscar/satellites/`.

Lei Shi and John J. Bates. Three decades of intersatellite-calibrated high-resolution infrared radiation sounder upper tropospheric water vapor. *Journal of Geophysical Research*, 116(D4), feb 2011. doi: 10.1029/2010jd014847.

A. J. Simmons, P. Poli, D. P. Dee, P. Berrisford, H. Hersbach, S. Kobayashi, and C. Peubey. Estimating low-frequency variability and trends in atmospheric temperature using ERA-interim. *Quarterly Journal of the Royal Meteorological Society*, 140(679):329–353, jan 2014. doi: 10.1002/qj.2317.

Typhon. Tools for atmospheric research, 2019-02-18. URL `https://github.com/atmtools/typhon`.

Cheng-Zhi Zou and Wenhui Wang. Intersatellite calibration of AMSU-a observations for weather and climate applications. *Journal of Geophysical Research: Atmospheres*, 116(D23):n/a–n/a, dec 2011. doi: 10.1029/2011jd016205.

Appendix A: SAACs with an applied cloud filter

The inter-satellite biases shown by the appended Figure 25 can be compared to those of Figure 20 to assess the impact of applying a cloud filter to the SAAC data. Particularly inter-satellite biases in channels 3 and 4 of NOAA19/NOAA18 are significantly larger when a cloud filter is applied. This only occurs when one of the two compared instruments is subject to significantly larger noise than the other instrument. This is due to artificially cutting off the negative noise fluctuations with the cloud filter, while leaving the positive fluctuations untouched. This effect is undesired, but explains some of the observed differences in inter-satellite biases from SAACs, where no cloud filter is applied, compared to the other methods, for which a cloud filter is applied.

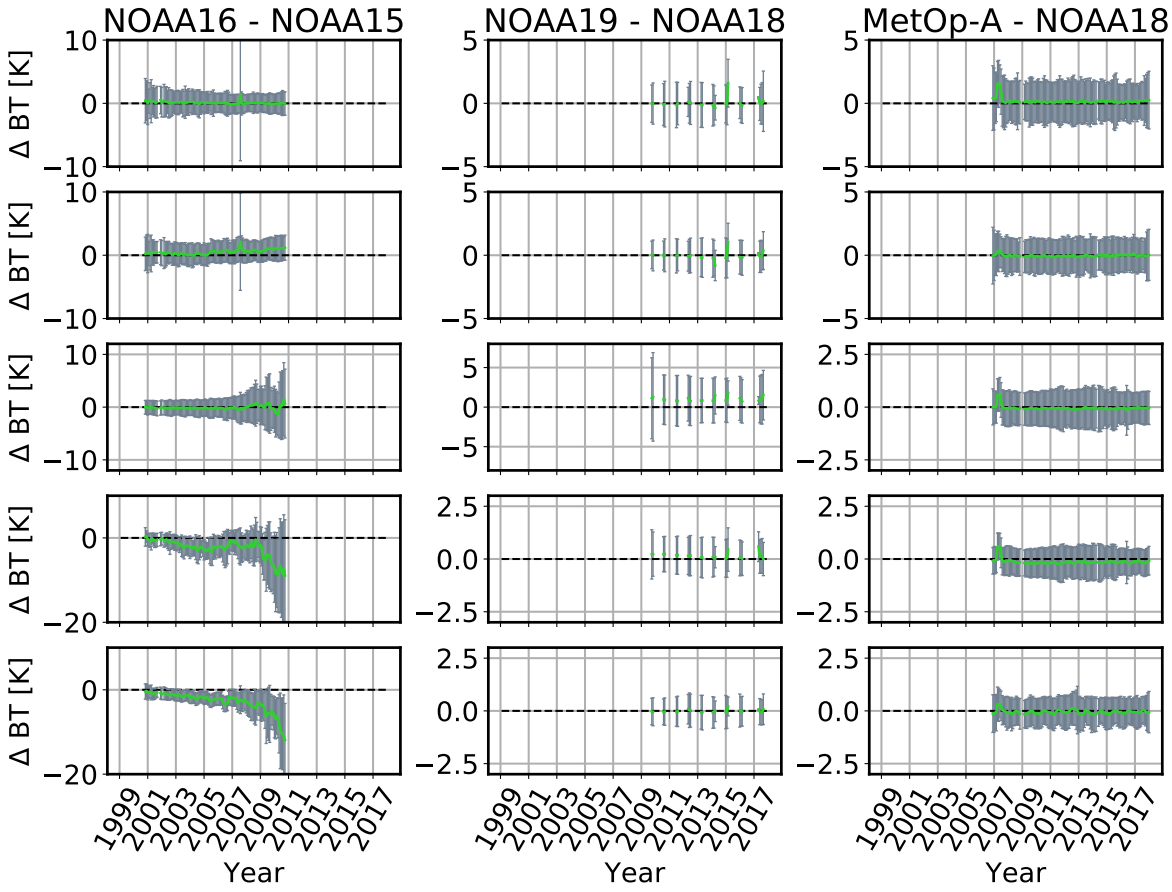


Figure 25: Same as Figure 20, but with cloud filter of Buehler et al. (2007) applied.

Versicherung an Eides statt

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Hilfsmittel - insbesondere keine im Quellenverzeichnis nicht genannten Internet-Quellen - benutzt habe. Zudem versichere ich, die vorliegende Arbeit vorher nicht in einem anderen Prüfungsverfahren eingereicht zu haben und dass die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht. Hiermit genehmige ich die Veröffentlichung dieser Arbeit in der entsprechenden Fachbibliothek der Universität Hamburg.

Datum: 20.02.2019 Unterschrift: