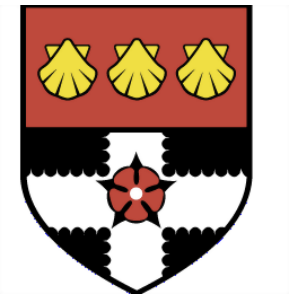
UNIVERSITY OF READING

Department of Meteorology



# Investigating the use of early satellite data to test historical reconstructions of sea surface temperature

Thomas W. Hall

A thesis submitted for the degree of Doctor of Philosophy

April 2020

# DECLARATION

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

T. W. Hall

# Acknowledgements

I am, first of all, extremely grateful to my supervisor Chris Merchant for your continued support and guidance during my PhD. You have helped me gain confidence in my work. I am also grateful to Liz Kent, David Berry, Helen Brindley and Richard Bantges for many valuable discussions we have had. Thanks also to David Ferreira and Ed Hawkins for the useful feedback in my monitoring committee meetings. I also want to thank Tom Eldridge for his help on the work related to the THIR.

I would also like to say thanks to my past and present officemates: Davi, Helene, Ying Ying, Will, Michael, Mark and Tom. You have all helped to make the last four years an enjoyable experience. I will remember all of the office trips in particular!

Last but not least, a special thanks to my parents, Val and Peter, for their constant support and encouragement.

# Abstract

An accurate record of global sea surface temperature (SST) is crucial for understanding the evolution of Earth's climate. Before the satellite era, developing a global picture of SSTs relies on reconstructions from relatively sparse, ship-based measurements. This thesis investigates the degree to which useful estimates of SST can be obtained from measurements made by the Infrared Interferometer Spectrometer (IRIS), for the purpose of testing the reliability of historical reconstructions of SST in the 'pre-satellite' era.

Global measurements of infrared spectral radiance were made by IRIS on board the Nimbus-4 satellite over a 10 month period in 1970-71. The measurements suffer from a systematic cold bias of 1-2 K in many channels relevant to SST retrieval. We use a combination of radiative transfer simulations and measurements from other instruments onboard Nimbus-4 to characterise the bias in a spectrally resolved sense, and apply this bias correction to the IRIS data. Our bias estimate agrees well with independent estimates derived from comparisons of IRIS with more recent hyperspectral sounders, in the spectral regions relevant to SST retrieval. However, the bias has more complex characteristics than previously understood, so further research is needed to fully quantify it.

An optimal estimation (OE) based SST retrieval scheme is then developed for IRIS. The retrieval scheme is tested on IASI measurements, made 'IRIS-like' by imposing the same error characteristics as IRIS. Drifting buoy SSTs are used to validate the retrieved IRIS-like IASI SSTs. We find that, in theory, SSTs can be retrieved from IRIS to an accuracy of 0.4-0.5 K in conditions known, a priori, to be clear-sky and 0.8-1.0 K in cases where the expected cloud fraction is up to 20 %. The SSTs retrieved from the latter type of cases suffer from cloud-related biases.

The retrieval scheme is applied to $\approx$1000 IRIS spectra identified as cloud-free and $\approx$70,000 spectra with an expected cloud fraction of up to 20 %. The retrieved SSTs

from each set of cases are used to produce datasets of monthly, gridded IRIS SSTs. The IRIS SSTs are compared with HadISST2 and HadSST3 to assess their level of agreement, given their respective uncertainties. There is found to be a lack of overall agreement between the IRIS SSTs and the in-situ based SST datasets for both the clear-sky and partly cloud-affected IRIS SSTs. Based on the clear-sky IRIS SSTs, this is a potential indicator that the uncertainties in the reconstructions are underestimated. However, biases in the partly cloudy IRIS SSTs remain too large for a robust test of the reliability of the reconstructions. Further testing of the retrieval scheme is required to investigate if these biases can be reduced to an acceptable level.

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# NOMENCLATURE

| | |
|---|---|
| $20CR$ | 20th Century Reanalysis |
| $AI$ | Artificial Intelligence |
| $AIRS$ | Atmospheric Infrared Sounder |
| $AMJ$ | April-June |
| $AOD$ | Aerosol Optical Depth |
| $ATSR$ | Along Track Scanning Radiometer |
| $AVHRR$ | Advanced Very High Resolution Radiometer |
| $BT$ | Brightness Temperature |
| $C3S$ | Copernicus Climate Change Service |
| $CERA-20C$ | ECMWF Coupled Reanalysis of the 20th Century |
| $CF$ | Cloud Fraction |
| $CFC$ | Chlorofluorocarbon |
| $COBE-SST$ | Centennial in situ Observation-Based Estimates SST dataset |

| | |
|---|---|
| $ECMWF$ | European Centre for Medium-Range Weather Forecasts |
| $ECV$ | Essential Climate Variable |
| $ENSO$ | El Niño-Southern Oscillation |
| $EOF$ | Empirical Orthogonal Function |
| $EOT$ | Empirical Orthogonal Teleconnection |
| $ERA-20C$ | ECMWF Twentieth Century Reanalysis |
| $ERA-40$ | ECMWF 45-year Reanalysis |
| $ERA-Interim$ | ECMWF Reanalysis from 1979 |
| $ERA5$ | Fifth generation of ECMWF atmospheric reanalyses |
| $ERI$ | Engine Room Intake |
| $ERSST$ | Extended Reconstructed Sea Surface Temperature dataset |
| $ESA$ | European Space Agency |
| $FOR$ | Field of Regard |
| $FOV$ | Field of View |
| $FWHM$ | Full Width at Half Maximum |
| $GADS$ | Global Aerosol Data Set |
| $GCOS$ | Global Climate Observing System |
| $HadISST$ | Hadley Centre Sea Ice and Sea Surface Temperature dataset |
| $HadSST$ | Hadley Centre SST dataset |
| $HRIR$ | High Resolution Infrared Radiometer |

| | |
|---|---|
| $IASI$ | Infrared Atmospheric Sounding Interferometer |
| $ICOADS$ | International Comprehensive Ocean-Atmosphere Data Set |
| $IDCS$ | Image Dissector Camera System |
| $IFOV$ | Instantaneous Field of View |
| $IMCC$ | Image Motion Compensation and Calibration |
| $IMG$ | Interferometric Monitor for Greenhouse Gases |
| $IR$ | Infrared |
| $IRIS$ | Infrared Interferometer Spectrometer |
| $ISA$ | International Standard Atmosphere |
| $JRA-55$ | Japanese 55-year reanalysis |
| $L1$ | Level 1 |
| $LBLRTM$ | Line-by-Line Radiative Transfer Model |
| $MAP$ | Maximum a posteriori |
| $MRIR$ | Medium Resolution Infrared Radiometer |
| $NEDT$ | Noise Equivalent Differential Temperature |
| $NER$ | Noise Equivalent Radiance |
| $NMAT$ | Night Marine Air Temperature |
| $NWP$ | Numerical Weather Prediction |
| $OE$ | Optimal Estimation |
| $OI$ | Optimal Interpolation |

| | |
|---|---|
| $PCA$ | Principal Component Analysis |
| $PCRTM$ | Principal Component-based Radiative Transfer Model |
| $PDF$ | Probability Density Function |
| $Q-Q$ | Quantile-quantile |
| $QA$ | Quality Assurance |
| $QC$ | Quality Control |
| $RMS$ | Root Mean Square |
| $RSOI$ | Reduced-Space Optimal Interpolation |
| $RTM$ | Radiative Transfer Model |
| $RTTOV$ | Radiative Transfer for TOVS |
| $SAA$ | South Atlantic Anomaly |
| $SRF$ | Spectral Response Function |
| $SST$ | Sea Surface Temperature |
| $SST\ CCI$ | Sea Surface Temperature Climate Change Initiative |
| $SVD$ | Singular Value Decomposition |
| $TCWV$ | Total Column Water Vapour |
| $THIR$ | Temperature Humidity Infrared Radiometer |
| $TIROS$ | Television Infrared Observation Satellite |
| $ToA$ | Top-of-Atmosphere |
| $TOVS$ | TIROS Operational Vertical Sounder |
| $VBPCA$ | Variational Bayesian Principal Component Analysis |

# CHAPTER 1

# INTRODUCTION

## 1.1 Motivation

Since the 1950s, the Earth's climate has changed at a rate unprecedented over timescales of decades to millennia (*IPCC*, 2013). A variety of atmospheric and surface observations are needed to comprehensively monitor these changes in the state of the climate. Observations of sea surface temperature (SST) form one of the key components of the climate record and SST is one of the Essential Climate Variables (ECVs), as defined by the Global Climate Observing System (GCOS) (*GCOS*, 2009). SST also provides the lower boundary forcing for atmospheric circulation models (*Wang*, 2004).

There are a number of different in-situ based datasets of SST extending back over 150 years. Figure 1.1 compares the global annual average SST (and night marine air temperature (NMAT)) anomaly from the main datasets during 1850-2012. Over most of the record, the SST reconstructions agree to within 0.1°C. However, they are not truly independent of each other because the observations they are based on are largely the same (*Berry et al.*, 2018). Datasets of SST retrieved from satellite radiometers exist for the 1980s onwards, providing an independent record of SST. Before this, SST reconstructions are based on ship-based measurements with highly variable coverage.

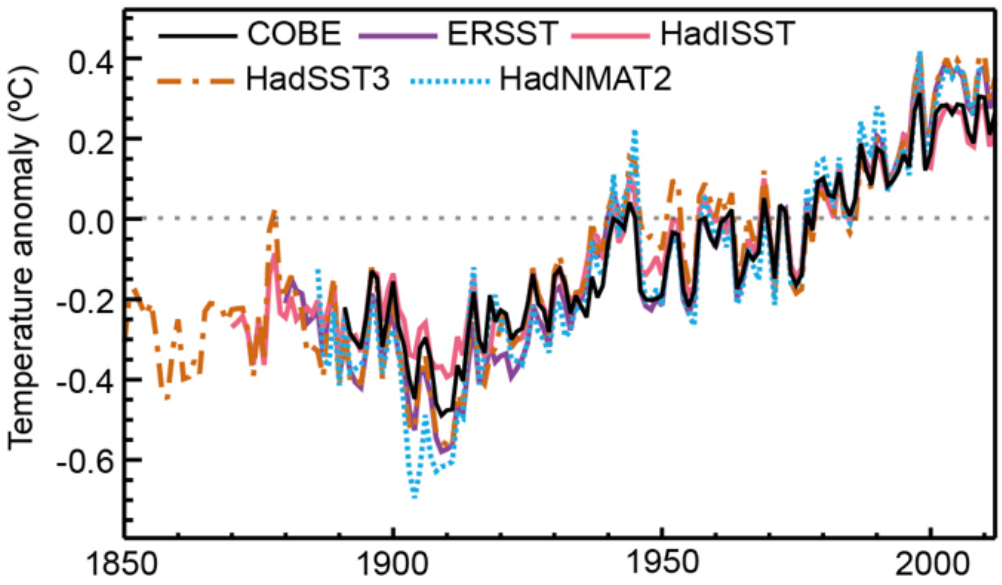



Figure 1.1: Global annual average sea surface temperature (SST) and Night Marine Air Temperature (NMAT) relative to a 1961-1990 climatology from different reconstructions. Spatially interpolated products are shown by solid lines and non-interpolated products by dashed lines. Figure reproduced from *IPCC* (2013), Figure 2.18.

Figure 1.2 shows the percentage of global ocean sampled annually by SST measurements from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS). ICOADS is the most extensive freely available archive of global surface marine data (*Freeman et al.*, 2017) and so is the preferred source for constructing historical SST analyses (*Kent et al.*, 2017). Figure 1.2 shows that the percentage coverage never exceeds about 60 %, even in the modern period. Many of the SST datasets are spatially interpolated, meaning that statistical methods are used to estimate the SSTs in regions with no observations. These datasets are referred to as reconstructions.

There were meteorological measurements being made from satellites in the 1960s and 70s, some of which can potentially be used to retrieve SST with near-global coverage. This can provide an independent SST record to test the reliability of the in-situ based reconstructions during this early period.

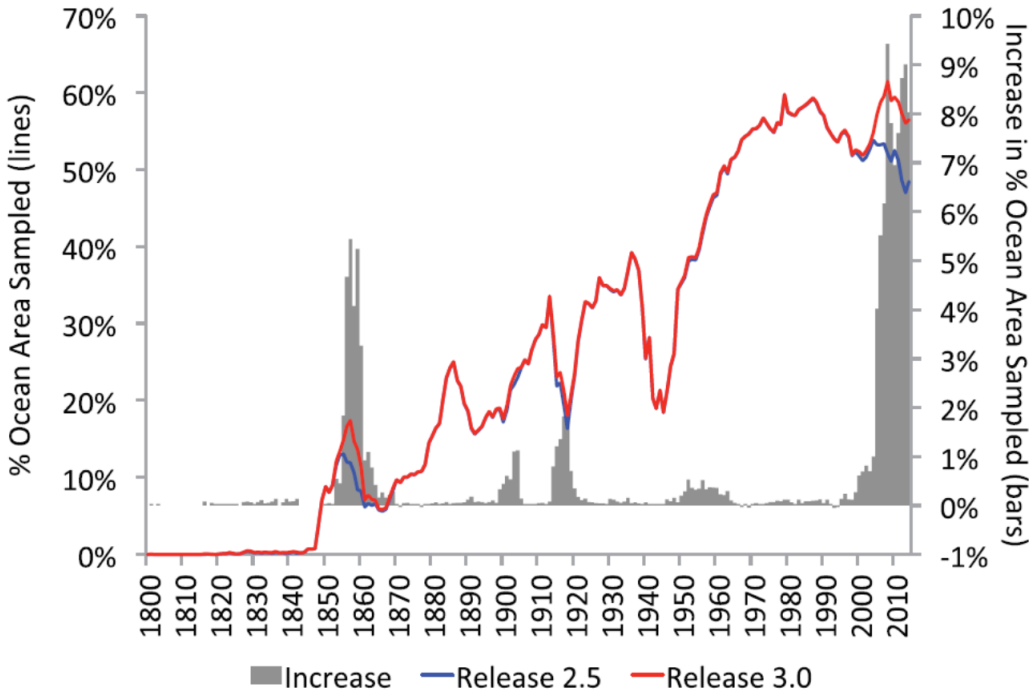



Figure 1.2: Annual percentage of global ocean and coastal area sampled for SST from ICOADS R3.0 (red curve) and R2.5 (blue curve, left axis). Annual percentage increase (or decrease) in global ocean and coastal area sampled for R3.0, compared to R2.5 (bars, right axis). Figure adapted from *Freeman et al.* (2017).

## 1.2 Thesis aim and objectives

The primary aim of this thesis is to investigate if SST can be retrieved from one of these early satellite instruments, to test the reliability of the SST reconstructions in the early 1970s. This instrument is the Infrared Interferometer Spectrometer (IRIS), which made measurements of atmospheric emission spectra on-board the Nimbus-4 satellite from April 1970 to January 1971. The IRIS instrument was designed primarily to provide information about the vertical structure of the atmosphere and surface emissivity *Sabatini* (1970). However, there is evidence from previous studies *Prabhakara et al.* (1974); *Iacono and Clough* (1996) that useful SST information can also be obtained from IRIS. The aim of this thesis can be achieved by fulfilling the following objectives.

1. Obtain IRIS dataset and apply quality assurance (QA). This includes identifica-

tion of biases in the data and bias correction.

2. Develop a method to retrieve SST from IRIS observations that is tolerant of partial cloud.

3. Test retrieval method on modern spectral data from the Infrared Atmospheric Sounding Interferometer (IASI).

4. Retrieve IRIS SSTs and process into a gridded dataset.

5. Compare gridded IRIS SSTs to gridded in-situ based datasets and assess level of agreement.

## 1.3 Thesis outline

Chapter 2 provides a review of relevant scientific literature. It starts with an overview of the historical in-situ based SST reconstructions and background on the process of retrieving SST from thermal satellite measurements. The remainder of the chapter takes a detailed look at the IRIS instrument and gives an account of past attempts to retrieve SST from IRIS.

Chapter 3 discusses the quality control (QC) tests applied to the dataset of IRIS spectra in order to remove erroneous spectra. It also explains how IRIS measurements over land and sea ice are identified and removed as these are not relevant to SST retrieval.

Chapter 4 is concerned with characterising a spectral bias in the IRIS measurements. Observation-simulation comparisons and comparison of IRIS with another sensor on-board the same satellite are used to characterise the bias and the resulting bias model is used to correct the IRIS measurements.

Chapter 5 discusses how we develop a method of retrieving SST from IRIS, both in clear-sky and partly cloudy conditions. The remainder of the chapter explains how

the retrieval method is tested on 'IRIS-like' IASI spectra, with the retrievals validated using quality-controlled drifting buoy SSTs.

Chapter 6 presents the results obtained from retrieving SST from IRIS. The process of constructing a monthly gridded dataset of IRIS SSTs is described and comparisons are made between the IRIS SSTs and in-situ based SST reconstructions to assess if there is agreement to within their stated uncertainties.

In chapter 7, our findings from the previous chapters are discussed and suggestions of potential future work are made.

# CHAPTER 2
# LITERATURE REVIEW

This chapter discusses a number of topics that provide the relevant background in this project. We start with a brief look at the role of SST in the climate system. This is followed by an overview of the historical SST reconstructions; how they compare and the sources of uncertainty. Section 2.3 then explains the process of retrieving SST from thermal satellite measurements, including radiative transfer theory and how to handle cloud. Section 2.4 provides a summary of early satellite SST retrieval which leads into a more detailed look at IRIS in section 2.5; its operation, spectral properties and previous attempts to retrieve SST from IRIS.

## 2.1 The role of SST in the climate system

Sea surface temperature plays a key role in the climate system, for it controls the exchange of heat, momentum, salt and gases between the ocean and the atmosphere (*Emery*, 2014). Figure 2.1 shows a diagram of the Earth's global mean energy balance (*Kiehl and Trenberth*, 1997; *Le Treut et al.*, 2007). About half of the total incoming solar radiation is absorbed by the surface with 80 % of that being absorbed by the oceans (*Hay*, 1996). About two-thirds of the energy received by the atmosphere is

transferred from the surface via conduction and convection (thermals), through latent heat (evapotranspiration) and longwave radiation. This highlights the importance of surface-to-atmosphere energy fluxes in the climate system, as they are a major driver of weather systems.

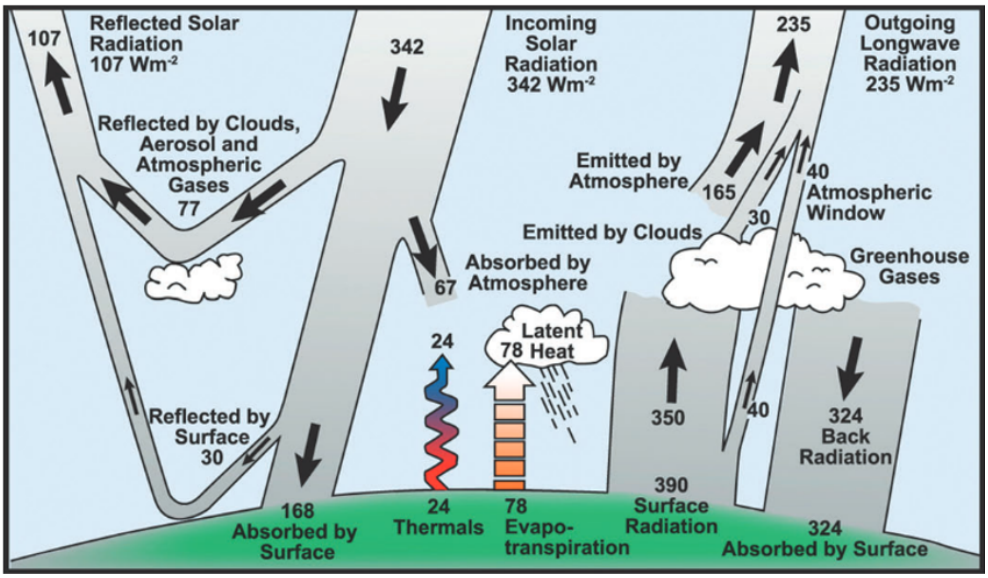



Figure 2.1: Earth's annual global mean energy budget. Units are in $\mathrm{W\,m^{-2}}$. Reproduced from *Le Treut et al.* (2007) (originally from *Kiehl and Trenberth* (1997).)

## 2.2 Historical reconstructions of SST

Historical SST reconstructions combine SST measurements made by different methods to produce datasets of global, gridded SST covering the past 100 years or more. There are a number of different reconstructions from different research institutions. Characteristics of the main reconstructions are given in Table 2.1, as well as for HadSST3 which is not a true reconstruction. This shows that they differ in terms of spatial resolution and reconstruction methods. HadSST3 differs from the other three reconstructions in that it is uninterpolated so there are missing grid values. A two stage reconstruction process is used in HadISST2: Reduced-space optimal interpolation (RSOI). The first stage is a large scale reconstruction based on Variational Bayesian Principal Compo-

| **Name** | **Period** | **Resolution** | **Reconstruction method** | **References** |
|---|---|---|---|---|
| HadSST3 | 1850-present | 5° monthly | Grid average with no interpolation | *Kennedy et al.* (2011a; 2011b) |
| HadISST2 | 1850-present | 1° monthly | Reduced-space optimal interpolation (RSOI) | *Kennedy et al.* (2019a) |
| ERSST v5 | 1854-present | 2° monthly | Empirical orthogonal teleconnections (EOTs) | *Huang et al.* (2017) |
| COBE-SST2 | 1850-2015 | 1° daily and monthly | Multiscale analysis based on EOFs | *Hirahara et al.* (2014) |

Table 2.1: Summary of historical SST datasets (only the most recent versions (as of May 2019) are shown).

nent Analysis (VBPCA) and the second stage is a small scale reconstruction using local optimal interpolation (OI) (*Kennedy et al.*, 2019a). OI estimates grid-point values as a linear combination of neighbouring observations, where the relative weights given to them are computed using the spatial autocorrelation function (*Gandin*, 1963). ERSST v5 uses the method of empirical orthogonal teleconnections (EOTs) for SST reconstruction. EOTs are a variant of empirical orthogonal functions (EOFs) where the derived modes have a clear physical interpretation (*Van Den Dool et al.*, 2000). COBE-SST2 uses a multiscale analysis based on EOFs. This consists of analyses on secular, interannual to interdecadal and daily time scales (*Hirahara et al.*, 2014). The secular trends are estimated first and then variations of SST on interannual to interdecadal time scales are reconstructed using EOFs, defined from an SST analysis with in-situ and satellite observations from 1961-2005. Day-to-day changes in SST are finally computed using an OI method.

Some criticisms of these reconstruction techniques are given in *Kennedy* (2014). One criticism that can be made about all of these reconstructions techniques is that they assume that observational errors are random and uncorrelated. They also assume that the SST fields can be parameterized in terms of a finite number of patterns or simple covariance relationships. Another concern is that SST variability in earlier periods

may not be accurately represented because the parameters of the statistical models are estimated from modern patterns of variability. Methods that use EOFs to describe the variability may impose nonphysical long-range teleconnections as EOF modes do not necessarily represent real modes of variability (*Dommenget*, 2007). The reconstruction methods also tend not to resolve small-scale structure in the variability. Methods using OI tend towards the climatological average in locations with no observations.

Before the routine use of satellite instruments to measure SST starting in the late 1970s, only in-situ SST measurements are used in these reconstructions. In-situ SST measurement methods have evolved over time. The earliest measurements from the $19^{\text{th}}$ Century were made using wooden, canvas or metal buckets from ships (*Kent et al.*, 2010). In the first half of the $20^{\text{th}}$ century, engine room intake (ERI) measurements became more prolific, which tended to have warm biases relative to bucket SSTs. Hull sensors were also used on ships and were found to provide more consistent measurements than ERI. From the early 1970s, moored and drifting buoys were used to measure SST and they are now the most numerous source of in-situ data. Figure 2.2 shows the changes in in-situ SST measurement methods over 1945-2007 in ICOADS.

During the IRIS period (1970-71) measurements were made mostly from buckets, with some from ERI and hull sensors. Figure 2.2b shows that relative biases exist between measurement methods, especially before 1970. This can result in spatially/temporally varying biases in the reconstructions, if the SST biases are not corrected (*Kennedy et al.*, 2011b). Knowledge of the biases associated with specific measurement methods together with detailed metadata (or other methods used to determine past measuring practices) allows the construction of SST bias models (*Kent et al.*, 2010). However, metadata is often missing, lacking in detail or even assigned to the wrong observations. Various assumptions must therefore be made, contributing uncertainty to the bias models. The three available families of SST datasets (HadSST/HadISST, ERSST, COBE) take somewhat different approaches to bias adjustment (*Kent et al.*, 2017). However, these approaches are all adaptations of methods

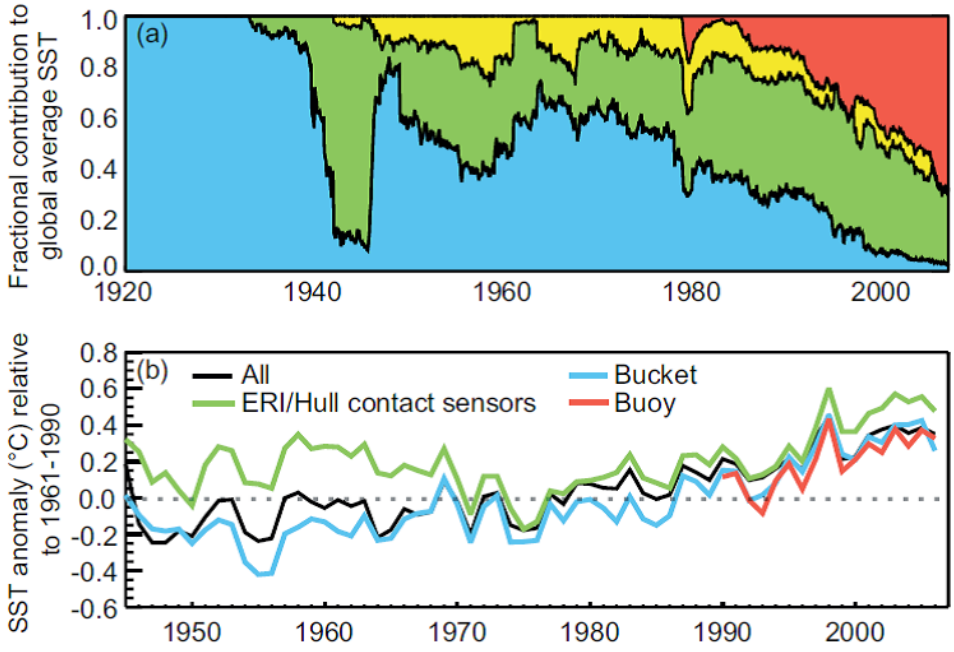



Figure 2.2: Temporal changes in the prevalence of different SST measurement methods in ICOADS. (a): Fractional contributions of observations made by different measurement methods: bucket observations (blue), engine room intake (ERI) and hull contact sensor observations (green), moored and drifting buoys (red), and unknown (yellow). (b): Global annual average SST anomalies based on different kinds of data: ERI and hull contact sensor (green), bucket (blue), buoy (red), and all (black). Reproduced from Figure 2.15 of *IPCC* (2013).

originally used decades ago. *Kent et al.* (2017) highlight that SST bias models can be improved. They also highlight the need for more diversity in bias correction methods, to improve understanding of structural uncertainty. This is a motivation for deriving early satellite SSTs to test the 'pre-satellite' reconstructions.

Figure 2.3 shows the cumulative effect of adding uncertainties from different error components in the global average SST anomaly for HadSST3. Sampling uncertainty is introduced into the gridded reconstructions due to estimating gridbox averages from a finite number of observations (*Kennedy et al.*, 2011a). Coverage uncertainty arises due to areas of ocean with no observations. It can be seen from Figure 2.3 that as the in-situ observation network has grown over time, the contribution from coverage uncertainty has decreased. Correlation between the errors in the observations (measurement

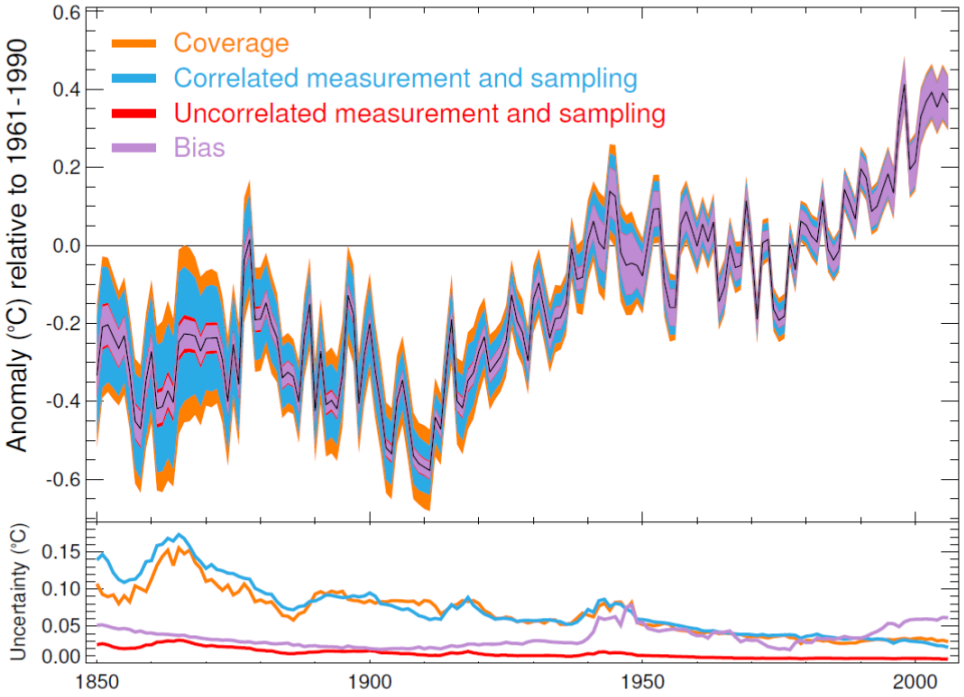



Figure 2.3: Median globally averaged SST anomaly from HadSST3 (black) with 2-sigma uncertainty arising from assumptions in bias adjustment (purple); measurement and sampling error, assuming these are uncorrelated between grid boxes (red); and the uncertainty including the intergridbox correlations (blue) and the total uncertainty including all the above terms and the coverage uncertainty (orange). The lower panel shows the sizes of the individual components as 2-sigma uncertainties. Reproduced from *Kennedy et al.* (2011b).

errors) has the effect of increasing the uncertainties in the gridbox averages (*Kennedy et al.*, 2011a). This increase is larger in gridboxes with many observations from only a few ships or drifting buoys. These correlations can also occur between gridboxes visited by the same ship or drifting buoy, increasing the uncertainty in the global average. The contribution to the total uncertainty from correlated measurement and sampling uncertainty has steadily decreased over time. Uncertainty originating from bias correction was a relatively minor component of the total uncertainty in the early record, but dominates the total since the 1960s. This is because from the 1940s, there were a greater variety of measurement methods resulting in increased uncertainty in how measurements were made, due to incomplete or incorrect metadata. It is important to

remember that the relative sizes of the different uncertainties depend on the time and spatial scales being considered. Over greater spatiotemporal scales, measurement and sampling uncertainties become less important relative to bias uncertainties (*Kennedy et al.*, 2011b).

### 2.2.1 IRIS vs in-situ coverage

It was shown in chapter 1 that the coverage of in-situ SST observations from ICOADS is far from complete, even in recent decades. Figure 2.4 compares the density of SST observations in ICOADS to the density of IRIS observations, during the IRIS period. For the purposes of this comparison, the quality control tests discussed in chapter 3 have been applied to the IRIS data to remove spurious observations. Observations over sea ice have also been removed.

The total number of SST observations from ICOADS is much higher than from IRIS. However, the coverage in ICOADS is highly heterogeneous. Major shipping routes are clearly visible in the map of the ICOADS coverage as lines with much higher concentrations of observations than surrounding areas. There is a general paucity of observations in the Southern Hemisphere, especially in the eastern South Pacific. In such regions, the reconstructions rely heavily on the statistical techniques previously mentioned to estimate the SSTs. There are no gaps in the IRIS coverage, except for the location of the South Atlantic Anomaly (SAA); a region where satellites in low-Earth orbit are exposed to strong radiation, causing instrument malfunction (*Anderson et al.*, 2018). IRIS observations in regions without in-situ observations are especially valuable if SST can be retrieved from them.

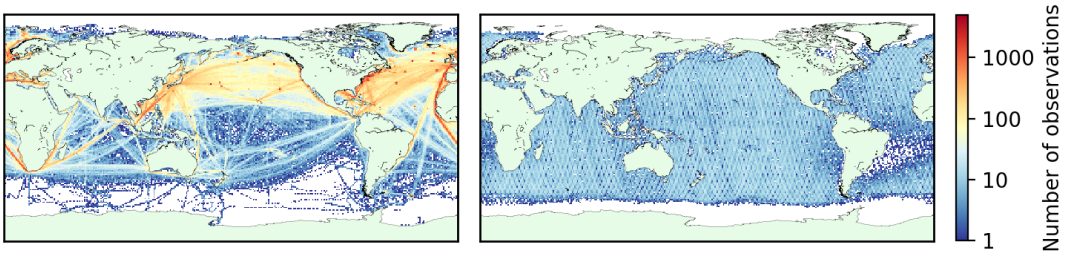



Figure 2.4: Total number of SST observations in ICOADS (Release 3.0) (left) and number of IRIS observations (QC applied) (right) per 1° cell during the IRIS period (April 1970 - January 1971).

## 2.3 How is SST retrieved from thermal satellite measurements?

The infrared (IR) radiance measured by satellite radiometers contains information about the state of the surface and/or atmosphere (*Merchant et al.*, 2014). How much information is contained about the SST mainly depends on the wavelength of the radiation and whether cloud obscures the field of view (FOV) of the radiometer. To explain the process of retrieving SST requires us to define the 'forward' and 'inverse' problems. The forward problem describes the problem of mapping the information about the atmospheric and surface state to the measured radiances. The physics of this mapping is encapsulated in the forward model. The inverse problem is that of finding an appropriate inverse mapping from the measured radiances to the atmospheric and surface state. However, the mapping is not unique, due to the presence of measurement uncertainty and also the fact that changes to the atmospheric and surface state do not always affect the measurements. A Bayesian approach can be used to solve the inverse problem (*Rodgers*, 2000) and this is set out in chapter 5.

### 2.3.1 Radiative transfer simulation

Simulation of satellite IR radiances is a key step in the process of SST retrieval. Simulation is carried out by a radiative transfer model (RTM), which simulates the physical

processes of emission, absorption and scattering of thermal radiation (*Merchant and Embury*, 2014). Any RTM must approximate the equations of radiative transfer, usually involving discretization and parameterization. RTM simulations therefore have uncertainty associated with these approximations. There is additional uncertainty associated with the characterisation of a particular sensor.

There are different types of RTMs with relative advantages and disadvantages. Line-by-line models generally run with minimal assumptions and approximations. As a result they are more accurate but more computationally expensive. There are also faster models with precalculated channel-integrated transmittance coefficients specific to each instrument. Partial derivatives of ToA radiance/BT with respect to atmospheric and surface variables can often be obtained with these type of models. RTTOV (Radiative Transfer for TOVS) is a widely used example of such a model (*Saunders et al.*, 2013). A subcategory of these faster models employ EOF analysis to identify a small subset of channels required to reconstruct the full spectrum, enabling even more efficient simulation. The Principal Component-based Radiative Transfer Model (PCRTM) (*Liu et al.*, 2006) is a key example of this type of model.

IRIS is a hyperspectral instrument, so efficient simulation is important. For this reason, PCRTM would be a preferred choice of RTM. However, due to issues with availability of this model, RTTOV (v11) will be used instead.

### 2.3.2 Retrieving SST in the presence of cloud

Clouds greatly affect observed IR radiances and therefore the retrieval of SST. There are three fundamental approaches for treating cloud effects on IR observations (*Susskind et al.*, 2003):

1. Identify clear scenes to avoid the problem

2. Solve for the radiative effects of cloud directly in the inversion process

3. Infer what the radiances are in the clear parts of the scene (clear-column radi-

ances)

Each of these approaches are discussed in turn in the following sections.

#### 2.3.2.1 Cloud detection

Cloud detection is covered extensively in the literature, with many different methods in use. A common method involves the use of BT thresholds. A combination of thresholds are typically used to detect different types of cloud (e.g. *Závody et al.* (2000)). These thresholds can exploit the variation of cloud emissivity with wavelength. For example, the BT difference between channels at 3.7 µm and 11 µm is used as a test for fog/low stratus. Modern threshold based cloud detection methods generally use dynamic thresholds, determined by RTM simulation for individual scenes (e.g. *Dybbroe et al.* (2005)). This means that thresholds are adapted to the actual state of the atmosphere and surface and instrument viewing conditions.

A probabilistic Bayesian method of cloud detection introduced by *Merchant et al.* (2005), has certain advantages over thresholds methods. Thresholds are typically defined using expert knowledge that may not be objective, and there may be inconsistencies between different threshold tests within the cloud detection scheme. Related to this is the fact that threshold schemes are often tailored to a specific sensor, so are difficult to update and improve. Additionally, users of an SST product may have different tolerance for cloud-related bias, but fixed threshold tests do not accommodate for this. *Merchant et al.* (2005) apply a Bayesian method of cloud detection to very high resolution sensors (O(1 km)). 'Texture measures' are used to quantify the local standard deviation in BT for a group of pixels, to exploit the fact that cloudy regions are generally more heterogeneous than clear-sky regions. IRIS has a resolution of $\approx 100$ km, which is much too coarse to be able to use texture measures to assist in cloud detection, suggesting that this method may not be suitable in this project.

Cloud detection methods normally used for hyperspectral sounders (IRIS being an early example) are based on detecting a cloud 'signature' in the observed spectra

(e.g. *Serio et al.* (2000), *McNally and Watts* (2003)). The *McNally and Watts* (2003) method is designed to identify clear channels within a potentially cloudy scene rather than completely clear scenes. The observed spectrum is first compared with a clear-sky simulated spectrum and the observed - simulated spectrum is termed the departure vector. To isolate the cloud signal in the departure vector, the channels are re-ordered according to their relative sensitivity to the presence of cloud. This is determined by defining the characteristic level for a channel as the altitude at which the ratio of the radiance effect of an opaque black cloud to the total clear radiance exceeds a pre-defined threshold. An algorithm is then used to identify the channel along the ranking where the cloud signal is first identifiable. All channels ranked more sensitive are flagged as cloudy and those ranked less sensitive are flagged clear. For SST retrieval, the channels that are most sensitive to cloud (window channels) are required. This means that the ability of this method to find additional scenes with some clear channels (not surface viewing) is not a benefit. For scenes with very thin or partial cloud cover, the cloud signal can be hard to detect. Older sensors like IRIS typically have a higher level of radiometric noise and are more prone to systematic effects, so detecting a cloud signature becomes harder still for these more difficult cases.

Imager based methods use high resolution imagery to improve the accuracy of other cloud detection methods. For example, *Eresmaa* (2014) used collocated Advanced Very High Resolution Radiometer (AVHRR) data to assist the cloud detection for the Infrared Atmospheric Sounding Interferometer (IASI). For each IASI FOV, an AVHRR based cloud flag was determined from statistical tests on the cluster of AVHRR pixels within the IASI FOV. In recent years, cloud detection methods based on artificial intelligence (AI) algorithms have been developed. (e.g. *Shi et al.* (2016); *An and Shi* (2015); *Johnston et al.* (2017)). However, most AI schemes are trained using visible wavelength cloud images only, limiting their usefulness (*Luo et al.*, 2019).

#### 2.3.2.2 Simulating cloud effects

*Eyre* (1989a) introduced a technique to extract information from cloud affected satellite sounding radiances for use in numerical weather prediction (NWP). This scheme simultaneously retrieves the cloud-top pressure and cloud amount with the temperature and humidity profiles and surface temperature. The problem of retrieving from cloudy radiances is highly nonlinear, so the retrieval must be iterated (*Rodgers*, 2000). This process is explained further in chapter 5. The theoretical errors using this retrieval scheme were calculated, based on one temperature/humidity profile with varying cloud conditions. The cloud parameters were retrieved to reasonable accuracy under most conditions. Accuracy is improved for high-level cloud with a high cloud fraction. However, the effect of cloud uncertainty on the parameters being retrieved (including SST) is much smaller for low-level cloud. It was also found that the uncertainty for surface skin temperature in cloud-free conditions (but not known to be cloud-free) is about three times greater than if it is known a priori to be cloud free. The retrieval scheme was also tested on real TIROS Operational Vertical Sounder (TOVS) data (*Eyre*, 1989b). The retrieved cloud parameters were not comprehensively validated, but they were shown to be qualitatively consistent with AVHRR imagery.

A more recent study built upon the work of Eyre to assess the feasibility of assimilating cloudy IR radiances (*Pavelin et al.*, 2008). This highlighted limitations in the forward modelling of cloud, when retrieving the cloud parameters. This is because the forward model (RTTOV) assumes a single-layer of grey[1] cloud with negligible depth, leading to error in the retrieved cloud parameters.

#### 2.3.2.3 Cloud-clearing

The 'cloud-clearing' approach can be applied to hyperspectral sounders. This approach was initially developed in *Chahine* (1974, 1977) and has since been applied in the Atmospheric Infrared Sounder (AIRS) Science Team algorithm for retrieving atmospheric

[1] cloud emissivity is independent of wavelength

and surface parameters in the presence of clouds (*Susskind et al.*, 2003). It has the advantages of avoiding the problem of clear-sky sampling bias from the first approach and does not require accurate modelling of the radiative properties of clouds, as required by the second approach. In this approach, cloud-cleared radiances are derived from a set of neighbouring fields of view (FOVs). *Susskind et al.* (2003) make the following assumptions. Firstly, the radiative properties of a given type of cloud are identical in the FOVs, which differ only in the relative amounts of these cloud types. The other key assumption is that the clear portions of the FOVs have the same temperatures, humidities etc. In the latest version of the AIRS Science Team algorithm (*Susskind et al.*, 2011), the cloud-cleared radiances are derived for all channels using 56 channels from the 15 µm $CO_2$ absorption band and 11 µm window region.

As the IRIS FOV is $\approx 100$ km in diameter, the assumption that the clear portions of neighbouring FOVs have identical characteristics is unlikely to be valid in most cases. Added to this is that IRIS is a nadir-only sensor, so there are no neighbouring across-track FOVs. This suggests that the cloud-clearing approach is unlikely to be effective for IRIS.

## 2.4 Early satellite SST retrieval

The idea of making sea surface temperature measurements from satellites was first discussed at the Woods Hole Oceanographic Institution, Massachusetts, in August 1964 (*Ewing*, 1965). In a study to determine the feasibility of making such measurements (*Greaves et al.*, 1966), it was demonstrated that IR data from the five-channel radiometer on TIROS-7 "provide good measurements of the gradients of sea surface temperature". Accurate absolute values of SST were unachievable at this time due to to large uncertainties in correcting for atmospheric attenuation. *Allison and Kennedy* (1967) evaluated the capability of the High Resolution Infrared Radiometer (HRIR) on the Nimbus-1 satellite to map SST. The HRIR (operating from 28th August to

23rd September, 1964) had a single channel in the 3.5 µm-4.1 µm window region and a ground resolution of 8 km at nadir. However, the sub-satellite position could only be determined to an accuracy of 16 km. The noise equivalent differential temperature (NEDT) was about 1 K; very poor by today's standards. Model atmospheres were used to compute theoretical values of outgoing radiation, in order to estimate the atmospheric absorption due to water vapour and carbon dioxide. *Allison and Kennedy* (1967) concluded that the HRIR "is capable of collecting limited amounts of useful oceanographic data." Under synoptic conditions where a cloud-free sea surface was assumed, the derived SSTs have an uncertainty of 1-3 K (after correcting for atmospheric absorption).

Over the next five years, methods of deriving SST from single channel radiometers such as the HRIR were developed. The first global analysis of satellite-derived SST was made by *Rao et al.* (1972). Data came from the first Improved TIROS Operational Satellite (ITOS 1), equipped with several TV cameras and radiometers. Launched in early 1970, it provided full day and night global coverage of the Earth's surface. SSTs were derived using the 10.5 µm-12.5 µm window channel. A technique developed by *Smith et al.* (1970) was used to determine mean SST over 2.5° latitude/longitude areas of ocean. These regions are generally larger than that covered by clouds, meaning that some of the measurements within these regions will be unaffected by cloud. From the histogram of all measurements within a region, a clear mode is identifiable, from which the mean SST for the region can be derived. A comparison between the satellite derived SSTs and ship measurements gave a root mean square (RMS) difference of 2-3 K. It was later found that the atmospheric attenuation corrections used in this study were too small. It was also found that the tape recorder introduced coherent noise into the data. A small-scale test on the corrected data showed an RMS difference of 1.8 K, still relatively large.

## 2.5 The Infra-Red Interferometer Spectrometer (IRIS)

Launched in April 1969, the Nimbus-3 satellite was the first in the Nimbus series to feature the Infra-Red Interferometer Spectrometer (IRIS). The IRIS instrument was a modified version of the Michelson Interferometer. It measured thermal emission spectra in the 5 µm-25 µm (400-2000 $\mathrm{cm}^{-1}$) spectral interval, with a resolution equivalent to 5 $\mathrm{cm}^{-1}$, although the 1400-2000 $\mathrm{cm}^{-1}$ range of channels became unreliable shortly after Nimbus-3 was launched (*Conrath et al.*, 1970).

An improved version of IRIS (IRIS-D) was carried on Nimbus-4, launched in early April 1970. The most significant improvement was an increase in the spectral resolution from 5 $\mathrm{cm}^{-1}$ to 2.8 $\mathrm{cm}^{-1}$ (*Hanel et al.*, 1971). This improvement was achieved partly by reducing the upper limit of the spectral range from 2000 $\mathrm{cm}^{-1}$ to 1600 $\mathrm{cm}^{-1}$. As a consequence of the increase in resolution, the FOV of the instrument decreased from 4° to 2.5° half cone angle. The circular area (footprint) under observation during the recording of one interferogram was reduced from a diameter of 150 km to 94 km, increasing the probability of seeing more homogeneous atmospheric and surface conditions. Data from IRIS on Nimbus-3 is reported as 'missing' (*Johnson et al.*, 2017) so only data from IRIS on Nimbus 4 is now available.

The original objectives of the IRIS experiment were to provide atmospheric profiles of temperature, humidity and ozone for use in studies of atmospheric circulation and synoptic scale meteorology (*Sabatini*, 1970).

### 2.5.1 IRIS operation and calibration

Background information on the IRIS instrument is documented in the Nimbus IV User's Guide (*Sabatini*, 1970). Figure 2.5 shows a schematic diagram of IRIS. The essential part of the interferometer is a beam splitter which divides the incoming radiation into two approximately equal components. After reflection, the two beams interfere with each other with a phase difference proportional to the optical path difference

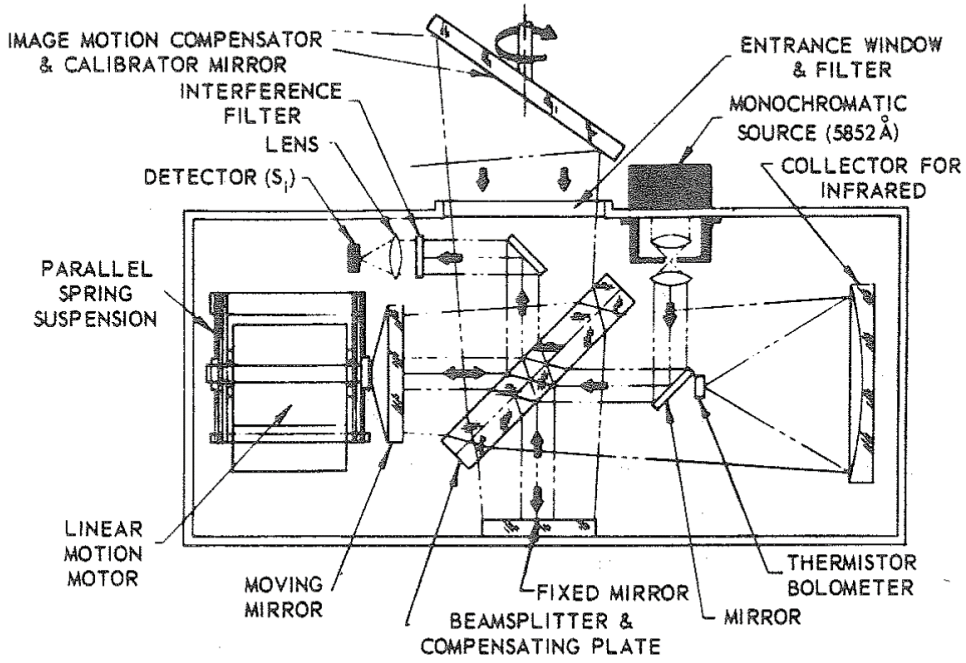



Figure 2.5: Schematic diagram of IRIS instrument. The monochromatic source is a neon discharge tube. Reproduced from *Sabatini* (1970).

($\delta$) between the beams. The recombined components are focused onto the bolometer detector where the intensity is recorded as a function of $\delta$. A circular fringe pattern occurs at the focal plane of the condensing mirror.

IRIS has an Image Motion Compensation and Calibration (IMCC) system which channels radiation from several sources to the radiometer. After 14 interferograms are taken in operating mode, one is taken from a built-in blackbody kept at spacecraft temperature (warm calibration target) followed by one taken from outer space (cold calibration target). These interferograms are used for in-flight calibration. The calibration mode is prohibited near the North Pole due to danger of viewing the sun. Various telemetry data are also generated each orbit including instrument temperatures and voltage calibrations.

To generate the calibrated IRIS spectra (level 1), the raw interferograms are screened for consistency and completeness and Fourier transformed into the spectral domain. An apodisation (smoothing) function is applied before transformation to reduce the side

lobes of the instrument function. The amplitude in the spectrum ($c_\nu$) is proportional to the difference in radiance between the target and instrument,

$$c_\nu = r_\nu(B_t - B_i) \tag{2.1}$$

where $r_\nu$ is the responsivity of the instrument. The responsivity is calculated using the warm and cold calibration spectra without knowledge of $B_i$, which is more difficult to measure due to thermal gradients within the interferometer (*Hanel et al.*, 1971). Figure 2.6 shows the responsivity measured at different times before launch and in orbit. The responsivity remains relatively constant, indicating the optical-mechanical stability of the instrument. The noise equivalent radiance (NER) can be determined

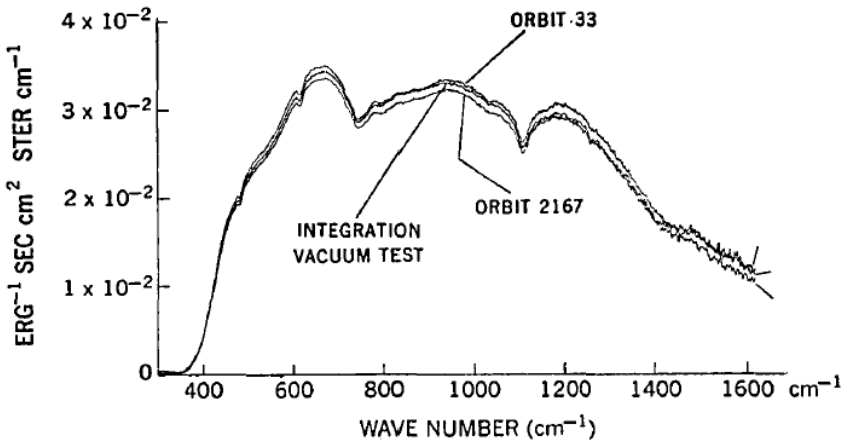



Figure 2.6: Spectral responsivity of IRIS-D during the thermal vacuum test of the spacecraft about 2 months before launch, shortly after launch (orbit 33) and after about $5\frac{1}{2}$ months (orbit 2167) in orbit. Reproduced from *Hanel et al.* (1971).

using the mean responsivity during an orbit ($\hat{r}$)

$$NER = \frac{\sqrt{2}s_r B_w}{\hat{r}} \tag{2.2}$$

where $s_r$ is the standard deviation of the responsivity during an orbit and $B_w$ is the

radiance of the warm blackbody. Systematic variations in the NER were reported due to small orbital temperature changes within the instrument, but the effect of these changes was removed in the final data reduction process. A few other corrections were also applied in the final data reduction process (*Hanel et al.*, 1972). A correction was made to compensate for the nonblackness of the warm calibration target. Figure 2.7 shows the spectral emissivity of the warm calibration target as measured in the laboratory. This shows that the warm calibration target is very close to an ideal blackbody except for a feature near 1100 $\mathrm{cm}^{-1}$ attributed to minerals in the black paint used. A correction was

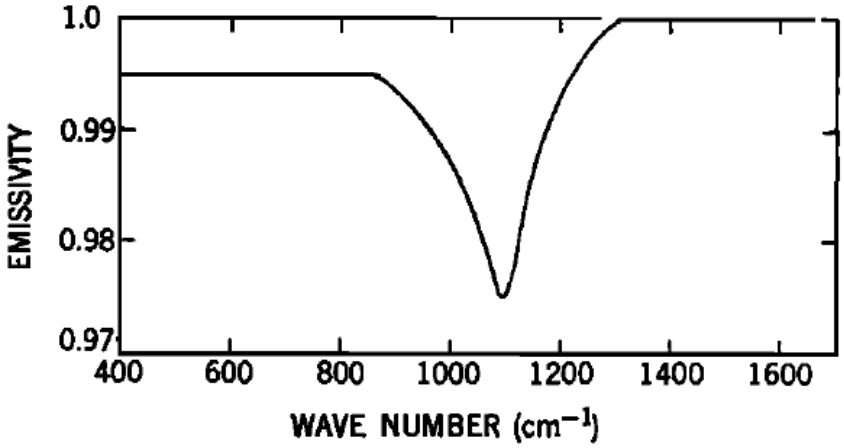



Figure 2.7: Spectral emissivity of the warm calibration target based on a laboratory measurement. Reproduced from *Hanel et al.* (1972).

also made for a deficiency in the cold calibration signal. The effect of this deficiency can be seen in Figure 2.8, which shows that the instrument temperature calculated from the warm and cold calibration spectra is somewhat lower than the thermistor measured value, deviating by several K above 1200 $\mathrm{cm}^{-1}$. Without this correction, negative radiances occur regularly at the high wavenumber end of the spectrum, but most disappear after the correction is applied.

A small smear of the FOV is also reported, beginning in September 1970. This was found not to have any effect on the quality of the spectra and so no correction was required. However, it did result in a slight elongation of the effective FOV. During the

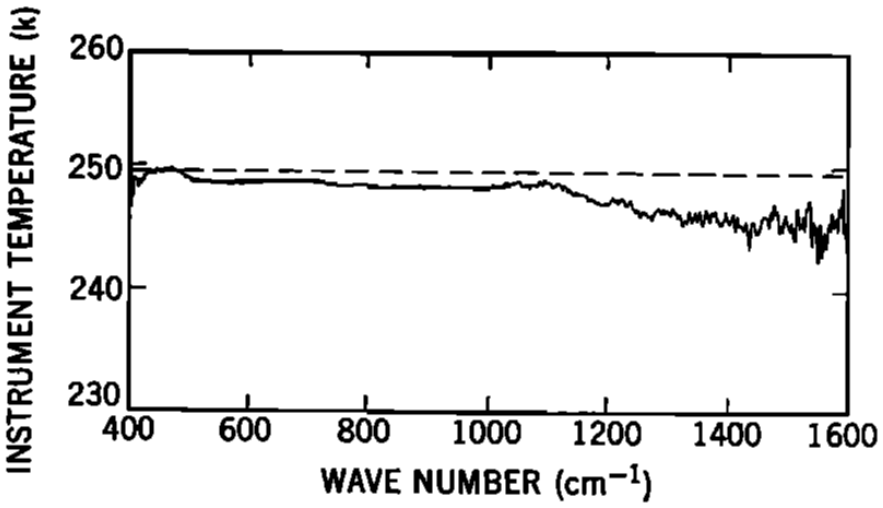



Figure 2.8: Instrument temperature calculated from warm and cold calibration spectra (solid line) as a function of wavenumber, compared to the thermistor measured value (dashed line). Reproduced from *Hanel et al.* (1972).

first months in orbit, the FOV was a circle 95 km in diameter, but it later increased in the direction of the satellite motion to cover an area of 95 by 120 km.

### 2.5.2 IRIS radiance spectra

Figure 2.9 shows typical IRIS radiance spectra from the tropical Pacific. Radiance curves for blackbodies at different temperatures are also shown. The main spectral absorption features are labelled. Absorption from the water vapour continuum affects all parts of the spectra to a varying degree (*Shine et al.*, 2012). The atmospheric window regions occur between 775-960 $\mathrm{cm}^{-1}$ and 1080-1200 $\mathrm{cm}^{-1}$ where absorption is minimal. Channels in the window regions are particularly sensitive to surface temperature, so are the most useful channels for SST retrieval. The radiance varies greatly in the $CO_2$ absorption band, as the radiation for channels in this band emanate from a range of levels in the atmosphere, from the lower troposphere to upper stratosphere (*Poli and Brunel*, 2016). The ozone absorption band is partly opaque with radiation from both the surface and the stratosphere (*Iacono and Clough*, 1996).

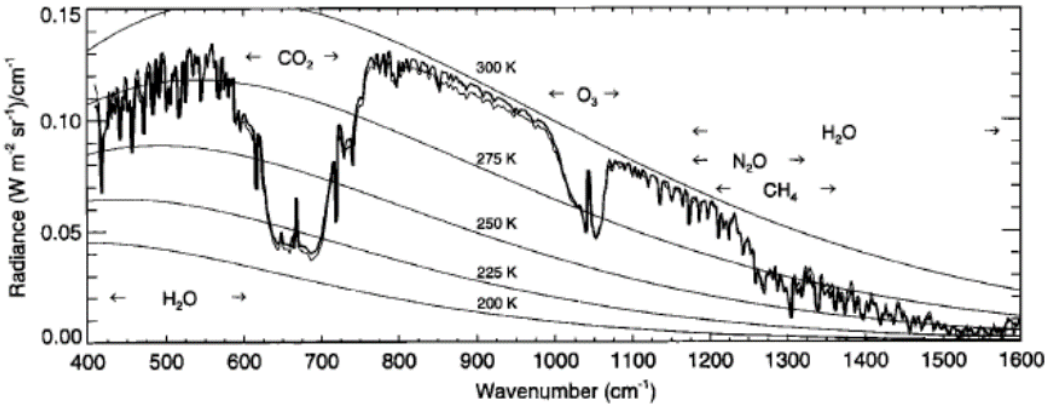



Figure 2.9: Typical IRIS radiance spectra from the tropical Pacific. For reference, radiance curves for blackbodies at different temperatures (Planck functions) are shown and the main spectral absorption features are labelled. Adapted from *Iacono and Clough* (1996).

### 2.5.3 Early attempts to retrieve SST from IRIS

SST estimation was never the primary focus of the IRIS experiment, but a few early attempts to retrieve SST from IRIS were made in the decades following it.

In the first main attempt to retrieve SST from IRIS (*Prabhakara et al.*, 1974), a split window technique was used. IRIS channels in the 775-960 $\mathrm{cm}^{-1}$ window region were averaged to simulate three broadband channels. The split window technique exploits the differential absorption properties of water vapour in this spectral interval. This enables the attenuating effect of water vapour to be removed from the IRIS measurements, allowing the SST to be determined directly in clear-sky conditions. Due to the early nature of this study, aerosols are not explicitly accounted for. A very small sample of eight clear sky IRIS spectra collocated with ship measurements were selected, by inspection of images from the Nimbus-4 Image Dissector Camera System (IDCS). The RMS difference between the retrieved SSTs and those from ship measurements was 1.1 K. *Prabhakara et al.* (1974) were able to obtain a larger set of about 100 cloud-free IRIS measurements from the Nimbus-3 version of IRIS, by using data from the Medium-Resolution Infared Radiometer (MRIR) for cloud detection. With a FOV of about 50 km at nadir, the MRIR had four IR channels as well as a channel spanning

visible and near-IR wavelengths. Measurements from this visible channel indicate the amount of reflected solar radiation. Since clouds reflect solar radiation much more than the ocean surface, this provides a method of cloud detection. The RMS difference between the retrieved and ship SSTs was 1.3 K. However, much of the difference is accounted for by uncertainty in the ship SSTs.

Over two decades later, *Iacono and Clough* (1996) used IRIS observations to investigate climate variability. The initial phase of the study involved validation of IRIS spectra against those simulated from a line-by-line radiative transfer model (LBLRTM). Soundings of temperature and water vapour from radiosondes were used as input to the model for both tropical and mid-latitude locations. Several numerical corrections were applied to the simulated spectra. First, the effect of the instrument FOV was corrected for, which results in considerable self-apodisation. Corrections were also made for spectral variations of ocean emissivity and the radiance reflected from the ocean surface. The IRIS data were processed to filter out spurious radiance spectra, many of these being Planck function calibration spectra. Three IRIS radiance spectra observed near Guam in the western Pacific were selected, as radiosonde measurements were taken from Guam within two hours of the IRIS observations. Figure 2.10 shows the observed-simulated (IRIS-LBLRTM) spectral radiance difference for these three cases. Overall, the radiance differences are mostly less than $3\,(\mathrm{mW\,m^{-2}\,sr^{-1}})/\mathrm{cm^{-1}}$, which equates to approximately 2 K in the 800-1000 $\mathrm{cm^{-1}}$ region. *Iacono and Clough* (1996) claim that known systematic errors in the radiosonde measurements explain much of the differences. The larger residual in Figure 2.10c across most of the 450-1250 $\mathrm{cm^{-1}}$ range is mainly accredited to diurnal relative humidity error in the radiosonde measurements. Specification of the SST is noted as another source of uncertainty and potential contamination from thin/partial cloud may also contribute to the residual in the 400-1000 $\mathrm{cm^{-1}}$ range.

To investigate climate variability, clear-sky IRIS observations were selected for large geographic regions. Cloud detection was approached using brightness temperature

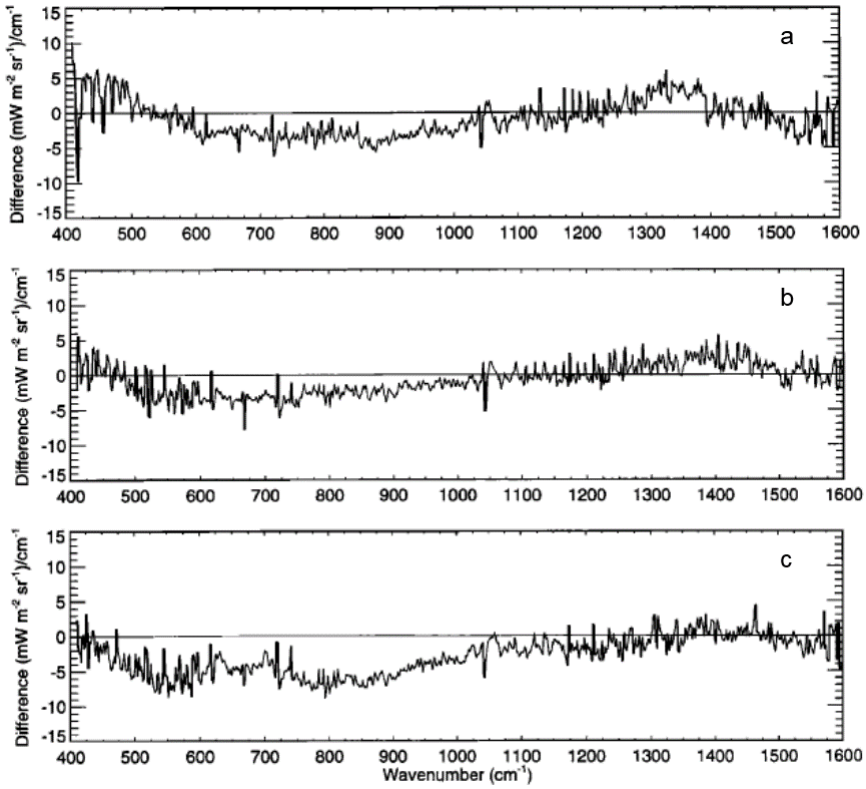



Figure 2.10: Differences in radiance spectra from a line-by-line radiative transfer model (LBLRTM) calculation and IRIS observation near Guam at (a) 0202 UT on 27$^{th}$ April, 1970; (b) 1200 UT on 12$^{th}$ July, 1970; (c) 0000 UT on 13$^{th}$ July, 1970. Adapted from *Iacono and Clough* (1996).

(BT) thresholds. In equatorial regions, water vapour variability is typically low. Since water vapour absorption is the main driver of differences between the SST and clear-sky BT for highly transparent window channels, the distribution of BT can be used to select an appropriate threshold. However, thin cirrus contamination remains a problem using this threshold method. In some places, there may be low-level cloud of similar temperature to the ocean surface which would be undetected using this method. The standard deviations of the set of 3500 clear-sky spectra in the western Pacific were found to be generally 1-2 K in the 800-1200 cm$^{-1}$ window region. *Iacono and Clough* (1996) state that this is consistent with the approximate 1 K standard deviation of tropical SSTs over annual mean conditions.

*Iacono and Clough* (1996) use a linear retrieval algorithm to quantify the signifi-

cant seasonal variations in SST and other atmospheric parameters from the clear-sky, tropical, mean radiance differences. The retrieval is performed on regionally averaged spectral radiances rather than individual radiance spectra to obtain regionally averaged atmospheric states. This is justified, as the standard deviations of the radiances in the 600-1200 $\mathrm{cm}^{-1}$ region are low in tropical regions. The least-squares retrieval algorithm minimises the radiance residual between the seasonal and annual mean spectra for the selected parameters. The retrieval is applied to the 430-1200 $\mathrm{cm}^{-1}$ spectral range excluding the $CO_2$ absorption band (610-725 $\mathrm{cm}^{-1}$). The 430-610 $\mathrm{cm}^{-1}$ spectral range is included as it makes up part of the water vapour pure rotation band, which contains information about middle and upper tropospheric water vapour. The differences between the retrieved and observed (in-situ based) seasonal SST changes are within 0.2-0.7 K for the two tropical Pacific regions studied. Both the retrieved and observed SSTs were found to decrease during the year relative to the 10-month mean. This is due to the onset of La Niña conditions in the summer of 1970. For the tropical Atlantic and Indian Ocean region studied, the retrieved-observed seasonal SST changes are within 0.1-0.7 K except for the Indian Ocean in spring and summer.

Up until now, there have been no further attempts to retrieve SSTs from IRIS. The results of *Iacono and Clough* (1996) suggest that it is possible to retrieve SSTs of sufficient accuracy to provide a useful diagnostic for climate variability. There is reason to believe that SSTs with improved accuracy and coverage can be retrieved in our work. This is because more recent satellite measurements of outgoing IR spectra, made by the IASI instruments starting in 2006 can be used to test and refine a method of retrieving SST from the IRIS measurements. RTMs have also improved since the 1990s (e.g. *Clough et al.* (2005)), meaning that IRIS spectra can be simulated from NWP estimates more efficiently and accurately. In addition to IRIS, Nimbus-4 carried many other instruments including the Temperature Humidity Infrared Radiometer (THIR) and IDCS (*Sabatini*, 1970). The THIR recorded top-of-atmosphere (ToA) brightness temperature in two broadband channels: a window channel centred at 11.5 µm and a

water vapour channel centred at 6.7 µm. It had a ground resolution of 8 km compared to 94 km for IRIS, so may assist in the process of cloud detection. The IDCS collected visible images of cloud cover with a ground resolution of around 3 km. These images may be used in combination with data from the THIR to improve cloud detection.

## 2.6 Chapter summary

SST is a critical variable in the Earth's climate system, and so keeping a reliable record of SST is crucial. There are several different historical SST datasets, based purely on in-situ measurements before the 1980s. These datasets have different methods of reconstructing SST in poorly-observed regions; all of which have their limitations.

Although the original objective of the IRIS experiment was never to derive measurements of SST, it has been previously demonstrated that SST information can be retrieved from IRIS to sufficient accuracy to study climate variability. Various factors suggest that higher-quality SSTs with improved coverage can now be retrieved from IRIS, relative to previous studies. This presents the opportunity to test the reliability of the historical SST reconstructions during the early 1970s.

# CHAPTER 3

# QUALITY CONTROL OF THE IRIS DATA

## 3.1 Introduction

Data quality control is the process of examining a dataset in order to find anomalies or errors in the data. Its purpose is to improve the overall quality of the dataset, so that it can provide more robust, reliable scientific findings. This chapter discusses the quality control procedures applied to the Nimbus-4 IRIS dataset.

We obtained the dataset from R. Bantges (pers. comm., 2016) in the form of monthly plain text files. The contents of these files originate from the 'IRIS/Nimbus-4 Level 1 Radiance Data V001' dataset (*Hansel*, 1994) and no changes have been made to the data from this. Accompanying the radiance spectra themselves are a variety of other data. These include warm and cold reference calibration spectra, noise equivalent radiance spectra, and temperatures of various components of IRIS. These extra data can be used as a means of quality controlling the IRIS radiance spectra.

In this chapter, we reveal the existence of blackbody spectra in the radiance data, which were probably mistakenly encoded as atmospheric measurements. The process of quality control can remove these spectra, as well as other erroneous spectra that are identified. The aim of this chapter is to describe the methods used to discard

such spectra in order to produce a dataset that enables retrieval of high quality SSTs. A variety of methods are used, with the QC divided into four stages. The first two stages use thresholds to discard outlying spectra. These thresholds are decided based on prior expectations of the data. QC stage 3 involves comparison of the spectra with NWP-based simulations, to check for anomalous cases. Finally, QC4 uses principal component analysis (PCA) to describe statistical properties of the data in order to identify outliers. Quality control of the level 1 (L1) IRIS dataset was carried out in *Poli and Brunel* (2016). QC stages 1-3 are replicated from *Poli and Brunel* (2016), apart from certain modifications which will be explained. After application of these QC procedures, the radiance data are assessed by comparison with simulated spectra (section 3.6). Finally, in section 3.7 RTTOV and PCRTM are compared to investigate potential forward model error.

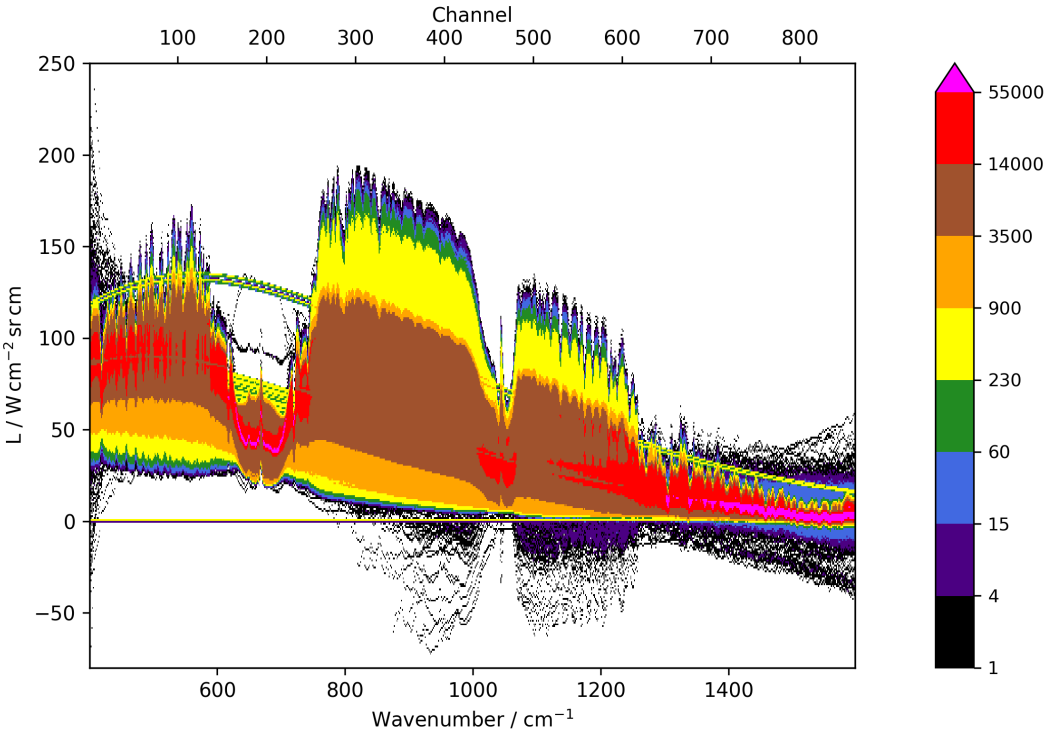
## 3.2 QC1 and 2

The tests involved in QC1 and 2 are summarised in table 3.1. Except in the case of test 1.1, the data are first converted to brightness temperatures (BTs) to make the tests simpler to apply.

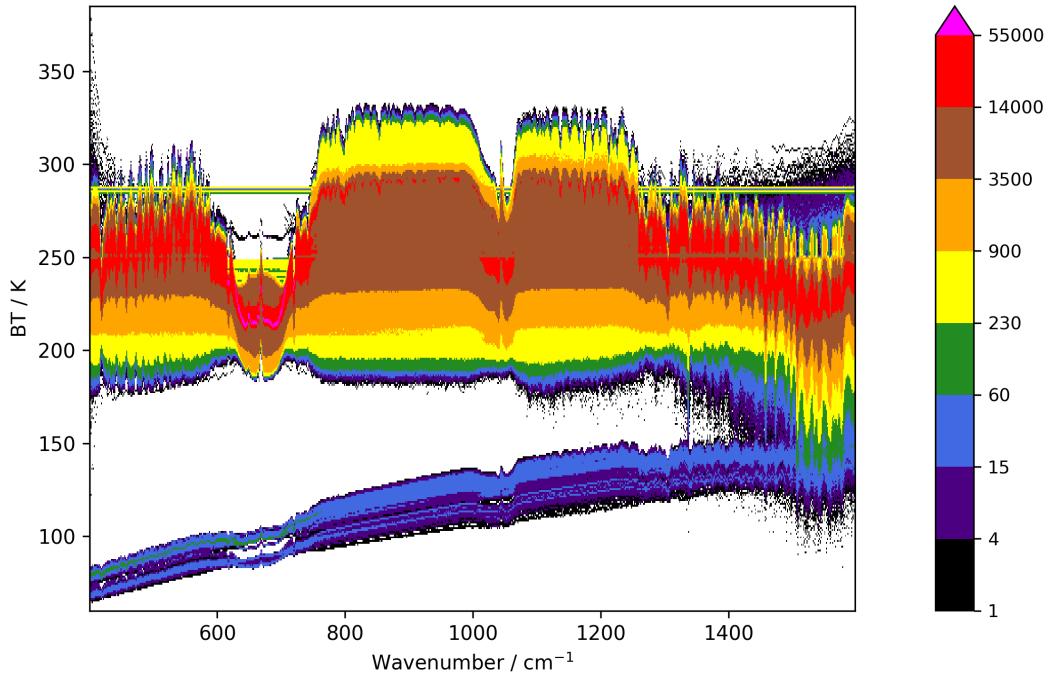
| **QC stage** | **Test** |
|---|---|
| 1.1 | Discard spectra where $L_{1-720} < 0$ or $> 220\,\mathrm{W\,cm^{-2}\,sr\,cm}$ |
| 1.2 | Require spectra to have $T_{\max}$ - $T_{\min} > 19\,\mathrm{K}$ |
| 1.3 | Require $150\,\mathrm{K} \leq T(677\,\mathrm{cm^{-1}}) \geq 250\,\mathrm{K}$ |
| 2.1 | Discard spectra if cooling surface temperature, IMCC subassembly temperature or mirror temperature are below 240 K |
| 2.2 | Discard spectra if IMMC position is greater than 3 |
| 2.3 | Discard spectra if the +0.6 V, 0.0 V, -0.6 V calibrations are outside the ranges 0.59 to 0.65, -0.01 to 0.05, -0.61 to -0.55 respectively, or if the transducer calibration is outside the range 0.034 - 0.040 |
| 2.4 | Discard all spectra from days 2 to 9 of January 1971 |

Table 3.1: Summary of tests making up QC1 and QC2

Figure 3.1a shows a density plot of all radiance observations from the L1 data before application of any QC. Characteristic absorption features can be seen in most spectra.




(a)




(b)

Figure 3.1: Density plot of all radiance observations (a) and brightness temperatures (only computed for positive radiances) (b). Numbers indicate how many observations are in each interval of 1 W cm$^{-2}$ sr cm or 1 K per channel.

Two main groups of blackbody spectra can be distinguished as they lack these absorption features. A number of nonphysical negative radiances are also seen, occurring in highest density in the highest wavenumber channels. QC1.1 involves discarding spectra if any radiance in channels 1-720 (400-1400 $\mathrm{cm}^{-}1$) is negative. This ensures that all remaining spectra are free from negative radiances in these channels. Channels 721-862 are of little importance in this work and have a much higher proportion of negative radiances so the test is not applied to these. After QC1.1, all remaining L1 radiances were converted to BTs using Planck's law, for it is easier to apply the other QC tests to the data in this format. All negative radiances in channels 721-862 had to be set to 'NaN' to calculate BTs in these channels. Figure 3.1b shows a density plot of all BTs. The blackbody spectra are now seen as horizontal lines of constant BT and so are all removed in QC1.2. Some spurious data can be seen in the $CO_2$ absorption channels and the spectra corresponding to these are removed in QC1.3. Figure 3.1b also reveals a group of spectra with consistently low BTs; these correspond to a line of data with constant radiance just above zero in Figure 3.1a. QC1.3 removes all of these spectra.

QC2 uses other data in the IRIS dataset to remove spectra measured when the instrument appeared to function outside of its normal range. *Poli and Brunel* (2016) show that the temperature for the whole instrument rose around October 1970, near the end of the mission. However, it is difficult to explain why this happened and the implications it may have on the radiance measurements. There is no evidence for any significant impact on the measurements. QC2 is made up of simple threshold tests summarised in table 3.1. *Poli and Brunel* (2016) identify anomalous orbit tracks in January 1971, suggesting incorrect geolocation. Figure 3.2 shows a map of the orbit tracks in January. The orbit tracks from days 2 to 9 are aligned differently to all other orbit tracks in January and the rest of the IRIS period, and so QC2.4 discards all spectra from days 2 to 9.

Figure 3.3 shows a density plot of all BTs after application of QC stages 1 and 2. The effect of QC1 on removing the types of spurious spectra previously mentioned

can be seen. The spectra removed in QC2 tend not to be obviously erroneous looking spectra so the effect of this QC stage can not clearly be seen in Figure 3.3.

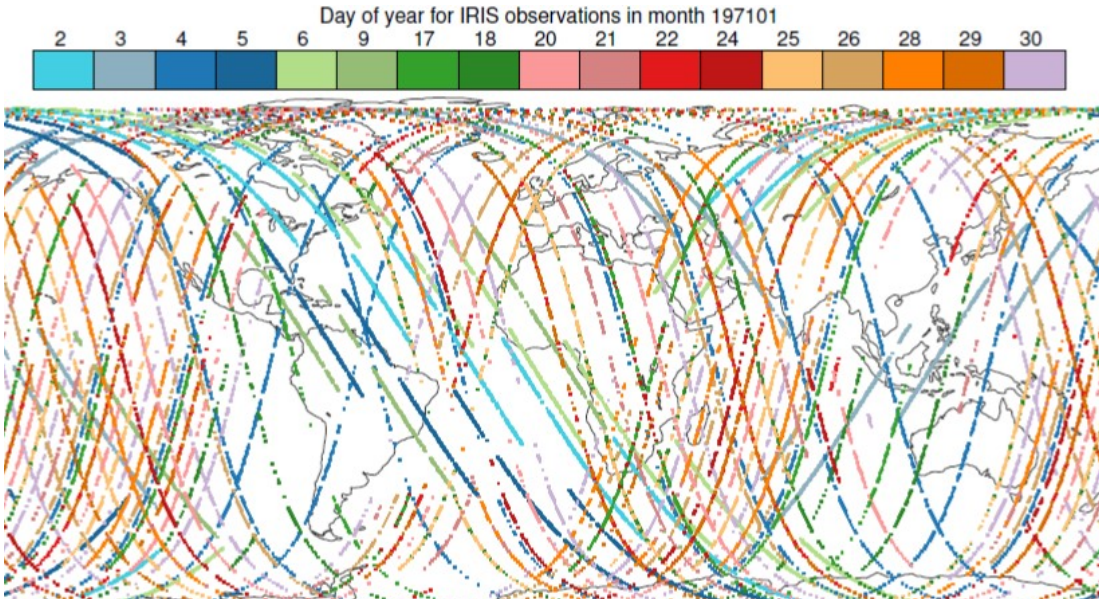



Figure 3.2: Map of IRIS observations during January 1971, colour coded according to day of the month. Orbit tracks for days 2 to 9 appear to be incorrect. Reproduced from *Poli and Brunel* (2016).

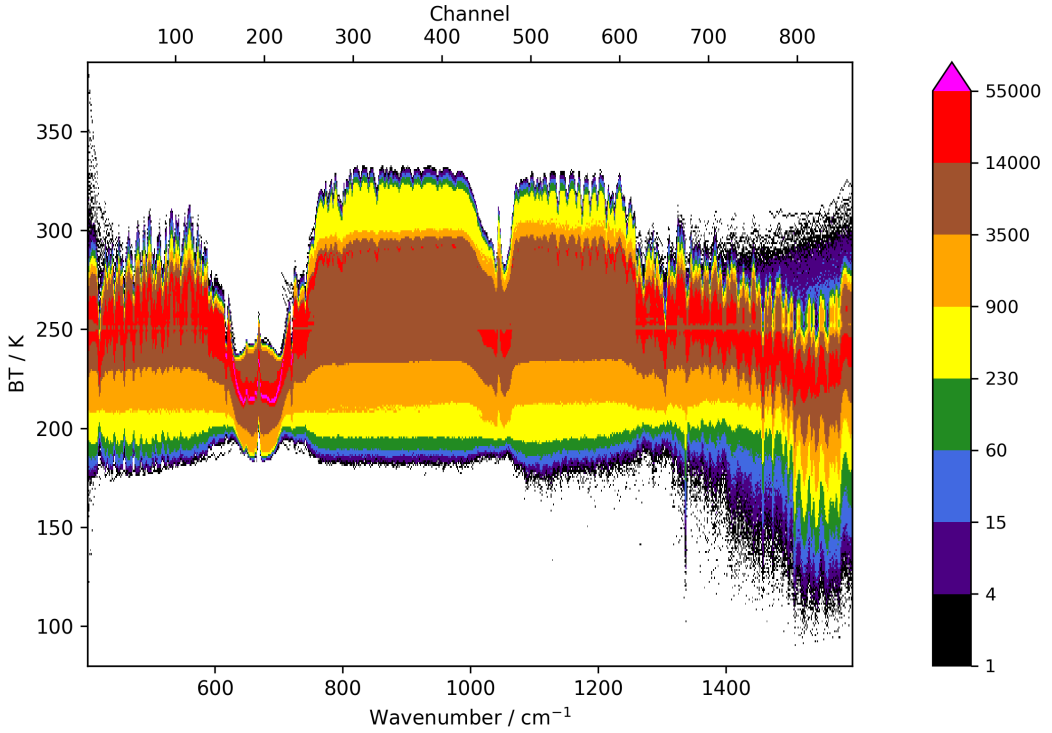
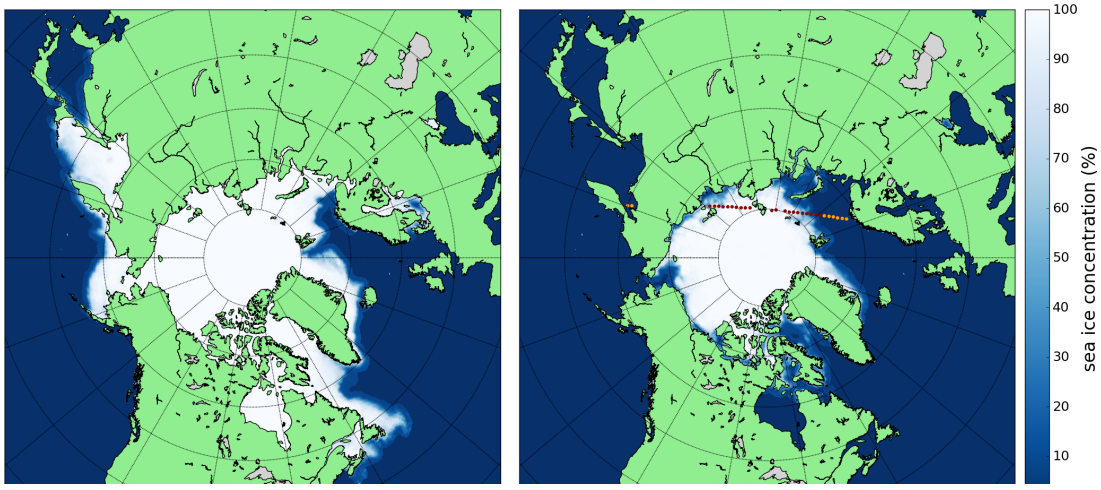



Figure 3.3: Density plot of all brightness temperatures for the data passing QC1 and QC2.

## 3.3 Land and sea-ice removal

As this research aims to retrieve only SST from the IRIS observations, it is necessary to discard spectra measured over land and sea-ice. To discard land based spectra, a 300 m resolution distance-to-land dataset (*Carrea et al.*, 2015) was used to compute the closest distance of each IRIS observation sub-satellite point from land. As the IRIS footprint is about 94 km in diameter, any observation within 47 km of land was discarded. Identifying spectra measured over sea-ice is more complex, as there are no spatially complete datasets of global sea-ice extent as early as 1970-71. Because of this, a more recent dataset derived from passive microwave data was used to estimate the maximum likely sea-ice extent for each day of the IRIS period. EUMETSAT's Global Sea Ice Concentration reprocessing dataset consists of gridded daily ice concentration from 1978-2015 (*EUMETSAT*, 2015). For each day of the year, the maximum ice

concentration over this period was calculated to give the likely maximum daily sea-ice concentration. This was smoothed slightly by computing a running 5-day median. From this smoothed series of daily sea-ice concentration, Figure 3.4 shows the sea-ice concentration in the northern hemisphere for the days of the year with maximum and minimum sea ice coverage. Part of an IRIS orbit from the 10$^{th}$ September is superimposed over Figure 3.4b to show which of the observations are discarded. IRIS observations are discarded for any gridsquare where the sea-ice concentration is non-zero. Due to the uncertainty of sea-ice extent on day-to-day timescales and allowing for the size of the IRIS footprint, any observations within 100 km of gridsquares with sea-ice are also discarded. At a later stage, it was discovered that this sea-ice masking fails to remove a few observations from locations where the ERA-20C skin temperature is far below -1° C. It is highly likely that these observations were measured over sea-ice, so any remaining observations with matched ERA-20C skin temperatures colder than -1° C were discarded.




(a) 13$^{th}$ March (b) 10$^{th}$ September

Figure 3.4: Sea ice concentration in the northern hemisphere from the smoothed series of daily maximum concentration (1978-2015), on the days with maximum (a) and minimum (b) coverage. The locations of the IRIS footprint for observations from the 10$^{th}$ September are also shown in (b) to indicate which observations are discarded (red dots) and which are retained (orange dots).

## 3.4 QC 3

*Poli and Brunel* (2016) show how simulation of IRIS spectra using a RTM can be used to provide further QC of the observations. They used RTTOV to provide simulations matched to each observed spectrum. The input came from atmospheric reanalyses, which provide atmospheric profiles of temperature, water vapour and ozone and important surface parameters (collectively called the state vector), all of which were used to simulate clear sky radiance/BT spectra. RTTOV is able to simulate the effects of cloud to simulate cloudy observations much more accurately, but this requires input of atmospheric profiles of cloud properties, which are not nearly as well constrained as temperature and water vapour profiles. Therefore, simulations of clear-sky IRIS observations are likely to be much more accurate than simulations of cloudy observations. *Poli and Brunel* (2016) chose the ERA-20C and ERA-40 reanalyses to provide the required state vectors to produce two sets of simulations. For each set of simulations, the linear correlation between each matched IRIS and simulated spectrum was computed. To limit the influence of instrument noise on the computed correlations, only channels 81-500 (9.1-19.5 µm) were considered where the instrument noise is not so large. Figure 3.5 shows a density plot of the IRIS noise equivalent radiance spectra. These spectra give the rms noise of each measurement and show that for channels 81-500, the noise is typically less than $1\,\mathrm{W\,cm^{-2}\,sr\,cm}$. ERA-20C has relatively high spatial and temporal resolution compared with other reanalyses that cover the IRIS period. The methodology of *Poli and Brunel* (2016) is replicated here, but using only ERA-20C to produce one set of correlations. The ERA-20C fields closest in space and time to the observation sub-satellite point were selected, whereas *Poli and Brunel* (2016) interpolate the fields bi-linearly in space (but not in time). We believe that interpolation is unimportant here, due to the level of spatial uncertainty in reanalysis fields from the IRIS time period. Figure 3.6 shows a histogram of the observation-simulation correlations ($r$). There is no clear way of choosing the right value of $r$ to set as a threshold to

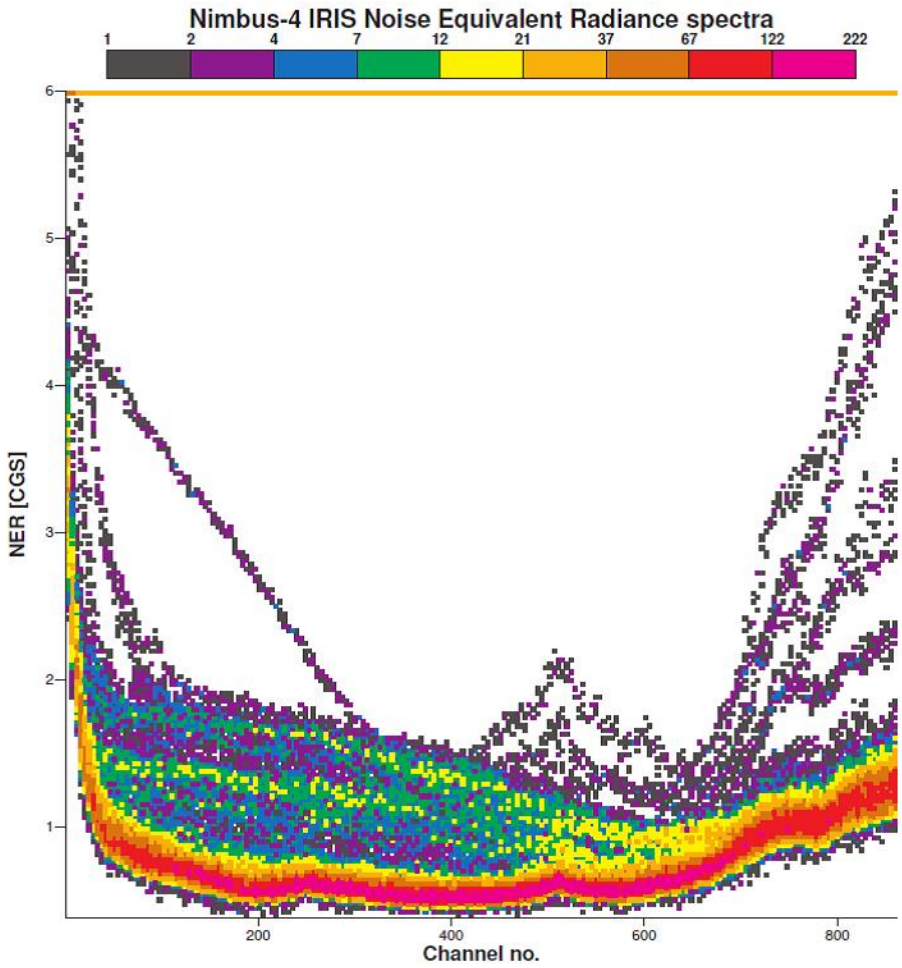



Figure 3.5: Density plot of IRIS noise equivalent radiance spectra. Reproduced from *Poli and Brunel* (2016).

choose which spectra to retain. *Poli and Brunel* (2016)'s version of QC3 is defined as the removal of spectra where correlations with ERA-20C and ERA-40 simulated spectra are both below 0.8. Our version of QC3 is based on only one set of correlations, so a more lenient threshold of $r = 0.5$ was chosen. This removes 0.4% of pre-QC3 IRIS spectra. This process has a tendency to remove very cloudy spectra as well as probably spurious spectra. This is because the correlation between very cloudy spectra and simulated clear-sky spectra tends to be weak. However, this is not an issue for SST retrieval, as it is difficult to retrieve SST from very cloudy scenes.

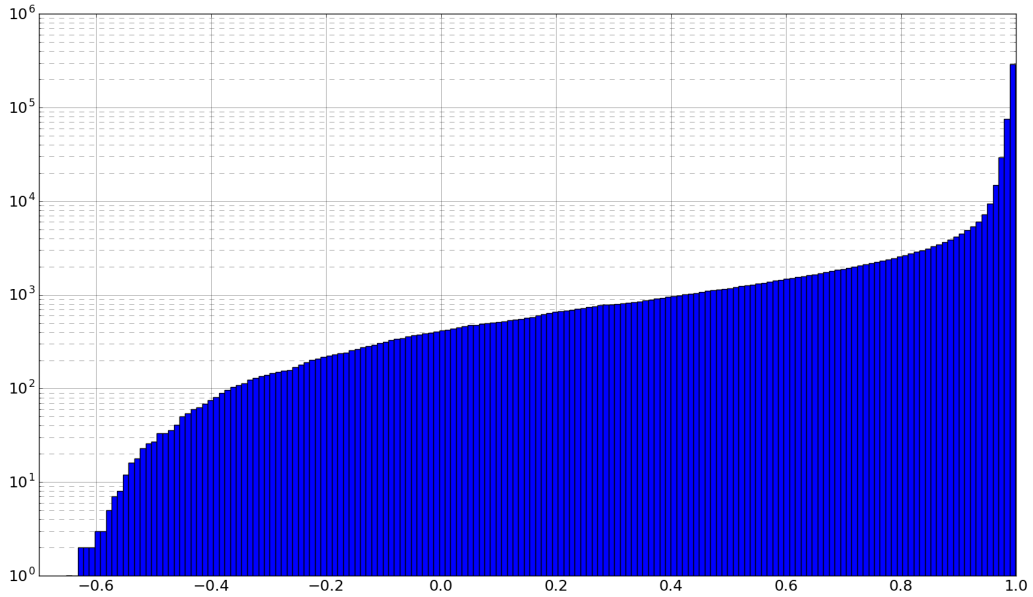



Figure 3.6: Histogram of $r$ (Pearson correlation coefficient) between all pairs of simulated clear-sky and observed IRIS spectra.

## 3.5 QC 4

We aim to go beyond the efforts of *Poli and Brunel* (2016) in quality controlling the IRIS data. An extra stage of QC is thefore added, QC4, which exploits the statistical method of PCA to find outliers in the IRIS radiance spectra. The use of principal components in quality controlling satellite observations has been demonstrated by *Matricardi and McNally* (2014). They developed a PC-based quality control as an extra check for residual cloud contamination in IASI radiance spectra. Singular value decomposition (SVD) is a powerful tool for extracting the PCs (EOFs) and PC scores from any data matrix (*Hannachi*, 2004). If the data matrix is an $m \times n$ matrix of $n$ IRIS BT spectra, $\mathbf{T}$, SVD factors it according to equation 3.1.

$$\mathbf{T} = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{V} \tag{3.1}$$

where $\mathbf{U}$ is an $m \times m$ matrix of EOFs, $\mathbf{\Lambda}$ is a $m \times n$ matrix of $m$ singular values (and will hence have $n - m$ columns composed entirely of zeros) and $\mathbf{V}$ is an $n \times n$ matrix

of PC scores. We applied SVD to a random sample of nearly 40,000 IRIS BT spectra. A sample of this size is statistically representative of the full set of IRIS spectra, so the calculated EOFs and PC scores are also representative of the full set of spectra. Normalized scores for the ten leading EOFs ($U_{1..10}$) were computed for every IRIS spectrum, as shown in equation 3.2.

$$s_{\mathrm{i,n}} = \frac{T_{\mathrm{resid_i}} \cdot U_{\mathrm{n}}}{\lambda_{\mathrm{n}} {n_{\mathrm{sam}}}^{-0.5}} \tag{3.2}$$

where $s_{\mathrm{i,n}}$ is the normalized score for the $i^{\mathrm{th}}$ spectrum and $n^{\mathrm{th}}$ EOF, $T_{\mathrm{resid_i}}$ is the $i^{\mathrm{th}}$ residual BT spectrum (after subtracting mean of all spectra) and $n_{\mathrm{sam}}$ is the number of spectra in the random sample. The denominator normalizes the PC scores, so that the score represents the number of standard deviations from the mean. Any spectra with negative/positive scores outside of a selected threshold in any of these EOFs were then discarded. If the PC scores are assumed to have a normal distribution, the expected number of spectra $n_{\mathrm{exp}}$ with z-scores greater than $|x|$ can be approximated by equation 3.3.

$$n_{\mathrm{exp}} = 2 n_{\mathrm{EOF}}\, n_{\mathrm{IRIS}} (1 - \mathrm{cdf}(x)) \text{ for } n_{\mathrm{exp}} \gg n_{\mathrm{IRIS}} \tag{3.3}$$

where $n_{\mathrm{EOF}}$ is the number of EOFs (ten), $n_{\mathrm{IRIS}}$ is the total number of IRIS spectra and *cdf* is the cumulative distribution function of the standard normal distribution.

Figure 3.7 shows histograms of all normalized PC scores for EOFs 1-10. The PC scores for EOFs 1-4 are not normally distributed and so the threshold method was not applied for these EOFs. For EOF 1, the positive scores appear to be normally distributed but the negative scores are not. The negative scores belong to spectra with are not greatly affected by cloud and the positive scores to colder, cloudier spectra. For the other EOFs it is difficult to provide a physical interpretation of the distributions of PC scores. Figure 3.8 shows a plot of the leading five EOFs and Table 3.2 shows the percentage of the total variance explained by these five EOFs. Almost 95% of the total variance is explained by EOF1, which exhibits the same spectral shape (but inverted)

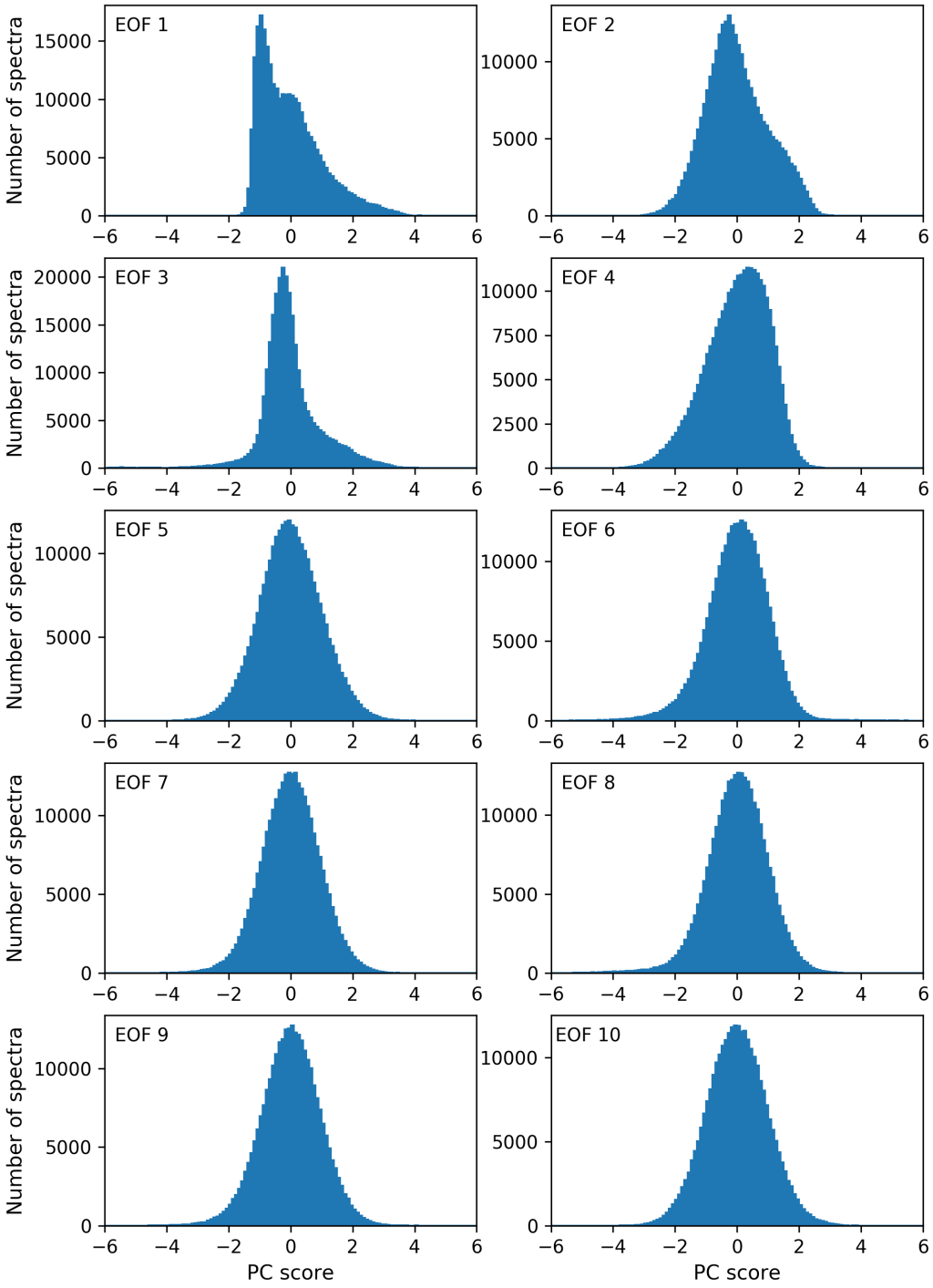



Figure 3.7: Histogram of scores for EOFs 1-10, computed from the whole set of IRIS spectra.

of a typical IRIS BT spectrum. The leading five EOFs explain over 99% of the total variance.

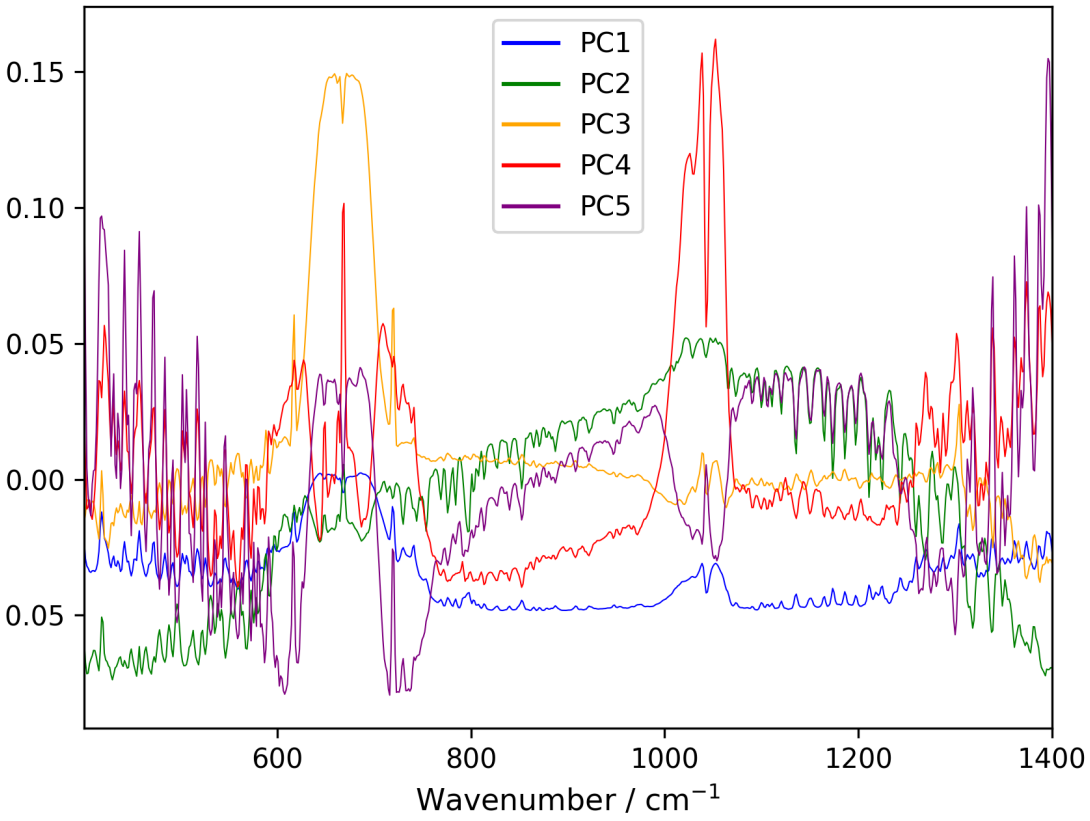



Figure 3.8: Leading 5 EOFs, computed from a random sample of ≈ 40,000 IRIS spectra.

| **EOF number** | **Variance (%)** | **Cumulative variance (%)** |
|---|---|---|
| 1 | 94.6 | 94.6 |
| 2 | 3.0 | 97.6 |
| 3 | 0.9 | 98.5 |
| 4 | 0.5 | 99.0 |
| 5 | 0.2 | 99.1 |

Table 3.2: Percentages of explained variance for the leading 5 EOFs.

Based on EOFs 5-10, expected and observed numbers of spectra with scores outside $|x|$ were calculated for a range of $|x|$. Starting with $|x| = 5$ and decreasing $|x|$ in increments of 0.1, the increases in expected and observed numbers of spectra with scores outside $|x|$ were compared until the increase in expected number surpassed half of the increase in observed number. This is a pragmatic way of making a compromise between the number of 'probably good' spectra lost and the number of 'probably bad' spectra discarded. This process decides the value of $|x|$ used as a threshold for discarding spectra in this QC stage. If SVD is carried out on the remaining spectra after this process, the recalculated EOFs and PC scores change, allowing further potentially bad spectra to be removed. This process was iterated, with the first and second iterations removing 2521 and 2228 spectra respectively. No spectra were removed in the third iteration, so further iterations were unnecessary.

After QC4 it was discovered that about 1600 of the remaining spectra are repeated at least twice in the IRIS dataset. These repeated spectra have near-identical radiances, times and locations although are often attributed to different orbits. This must be the result of observations being recorded twice in the original IRIS dataset and so these spectra were discarded. A density plot of BTs for the remaining spectra is shown in Figure 3.9. Many spectra observed from the coldest and warmest scenes have been lost. The test used in QC3 largely explains the former and the removal of land-based spectra largely explains the latter. A possible artefact remains in the data, however. This is seen as a line at 250-251 K where there are fewer observations than the surrounding points. This artefact was present before the QC and is also seen before and after the QC of *Poli and Brunel* (2016). At this stage, it can not be explained but should not have any implications for the rest of our work.

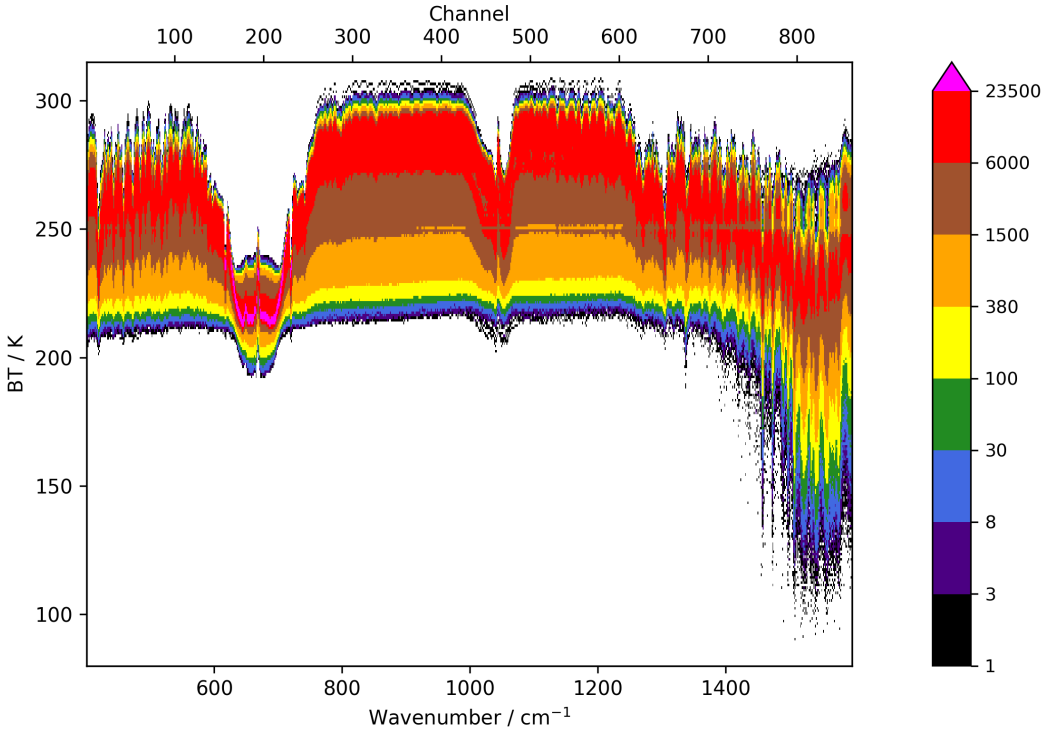



Figure 3.9: Density plot of all brightness temperatures for the data passing all QC stages (1-4).

## 3.6 Assessment of QC'd IRIS data

Simulating the IRIS spectra with RTTOV enables the quality controlled spectra to be evaluated to some extent. RTTOV is able to simulate the radiative effect of clouds, including scattering, if profiles of relevant cloud properties are included in the state vector. ERA-20C provides profiles of cloud liquid water, cloud ice water and cloud fraction during the IRIS period, allowing 'all-sky' simulations of the IRIS observations to be run. RTTOV requires the type of cloud (cumulus or stratus) to be specified for cloud liquid water and assumes that any cloud ice water is due to cirrus cloud. It was found that the type of cloud has an impact on simulated ToA BTs only of the order of 0.01 K. Therefore, the simplistic but adequate assumption was made that any cloud below 3 km was stratus and above 3 km, cumulus. The ERA-20C profile nearest in space and time to each IRIS observation was used for these simulations. It was

later found that this method occasionally took profiles over land and therefore any simulations of IRIS observations with a sub-satellite point within 100 km of land were discarded.

Figure 3.10 shows density plots of simulated BTs and the corresponding observed-simulated BT differences. Comparing the simulations to the observations shown in Figure 3.9, the most obvious difference is the absence of BTs below about 190 K in the simulations. The very low BTs in the observations are unphysical and due to the very weak signal to noise ratio of the radiances at the high wavenumber end of the spectrum. This means that observations at wavenumbers $>1400\,\mathrm{cm}^{-1}$ have to be treated with caution. Figure 3.10 shows that in the window channels especially, the distribution of BT differences is very wide, with a large proportion of differences exceeding 10 K. This reflects the difficulty in accurately simulating the IRIS observations for cloudy scenes, due to errors in the cloud amounts in ERA-20C. The mean observation-simulation difference is -3 to -4 K over most of the window channels. It is unlikely that the observations are systematically 3 to 4 K too cold, so this suggests that there is a warm bias in the simulations which accounts for at least part of the observation-simulation difference. Such a warm bias could be due to ERA-20C systematically under-representing cloud. Alternatively, the bias could be due to RTTOV, but RTTOV is a well-established model whose ability to simulate under clear and cloudy conditions has been well-validated (*Saunders et al.*, 2013). In section 3.7, clear-sky RTTOV and PCRTM simulations are compared to investigate potential forward model error. Chapter 4 presents evidence for there being a bias in the IRIS radiances, which could explain part of the mean observation-simulation difference seen here.

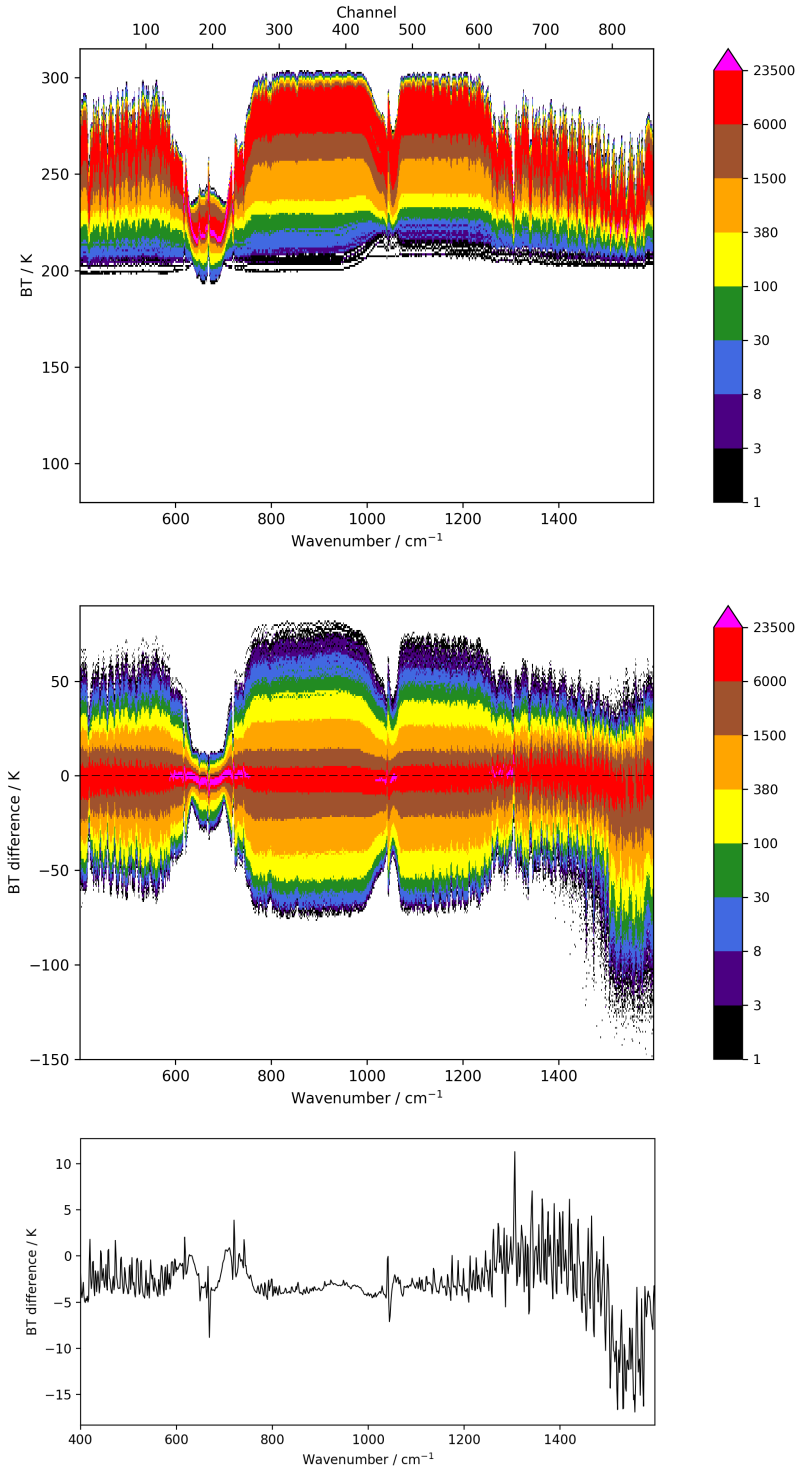



Figure 3.10: Density plot of 'all-sky' simulated brightness temperatures (top), observation-simulation differences (middle) and plot of the mean difference spectrum (bottom).

## 3.7 Comparison of RTTOV with PCRTM

PCRTM-based clear-sky IRIS simulations for all observations in April 1970 were obtained from R. Bantges (pers. comm., 2016). The PCRTM and RTTOV-based simulations both use ERA-20C to provide the state vectors, but it was necessary to re-run the RTTOV simulations using the same level of temporal interpolation between ERA-20C time-steps as used in *Bantges et al.* (2016). There is some difference in the vertical resolution of the ERA-20C profiles used for RTTOV and PCRTM. The ERA-20C profiles for PCRTM were interpolated from 37 fixed pressure levels to the 101 levels required for PCRTM. The lowest pressure level is at 1000 hPa, so profile values at pressures $>1000$ hPa were assumed to equal those at 1000 hPa. The ERA-20C profiles for RTTOV were defined on the native 91 model levels and interpolated to the 101 levels required for RTTOV. The effect of this difference is likely to be negligible, as clear-sky simulations are far more sensitive to SST than atmospheric temperature and water vapour. Four different groups of PCRTM simulations corresponding to four different orbits were found to be erroneous so were removed before comparing the RTTOV and PCRTM simulations.

The RTTOV and PCRTM simulations were compared in the 10-12 µm window region, by integrating the radiances over this interval using Simpson's rule. The RTTOV and PCRTM window radiances were then gridded at 10° resolution. Figure 3.11 shows the gridded RTTOV-PCRTM radiance difference. It shows that the two sets of clear-sky simulations agree to within 0.5 W cm$^{-2}$ sr cm (0.3 K at 300 K) over nearly all gridboxes, with larger differences corresponding to gridboxes with fewer observations. Figure 3.12 shows the mean RTTOV-PCRTM BT spectrum, calculated from all cases in April 1970. Across the window channels (780-980 cm$^{-1}$ and 1070-1220 cm$^{-1}$), the difference is very close to zero on average. These are the main channels used for SST retrieval, so it is encouraging that the two RTMs agree so well here. The large differences across the $CO_2$ absorption region (600-780 cm$^{-1}$) are probably due to different

assumptions about $CO_2$ concentrations, and does not indicate a discrepancy between the physics of the two RTMs.

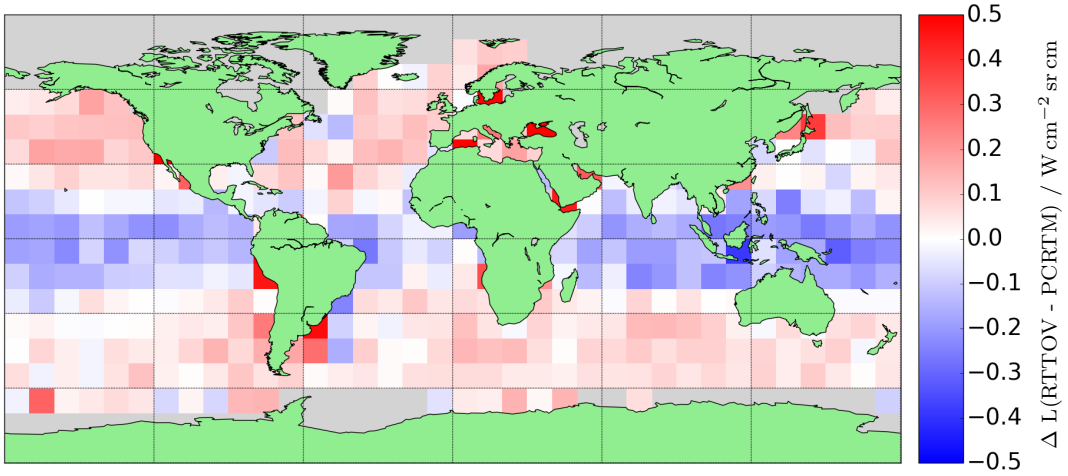



Figure 3.11: Gridded RTTOV-PCRTM clear-sky radiance differences for April 1970.

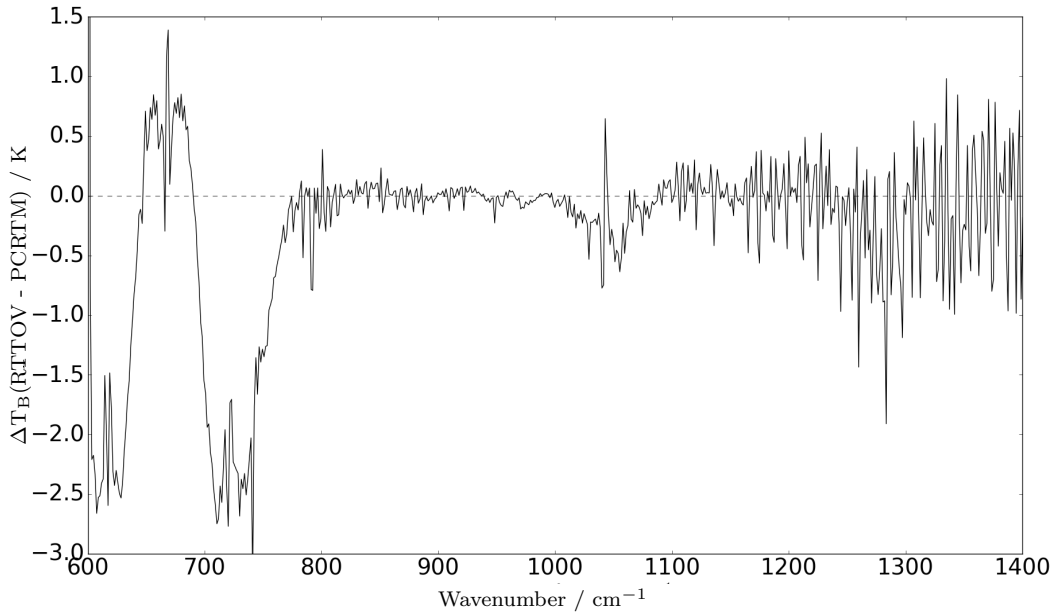



Figure 3.12: Mean RTTOV-PCRTM clear-sky brightness temperature difference spectrum. This represents a near-global average for April 1970.

## 3.8 Chapter summary

This chapter has explained, step-by-step, the quality control tests and processing applied to the dataset of IRIS spectra. Table 3.3 shows how many spectra were removed in each QC/processing stage. Over half of the original set of spectra were discarded in the land and sea-ice removal stage. QC3 removes far fewer spectra than the other QC stages, which possibly suggests that a stricter threshold could have been imposed for the minimum IRIS-simulated correlation.

| **QC stage** | **Number of spectra remaining** | **Number of spectra removed** |
|---|---|---|
| pre-QC | 654,488 | — |
| QC1 | 645,668 | 8,820 |
| QC2 | 631,555 | 14,113 |
| land and sea-ice removal | 286,846 | 344,709 |
| QC3 | 285,681 | 1,165 |
| QC4+ | 279,300 | 6,381 |

Table 3.3: Summary of numbers of IRIS spectra removed in each stage of QC/data processing. QC4+ includes removal of about 1600 repeated spectra.

Assessment of the quality of the QC'd IRIS spectra is difficult due to the absence of direct validation data, but a comparison was made with all-sky RTTOV simulated spectra. Assessment of individual IRIS spectra by comparison with individual simulations is not possible, because of the inability of reanalyses to accurately represent cloud over small spatiotemporal scales. However, comparison of spectral averages over large areas and longer timescales is worthwhile. Averaging over all observation-simulation pairs reveals that the observations have a relative cold bias of 3-4 K over the window channels. This average represents a near annual and near global mean.

Chapter 4 will discuss evidence for there being a bias in the IRIS radiances, which could explain part of this 3-4 K bias. This will need to be quantified and removed from the IRIS data in order to retrieve high-quality SST estimates. A warm bias in the simulations accounts for the remainder of the 3-4 K bias. The possibility of

forward model error contributing to this bias was investigated by comparing RTTOV and PCRTM simulations. RTTOV and PCRTM agree to within 0.2 K on average across most of the window channels in clear-sky conditions, though there may be greater differences in cloudy conditions.

The warm bias in the simulations is expected to result from under-representation of cloud in ERA-20C. The SST retrieval method will rely upon good quality a priori information about the surface and atmospheric state at the locations of the IRIS observations. It will therefore be beneficial to consider other sources of reanalysis data, with the aim of achieving better quality simulations in cloudy conditions.

# CHAPTER 4

# THE IRIS CALIBRATION BIAS

## 4.1 Background and origin of the bias

An assessment of the quality of the IRIS dataset was made in chapter 3, using a series of QC tests to remove 'bad' spectra. *Hanel et al.* (1972) claimed that the IRIS dataset is 'high quality', owing to the work put in to calibrate and screen the data during and shortly after the IRIS period. The instrument was exposed to a warm blackbody calibration source and the cold interstellar background roughly every four minutes while in orbit. The temperature of the warm blackbody was monitored to a precision of 0.2 K with a platinum wire element (*Hanel et al.*, 1971). Higher order corrections to the calibration involved measurement of the emissivity of the on-board calibration source and correcting effects caused by orbital variations. Despite these efforts to calibrate the raw measurements made by IRIS, this chapter presents evidence that a substantial bias remains in the IRIS radiance spectra, that if left uncorrected will prevent us from being able to retrieve SSTs from IRIS to the desired level of accuracy.

There is some evidence for such a bias in *Bantges et al.* (2016), where the global mean brightness temperature spectrum for IRIS was compared with two more recently operating hyperspectral radiometers: the Interferometric Monitor for Green-

house Gases (IMG) and IASI. The comparison is made over the April-June period using data from the years 1970 (IRIS), 1997 (IMG) and 2012 (IASI). In this comparison, the IMG and IASI spectra were degraded to match the IRIS spatial and spectral resolution. Figure 4.1 shows the spectral difference in BT between each pair of instruments. Differences of up to 2 K are found in the window channels in the interval 800-1000 $\mathrm{cm}^{-1}$ for IMG-IRIS and IASI-IRIS. This is despite differences being $< 0.25$ K

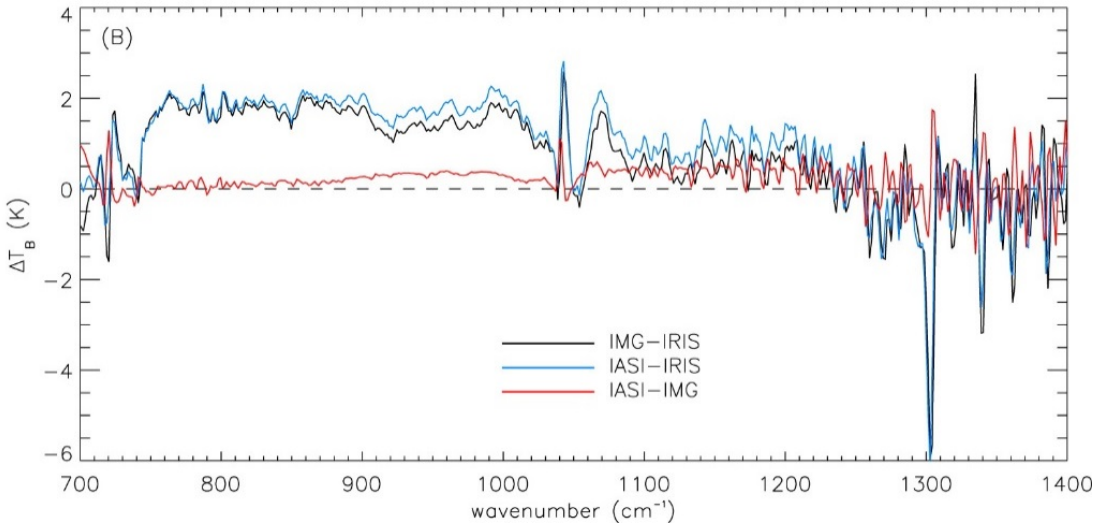



Figure 4.1: Area-weighted global mean brightness temperature differences averaged over 3 months (AMJ), between IRIS (1970), IMG (1997) and IASI (2012). Reproduced from *Bantges et al.* (2016).

in this spectral region for IASI-IMG. Interannual variability in global mean BT is driven largely by interannual variability in the spatial distribution of cloud in terms of cloud amount and cloud height. It is also driven to a lesser extent by interannual variability in global mean temperature at the altitudes sensed by the satellite instruments, which varies spectrally. Changes in concentrations of trace gases between 1970 and 1997 have been shown to have impacts on BT at some wavelengths of interest (*Harries et al.*, 2001). For example, the increase in concentration of CFC-11 and CFC-12 over this period has reduced the global mean BT by the order of 0.5 K in the intervals 840-850 $\mathrm{cm}^{-1}$ and 910-930 $\mathrm{cm}^{-1}$. However, these changes explain only a very small fraction of the IMG-IRIS difference in Figure 4.1. Changes in global mean cloud amount/height

and tropospheric/surface temperature may be able to explain a larger fraction of the IMG-IRIS and IASI-IRIS differences but the radiative impact of such changes does not show a strong differential window signal, as seen in the observed differences with IRIS. It is therefore highly unlikely that the differences with IRIS are a result of interannual variability alone.

From these results, *Bantges et al.* (2016) suggest that IRIS' calibration should be investigated, but as the authors note, the IRIS dataset lacks the raw interferograms and calibration parameters needed to accurately assess the calibration. Figure 2.7 (shown in chapter 2) shows a small step change in the emissivity estimate of the onboard blackbody calibration source. *Bantges et al.* (2016) point out that this step change is consistent with the apparent step change in the IMG-IRIS and IASI-IRIS difference spectra in Figure 4.1. Therefore, it could be that an error in the spectral emissivity applied to the on-board blackbody source explains the apparent IRIS calibration bias. This kind of error should produce a bias dependent on the scene temperature.

This theory was tested in *Bantges et al.* (2016) by comparing the brightness temperature spectra from IRIS, IMG and IASI as a function of scene temperature. The spectra were separated into 'cold', 'mild' and 'warm' scenes according to peak BT. The distributions of BT difference between two window channels either side of the step were then compared for each type of scene. It was found that the mean of the distribution for each instrument is nearly the same for the cold scenes, but for the mild and warm scenes the IRIS mean is increasingly less than for IMG and IASI. This suggests that although IRIS may have an overall bias relative to the other instruments for the cold scenes, there is no step change in this bias as seen in Figure 4.1. It suggests that the step becomes more pronounced the warmer the scene is. However, the BT difference between just two channels is not necessarily a robust measure of the size of this step.

For this reason we have repeated this analysis, using the difference in mean BT between the spectral intervals 870-970 $\mathrm{cm}^{-1}$ and 1100-1200 $\mathrm{cm}^{-1}$ as a more robust measure of the step change. We compared the IRIS observations with RTTOV simulations

using ERA-20C profiles, instead of IMG and IASI. The full set of IRIS observations remaining after QC was used, except for cases where the sub-satellite point was within 100 km of land, to ensure that the matched NWP profiles are not over land. All-sky simulations were generated for each observation as described in §3.6. The simulations do not take into account the potential instrument calibration errors. The results are shown in Figure 4.2. The expected shift in the distribution to the right for warmer

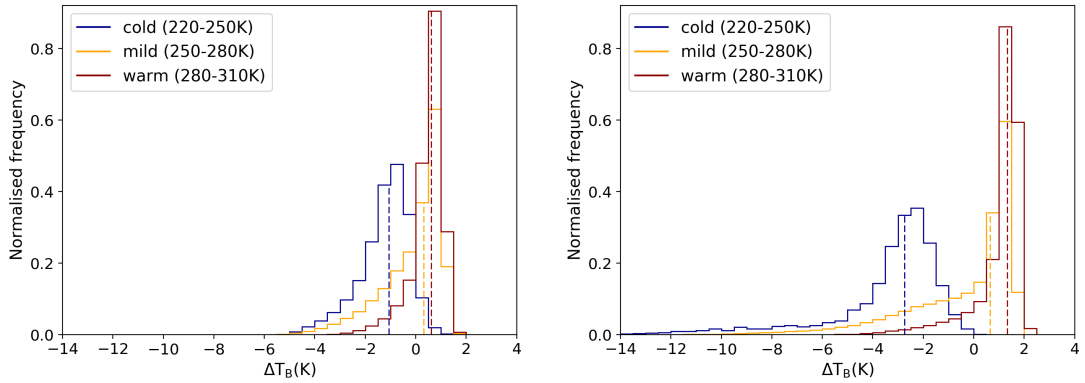



Figure 4.2: Histograms of the step change in BT (870-970 $\mathrm{cm}^{-1}$ minus 1100-1200 $\mathrm{cm}^{-1}$ mean BT) for the IRIS observations (left) and ERA-20C simulations (right), shown for the three different scene types. The median of each distribution is indicated by the dashed lines.

scenes is seen for both the observations and simulations. However, the shift in median BT is much greater for the simulations. For the simulations, the median step change is nearly 2 K more negative for the cold scenes and nearly 1 K more positive for the warm scenes. As the IRIS, IMG and IASI distributions for the cold scenes in *Bantges et al.* (2016) had nearly the same mean, this suggests that there may be errors in the ERA-20C simulations, although a very different measure of the step change was used.

To investigate the spectral characteristics of IRIS, mean global (ocean only) spectra were generated from the IRIS observations and simulations, and also from a set of 'IRIS-like' IASI observations for each category of scene temperature and compared. We explain the process of generating IRIS-like IASI observations in §5.4.1. For the sake of better comparability, the mean global spectra were generated in the following

way. Each set of BT spectra were converted to radiance spectra and then gridded at monthly, 5° by 5° resolution. For each scene type, the gridded spectra were averaged for each month, averaging only the gridboxes common to all three sets of cases. The mean monthly global spectra were then averaged to generate mean global spectra for the entire period (April-December 1970/2011). The mean global spectra are compared in Figure 4.3. For the cold scenes, the IRIS and IRIS-like IASI observations agree to

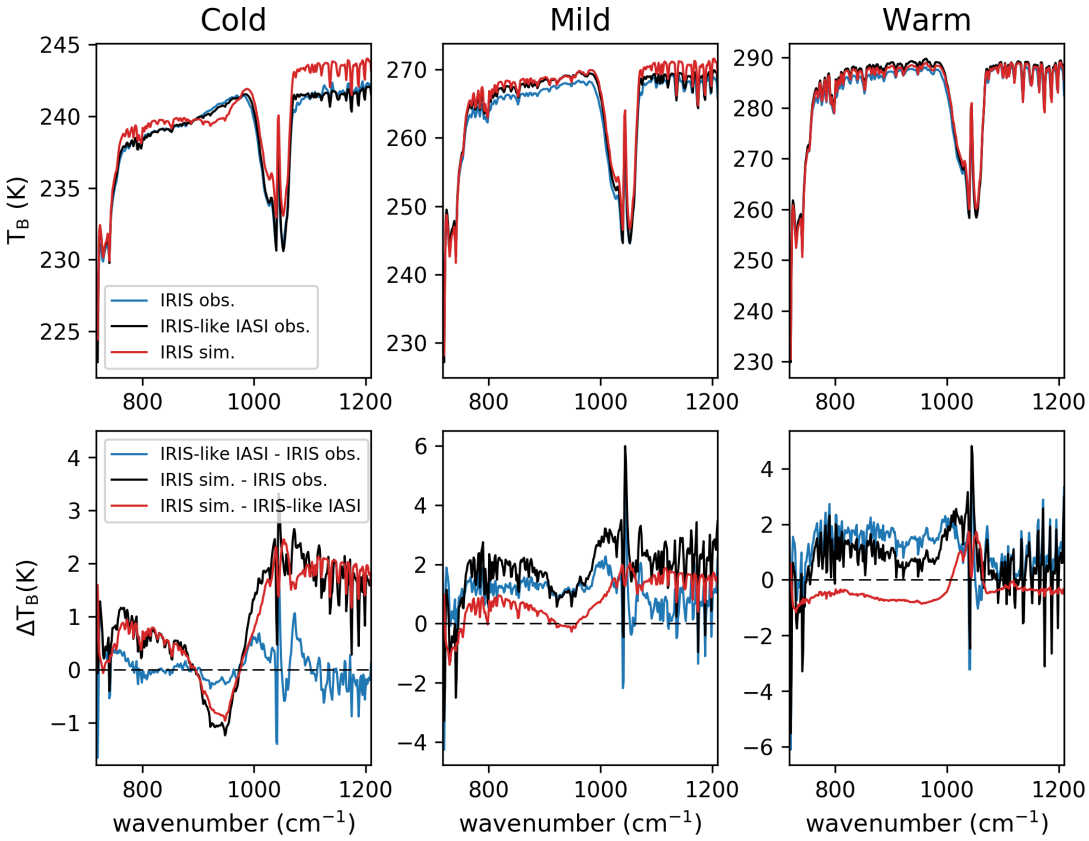



Figure 4.3: Area-weighted global (ocean only) mean BT spectra for April-December averages of the IRIS observations, IRIS simulations and IRIS-like IASI observations (top panels) and corresponding difference spectra (bottom panels) for the cold, mild and warm scenes.

within 0.5 K for most channels. The simulated spectrum is warmer than the observed spectra except where there is a noticeable dip centred at 930 cm$^{-1}$. This may be an indication of forward model (RTTOV) error that more strongly affects colder scenes

and helps to explain the magnitude of the median in the cold distribution for the simulations in Figure 4.2. The differential window signal in the IRIS-like IASI - IRIS difference becomes increasingly stronger for the mild and warm scenes, being most similar to the IASI-IRIS differential window signal in Figure 4.1 for the warm scenes.

Although Figure 4.3 tells us something about the spectral behaviour of IRIS with scene temperature, it does not give us a good estimate of the magnitude of the bias in the IRIS observations. This is because the IRIS and IRIS-like IASI observations are being compared in different years and due to potential biases in the IRIS simulations. In the following sections, other types of comparisons with IRIS will be made in order to better quantify the magnitude of the bias.

## 4.2 Comparison with clear-sky simulations

One way in which we can try to better quantify the IRIS bias is by comparison of observed and simulated spectra for clear-sky scenes only. The main source of error in the simulation of cloud affected IRIS observations is from error in the NWP cloud data, because cloud distributions are less well represented in NWP reanalyses than atmospheric and surface temperature, and water vapour. This means that if the IRIS observations can be successfully screened for cloud, comparison of the remaining observed and simulated spectra should provide an improved estimate of the spectral bias.

### 4.2.1 a priori state

Data from the CERA-20C reanalysis became available at this time. CERA-20C is the ECMWF 10-member ensemble based on the ocean-atmosphere coupled data assimilation system CERA (*Laloyaux et al.*, 2016). Using a combination of ERA-20C and CERA-20C to provide the profiles for simulation of IRIS spectra should improve the quality of the simulations. For each IRIS observation, an ensemble of ERA-20C and CERA-20C states were obtained. ERA-20C is deterministic, so an ensemble was con-

structed by taking the state at each gridpoint within 100 km of the centre of the IRIS footprint for the three time steps closest to the observation time. This was combined with the ensemble of ten CERA-20C members taken from the nearest gridpoint and time step to the IRIS observation. There are generally different numbers of ERA-20C and CERA-20C profiles in the combined ensemble, so the weighted median of the profiles was computed to provide the best estimate.

### 4.2.2 Simulating aerosol effects

Aerosols can significantly attenuate ToA radiance at IR wavelengths in clear-sky conditions. RTTOV is able to simulate the effects of aerosols. Number density profiles of specific aerosol particle types must be specified. It is not possible to obtain these profiles from NWP data, but climatological profiles can be generated for ten different compositions (*Hess et al.*, 1998). There are three relevant compositions for marine environments: tropical, clean and polluted. The first two compositions consist of water soluble aerosols and sea salt. The polluted composition exhibits a much higher concentration of water soluble aerosols and also some soot. At infrared wavelengths, the effects of water soluble aerosols and soot on measured ToA BT are minimal, because of the small size of these particles. This means that there is very little difference between the clean and polluted profiles, in terms of the effect on ToA BT. It should be noted that these three compositions do not include mineral dust, which has a strong attenuating effect on the ToA BT in typically dusty regions such as the East tropical Atlantic. Observations located in dusty regions are discarded as described in §4.2.3. For each observation, either tropical or clean aerosol profiles were generated, depending on the latitude.

RTTOV aerosol coefficients are unavailable for IRIS, but simulations can be run using IASI aerosol coefficients at the spectral resolution of IASI. These simulations can be spectrally degraded to IRIS spectral resolution, as described in §5.4.1. It was found that the mean effect of aerosols on a large set of simulated IRIS spectra was about

0.09 K in the window regions. This is more than double that found in *Embury et al.* (2012), when simulating the effect of marine aerosols on BTs for the 11 and 12 µm channels of the Along Track Scanning Radiometer (ATSR). Part of this discrepancy can be explained by differences in the aerosol profiles used. *Embury et al.* (2012) take aerosol profiles from the Global Aerosol Data Set (GADS), where the distribution of different types of aerosol vary geographically. This is likely to result in more accurate aerosol profiles compared with using either standard clean or tropical compositions. In light of this, the aerosol concentrations generated by RTTOV were halved before running simulations with aerosols.

### 4.2.3 Cloud detection

Cloud detection methods were discussed in chapter 2. Our choice of cloud detection method is restricted due to the size of the IRIS footprint and the presence of the spectral bias.

To bypass these challenges, we use a method of cloud detection independent of the IRIS observations, exploiting visible images of cloud cover produced by the Image Dissector Camera System (IDCS) onboard the Nimbus-4 satellite. These images are collocated with the IRIS measurements and cover nearly all of the IRIS period. Figure 4.4 shows an example of an IDCS image. Areas of cloud are easily distinguished from much darker regions of clear sky. However, there is not a distinct threshold separating clear and cloudy pixels. A higher threshold for pixel greyscale value will produce a larger set of IRIS observations declared as clear-sky, but with more cloud contamination. A closer look at the IDCS data reveals that the images are not geolocated correctly. Geolocation errors of up to 50 km are typical, meaning that for each IRIS observation, a much larger area must be scanned for the presence of cloud to ensure that the IRIS observation is definitely cloud free. For the 30% of IDCS images where there is collocated THIR window channel data, a correlation-based method was used to partially correct these errors. The THIR measurements are known to be geolocated

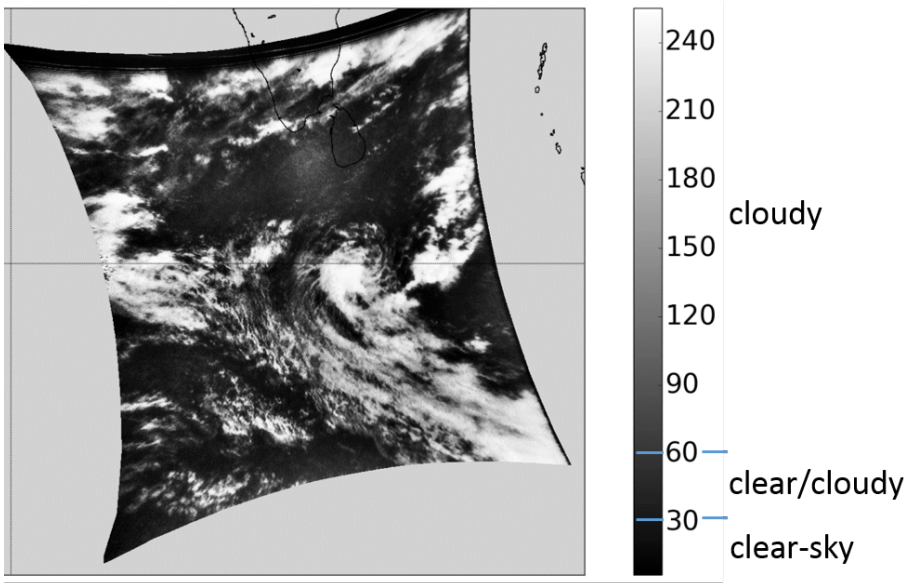



Figure 4.4: An IDCS image showing cloud cover to the south of India on 30$^{\text{th}}$ June, 1970. The scale shows the greyscale value of each pixel. From visual inspection, values of 30-60 correspond to pixels that may be clear or cloudy.

accurately (T. Eldridge, pers. comm., 2017). The collocation method exploits the negative correlation between IDCS pixel brightness and THIR window channel brightness temperature. Brighter pixels (higher greyscale value) are more cloudy, so generally correspond to lower brightness temperatures. Firstly, we look for IRIS scenes where there is sufficient IDCS and THIR data covering a 6° by 6° square with the IRIS observation at its centre. For these cases, the IDCS and THIR data were re-gridded to a common grid with 0.05° resolution. The correlation between the two sets of data was computed for different latitude/longitude offsets, in intervals of 0.05°. The offset producing the strongest correlation gave the correction applied to the IDCS geolocation. This method assumes that the IDCS geolocation can be corrected by translation of the latitude/longitude grid local to each IRIS scene. In reality, the geolocation error varies across the scale of a 6° by 6° square, so this method only partially corrects the IDCS geolocation.

From the geocorrected and uncorrected IDCS data, a set of clear-sky IRIS observations were found. A constant greyscale threshold of 50 was used to detect cloudy

pixels. The IDCS data have many small artefacts which are detected as cloudy pixels. This was allowed for by declaring an IRIS observation as clear if fewer than 1.5% of pixels within the footprint are cloudy.

Some cloud contamination is likely to remain in the cases declared as clear-sky. This can happen if geolocation error exceeds the amount allowed for or if the brightness threshold is not chosen correctly. Figure 4.5 shows the distribution of observed-simulated BT averaged over all window channels. The cold tail is evidence of cloud

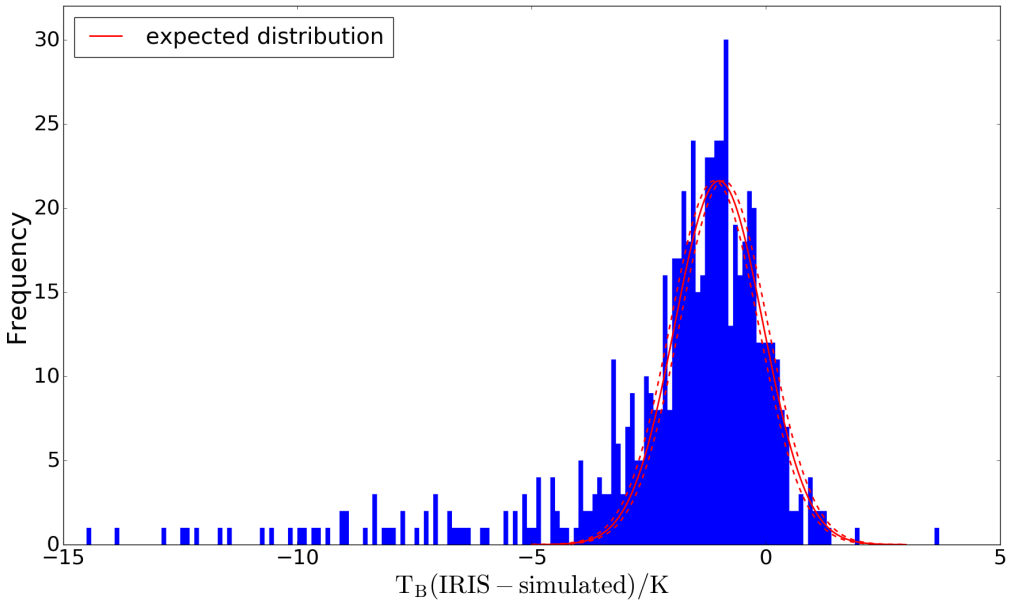



Figure 4.5: Distribution of IRIS - simulated BT averaged over all window channels for the observations identified as clear-sky from the IDCS data. The solid red curve indicates the expected normal distribution with no cloud contamination. The red dashed curves indicate the estimated uncertainty in this distribution.

contamination in some cases. We can assume that with no cloud contamination, the distribution of BT differences is normal. We can therefore assume that the right hand side of the distribution in Figure 4.5 is normal, since clouds rarely have a warming effect on the observed BT. This allows us to fit a normal curve to the distribution, using the data from the right hand side only. We have to estimate by eye where the peak of the distribution is, so this introduces some uncertainty into our bias model. We have assumed that the fitted normal distribution is centred at -1 $\pm$ 0.1 K. Any cases

outside of the central 80% of this normal distribution are discarded.

In addition, any observations located in typically dusty regions are discarded. This was done using a climatology of monthly 10 µm dust aerosol optical depth (AOD) from IASI, based on the period 2008-2017 (*Capelle et al.*, 2018). IRIS observations in locations with AOD $> 0.1$ were discarded. This leaves us with a set of about 420 IRIS observations with very little cloud or aerosol contamination. The next section describes how these observations are used to develop a model of the calibration bias.

### 4.2.4 Bias model

To determine a model of the spectral calibration bias, this set of 420 IRIS observations were compared to the corresponding set of clear-sky simulations. RTTOV appears unable to correctly simulate the amplitude of the water vapour absorption lines. It is also evident that these absorption lines are not simulated at exactly the same wavenumbers as they are observed. These two issues will be dealt with in turn.

#### 4.2.4.1 Spectral shift

When comparing IRIS and IMG spectra, *Harries et al.* (2001) derived an empirical multiplicative factor of 0.9995 to apply to the nominal IRIS wavenumber scale. A spectral shift of this order is expected due to angular extension of the detector in the interferometer field (C. Standfuss, pers. comm., 2019). This means that the IRIS wavenumber scale should be divided by this factor when simulating IRIS spectra with RTTOV. A new RTTOV optical depth coefficient file was created for IRIS by J. Hocking with this wavenumber correction applied. Figure 4.6 compares the median observed and simulated clear-sky spectra generated using the original and new ('shifted') optical depth coefficient files. The simulated water vapour absorption lines are much more closely aligned with the water vapour absorption lines in the observations using the shifted coefficient file. This greatly reduces the amplitude of the spikes in the difference spectrum.

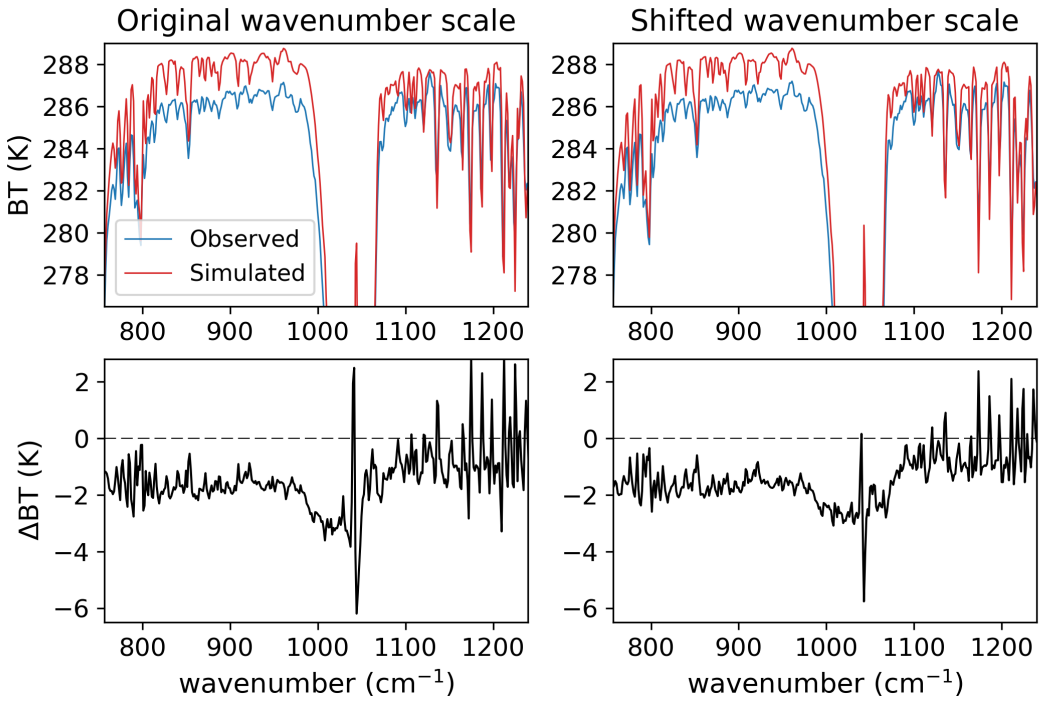



Figure 4.6: Median observed and simulated clear-sky BT spectra and difference spectrum (black line) generated using the original and 'shifted' (shifted wavenumber scale) optical depth coefficient files.

#### 4.2.4.2 Spectral resolution

The discrepancy between the amplitude of the observed and simulated water vapour absorption lines could be an indication of a positive bias in the NWP water vapour amounts. This possibility was investigated by reducing the water vapour amounts in the RTTOV profiles used to generate the simulations by 20 % at all levels. Figure 4.7 shows the effect of this reduction on the median observed-simulated BT spectrum. The amplitudes of the spikes in the difference spectrum are reduced by a relatively large amount. However, it is highly unlikely that the NWP water vapour is biased by as much as 20 %. This is because the range in monthly mean total column water vapour (TCWV) between three reanalysis products (ERA-20C, 20CRv2c and JRA-55) and an atmospheric model integration (ERA-20CM) is less than 4 % in 1970 (*Poli et al.*, 2016).

It is far more likely that the discrepancy in the amplitude of the water vapour ab-

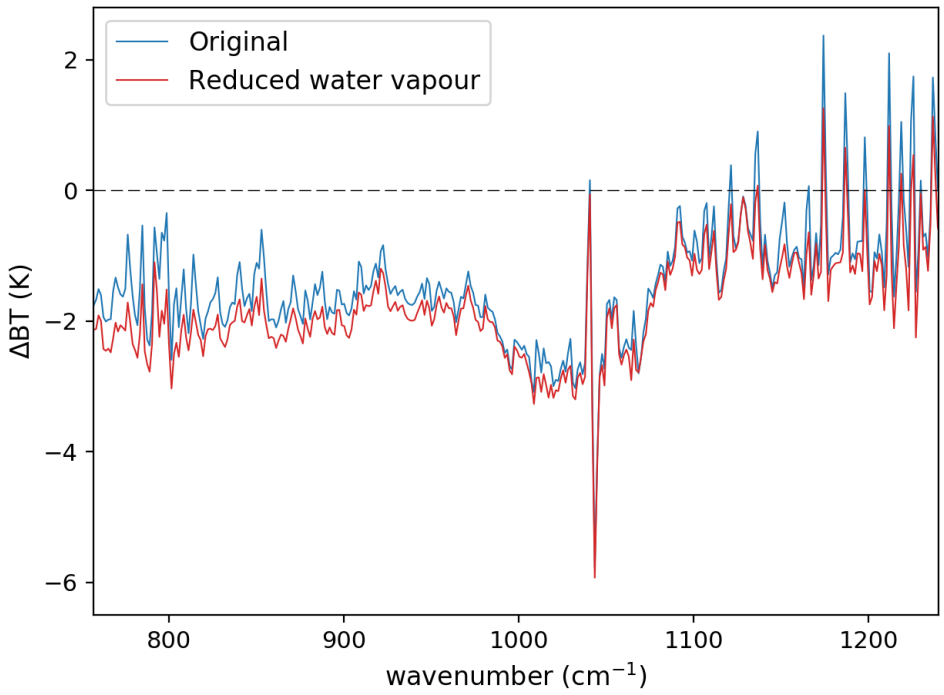



Figure 4.7: Median observed-simulated clear-sky BT spectrum for the original set of simulations and for the set of simulations with water vapour reduced by 20 %.

sorption lines is due to the nominal IRIS spectral response function (SRF) being incorrect. *Hanel et al.* (1971) suggest that the spectral resolution of IRIS after apodization is in reality closer to 3 cm$^{-1}$ than the nominal value of 2.8 cm$^{-1}$. This would increase the width of the SRF and, on average, reduce the amplitude of the water vapour absorption lines. In order to directly simulate IRIS spectra with a modified SRF, a new optical depth coefficient file must be created, which was not possible at this time. However, this can be done indirectly by simulating IASI spectra using the same NWP profiles and then spectrally degrading them to the desired spectral resolution (see §5.4.1). This method was used to investigate the effect of reducing the spectral resolution to 3 cm$^{-1}$. IRIS spectra were simulated in this way assuming spectral resolutions of 2.8 cm$^{-1}$ and 3 cm$^{-1}$. Reducing the spectral resolution from 2.8 cm$^{-1}$ to 3 cm$^{-1}$ was found to reduce the amplitude of the spikes in the median observed - simulated difference spectrum only very slightly. This means that a much larger change from the nominal spectral reso-

lution is required to remove these spikes. Sets of IASI simulations were produced for a range of different spectral resolutions, by varying the multiplicative factor applied to the IRIS SRF from 1 to 1.5 in increments of 0.01. Figure 4.8 shows how the mean observed - simulated radiance difference spectrum varies with SRF factor. There is greater sensitivity to changes in the SRF factor in spectral regions where there is more channel-to-channel variability in the radiance, such as the wing of the $CO_2$ absorption band ($< 770\,\mathrm{cm}^{-1}$) and water vapour absorption lines. A first attempt was made to

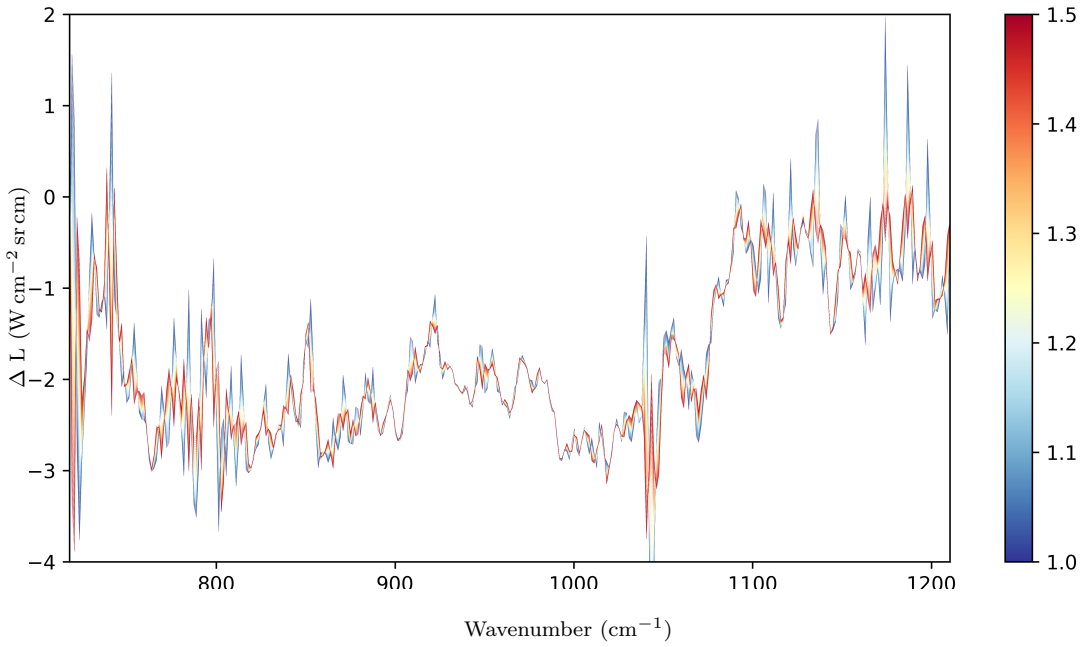



Figure 4.8: Mean IRIS - simulated spectral radiance difference for different choices of the multiplicative SRF factor.

estimate the correct SRF factor by minimizing the standard deviation in the mean difference spectrum. However, visual inspection of the resulting difference spectrum showed that there were some spectral artefacts, such as pairs of adjacent spikes when a single spike is expected. It was found that the IMG-IRIS and IASI-IRIS difference spectra from Figure 4.1 could be used as a reference to estimate the optimal SRF factor by eye. To illustrate this, Figure 4.9 shows two zoomed-in sections of Figure 4.8. The black line shows our best estimate of the SRF factor, which varies with wavenumber.

The factor was observed to increase from 1.20 at 720 $\mathrm{cm}^{-1}$ to 1.32 at 1210 $\mathrm{cm}^{-1}$, and was assumed to increase linearly in this spectral interval.

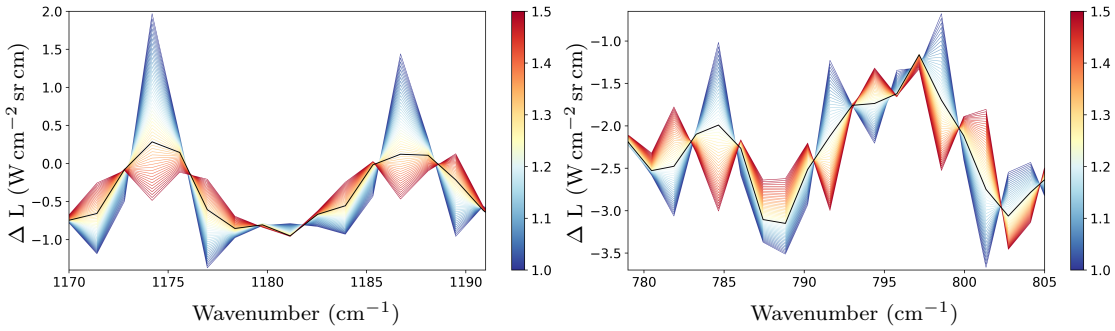



Figure 4.9: As figure 4.8 but zooming in to 1170-1191 $\mathrm{cm}^{-1}$ (left) and 779-805$\mathrm{cm}^{-1}$ (right). The black line shows the estimate of the optimal SRF factor.

#### 4.2.4.3 Preliminary bias model

Having made corrections to the IRIS wavenumber scale and the spectral resolution, the resulting median radiance difference spectrum forms our preliminary estimate of the spectral calibration bias. If we assume that an error in the spectral emissivity applied to the on-board blackbody is the cause of the bias, it follows that the bias should be the same for all IRIS observations, if measured in terms of radiance. This means that in terms of brightness temperature, the magnitude of the bias decreases with observed scene temperature. This is in opposition to the increase in the relative IRIS minus IRIS-like IASI bias with scene temperature seen in Figure 4.3. As stated previously, the difference in the years being compared mean that this comparison does not necessarily provide a good estimate of the exact magnitude of the bias, so further investigation is needed in order to determine the correct dependence of the bias on scene temperature. Our preliminary bias model makes the assumption that the spectral radiance bias is the same for all IRIS observations.

Our estimate of the average spectral bias is compared to the independently derived IRIS-IMG and IRIS-IASI BT differences, before and after correcting the simulations

(Figure 4.10). For the sake of better comparability, the average bias estimate was

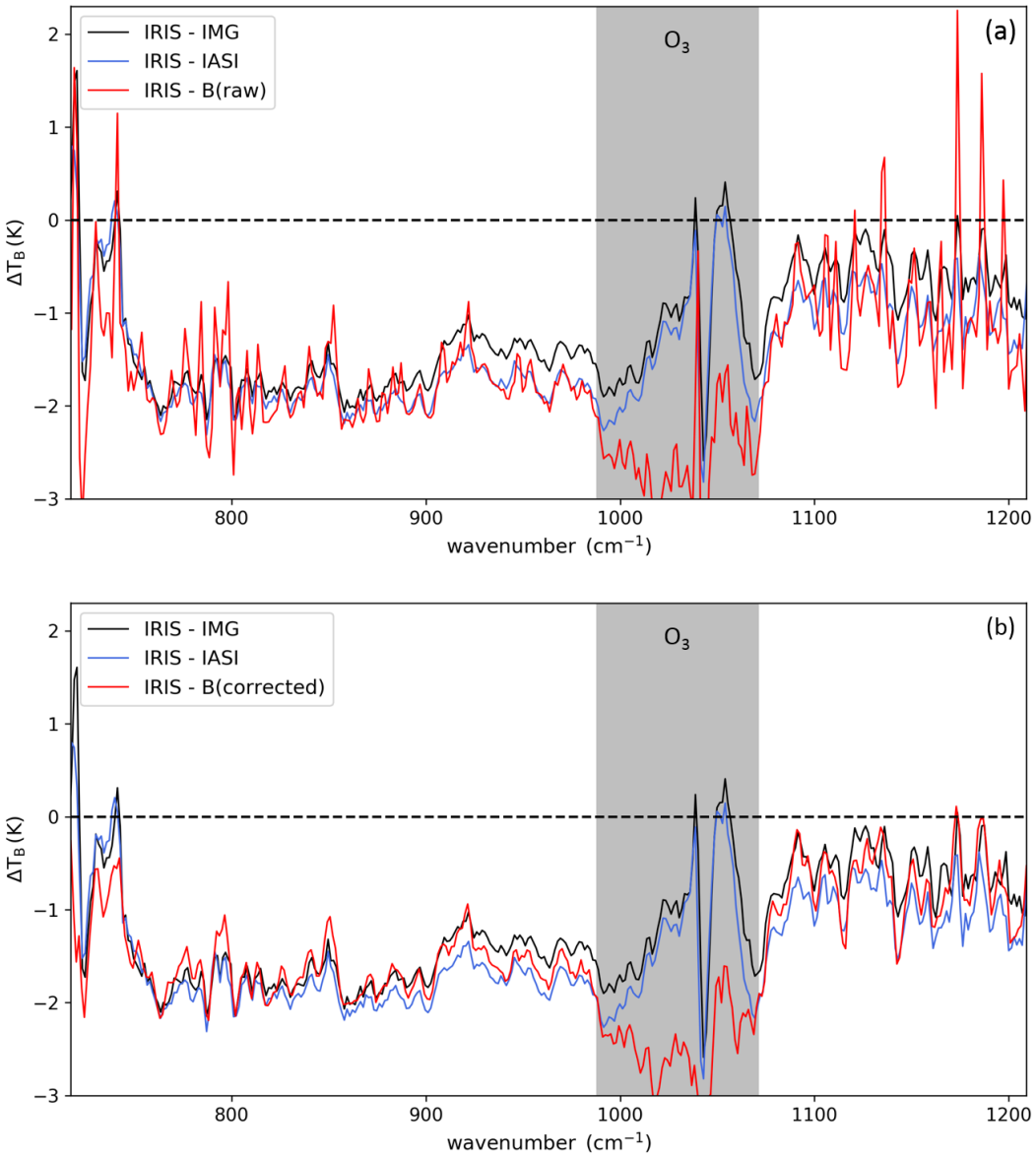



Figure 4.10: (a): IRIS bias estimate, area-weighted and averaged over AMJ, before correcting the simulations (red line). This is compared with area-weighted global mean BT differences averaged over AMJ between IRIS (1970), IMG (1997) and IASI (2012) (black and blue lines). (b): As (a), but after correcting the simulations. The ozone absorption band is shaded grey.

computed from IRIS spectra averaged over 5° by 5° cells and averaged over April-June (AMJ). The corrections to the simulations markedly improve the agreement between

our bias estimate and the IRIS-IMG and IRIS-IASI differences in most spectral regions. Apart from the ozone absorption band, which is irrelevant for SST retrieval, the greatest discrepancies are seen in the $CO_2$ absorption band wing. This is due to differences in global mean $CO_2$ concentration between 1970 and 2012. The level of agreement between our bias estimate and the IRIS-IMG and IRIS-IASI differences across the window regions suggests that our bias model largely corrects the biases in these channels, if the assumption about the constancy of the radiance bias is correct.

In the next section our preliminary bias model will be applied to the IRIS observations and comparisons will be made with BT measurements from the THIR, introduced in chapter 2. This will provide an independent assessment of the bias correction and can be used to evaluate our assumption about the constancy of the radiance bias.

## 4.3 Comparing IRIS with THIR

In order to compare IRIS and THIR BTs, the relevant IRIS channels must be combined to simulate the two THIR channels. The IRIS channels in the 1350-1600 $\mathrm{cm}^{-1}$ interval are required to simulate the THIR vapour channel, but in §3.6 it was shown that the IRIS BTs are often unphysically low at wavenumbers >1400 $\mathrm{cm}^{-1}$. Therefore, no meaningful comparison can be made for the THIR vapour channel so we compared IRIS and THIR for the THIR window channel only. As the THIR has a much higher spatial resolution than IRIS, it is also necessary to combine the THIR pixels located within the IRIS footprint. The spectral response function (SRF) defines the sensitivity of a sensor to incoming radiation at specific wavelengths over the range of a spectral band. The SRF is available for the THIR window channel and is used to weight the relevant IRIS channels combined to simulate this broadband channel. The THIR window channel SRF is plotted in Figure 4.11, superimposed on an IRIS radiance spectrum to show which part of the spectrum is used to simulate this THIR channel. The IRIS sensor has a relatively narrow SRF for each channel (the full width at half maximum (FWHM)

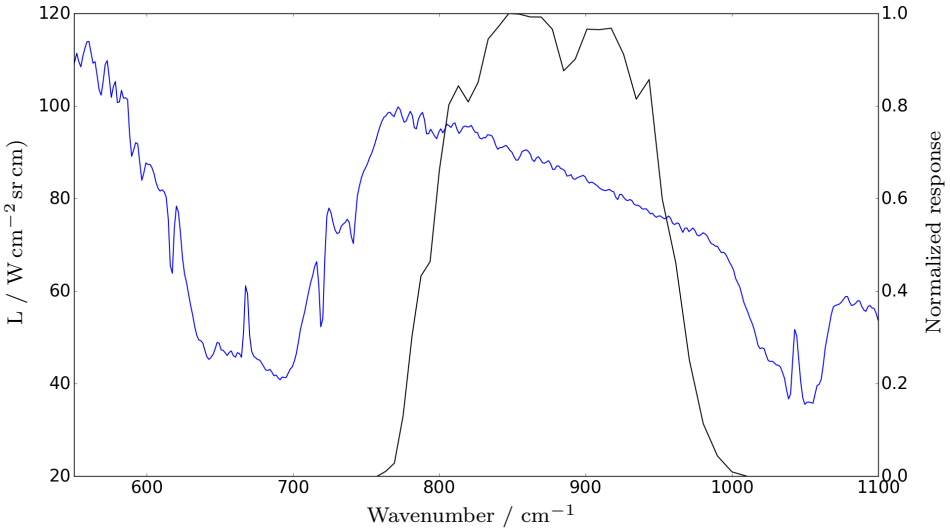



Figure 4.11: THIR window channel normalized SRF (black line) superimposed over an IRIS radiance spectrum (blue line). The IRIS channels inside the SRF curve are weighted by the SRF values at the centre of each channel in equation 4.1.

is 2.5 cm$^{-1}$) and so the approximation was made that each channel represents a single wavelength at the channel centre. Equation 4.1 describes how the IRIS radiances $L_{\mathrm{IRIS}}$ are simulated for the THIR window channel.

$$L_{\mathrm{IRIS}} = \frac{\sum_i L_i SRF_i}{\sum_i SRF_i} \tag{4.1}$$

where $L_i$ is the IRIS radiance for the $i^{\mathrm{th}}$ channel and $SRF_i$ is the THIR SRF interpolated to the $i^{\mathrm{th}}$ channel wavenumber. In order to average the THIR pixels over each IRIS footprint, the THIR BTs were converted to radiances using a lookup table (T. Eldridge, pers. comm., 2017). This computes the THIR channel radiance for BT values in 0.1 K increments in the range 150-350 K. Equation 4.2 describes how the THIR radiance $L_{\mathrm{THIR}}$ for the $j^{\mathrm{th}}$ BT value is calculated.

$$L_{\mathrm{THIR_j}} = \frac{\sum_i B(\nu_i, T_j) SRF_i}{\sum_i SRF_i} \tag{4.2}$$

where $B(\nu_i, T_j)$ is the Planck Function for the $i^{\text{th}}$ IRIS channel wavenumber $\nu$ and $j^{\text{th}}$ brightness temperature $T$. The mean of the THIR radiances over the IRIS footprint provides a radiance directly comparable with that obtained for IRIS from equation 4.1. These radiances were converted to brightness temperatures via the lookup table before comparisons were made.

In order to quantify the relative bias between IRIS and THIR it is important to ensure that for the cases compared, the THIR BTs are both relatively complete and reasonably homogeneous over the IRIS footprint. The first of these criteria can be quantified by ensuring that the number of THIR pixels in the IRIS footprint exceeds a given threshold. A threshold of 130 pixels was chosen, representing approximately the number of THIR pixels needed to cover at least half of the IRIS footprint. The second of these criteria aims to ensure that either there is zero cloud contamination in a scene or that any cloud is homogeneous throughout the scene. In such scenes, the observed variability in THIR BT must result from a combination of instrument noise and background variability in a combination of SST, atmospheric temperature and water vapour. Both of these terms can be estimated. For these scenes, the standard deviation in BT, $s_{\text{hom}}$, is given by $s_{\text{hom}} = (s^2_{\text{noise}} + s^2_{\text{nat}})^{0.5}$, where $s_{\text{noise}}$ and $s_{\text{nat}}$ are the standard deviations in BT due to instrument noise and natural variability of SST, atmospheric temperature and water vapour.

Estimates of the noise $\hat{\sigma}_{\text{noise}}$ in the THIR window channel BT have been calculated as a function of BT (i.e. scene temperature) (T. Eldridge, pers. comm., 2017). These noise estimates represent the population standard deviation (due to instrument noise alone). The expected value of $s_{\text{noise}}$ ($\hat{s}_{\text{noise}}$) in a sample matched to an IRIS footprint, will be smaller by a factor of $(1 - \frac{1}{n_{\text{THIR}}})^{0.5}$, where $n_{\text{THIR}}$ is the number of THIR pixels in the sample. Since $n$ is required to be at least 130, we make the approximation that $\hat{s}_{\text{noise}} = \hat{\sigma}_{\text{noise}}$. $s_{\text{nat}}$ is related to the natural variability in SST, atmospheric temperature

(T) and water vapour (q) via equation 4.3:

$$s_{\mathrm{nat}} = \sqrt{(\frac{\partial T_{\mathrm{B}}}{\partial SST} s_{\mathrm{SST}})^2 + \sum_n (\frac{\partial T_{\mathrm{B}}}{\partial T} s_{\mathrm{T}})^2 + \sum_n (\frac{\partial T_{\mathrm{B}}}{\partial q} s_{\mathrm{q}})^2} \tag{4.3}$$

We make the approximation that $s_{\mathrm{nat}} \approx s_{\mathrm{SST}}$, which is reasonable for clear-sky scenes for the case here of window wavelengths. To estimate the background SST variability at the THIR spatial resolution (7 km at nadir) within the sample requires high quality SST data with a similar or higher spatial resolution than the THIR. Such data is available from the European Space Agency (ESA) Sea Surface Temperature Climate Change Initiative project (SST CCI) for the period 1992-2010 (*Merchant et al.*, 2014). Specifically, the L4 analysis was used, which is derived from ATSR and AVHRR data and is produced daily with a grid resolution of 0.05°. Firstly, this data was regridded to 0.1° resolution to make it more similar to the THIR window channel resolution. For the 15$^{\mathrm{th}}$ day of each month of each year from 1992-2010, the standard deviation of the gridded SSTs within each larger 1° gridbox was computed. This gridbox size is similar to the size of the IRIS footprint. We then make the approximation that for each month, the gridded SST standard deviations in 1970 are equal to the mean of the gridded SST standard deviations in the years 1992-2010. We now have a method for estimating the expected value of $s_{\mathrm{hom}}$ ($\hat{s}_{\mathrm{hom}}$). The threshold for the THIR BT standard deviation ($s_{\mathrm{thresh}}$) is set to $\hat{s}_{\mathrm{hom}}$ plus one standard deviation of the gridded standard deviations in 1992-2010, to ensure that the majority of homogeneous scenes are declared as homogeneous. As an extra test for homogeneity, the range of THIR BTs in the sample was required to exceed a threshold of $8 s_{\mathrm{thresh}}$.

Table 4.1 summarises the thresholds used to select cases for comparison of IRIS and THIR.

This method was applied to find a set of about 4000 matched IRIS and THIR BTs for the THIR window channel. The IRIS BTs were calculated with and without the preliminary bias correction. In some cases, IRIS-THIR BT differences exceed 10 K. In

| **Statistic** | **Threshold** |
|---|---|
| number of pixels | $\geq 130$ |
| standard deviation | $< s_{\text{thresh}}$ |
| range | $< 8s_{\text{thresh}}$ |

Table 4.1: Summary of thresholds applied to the THIR data to choose cases for comparison with IRIS data.

many of these cases, these extreme differences are probably due to errors in the THIR BTs. For some orbits, there are systematic errors of several K in the THIR BTs. As a result, a test was applied to remove outliers in the IRIS-THIR BT differences. Cases where the modified z-score based on the median absolute deviation was greater than 3.5 were discarded.

Figure 4.12 shows the IRIS-THIR BT differences as a function of IRIS BT (effectively scene temperature) for the remaining set of cases, before and after the preliminary bias correction. Over most of the IRIS BT range, IRIS is about 2 K colder than THIR on average before bias correction. This reduces to about 0.5 K after bias correction. The remaining difference may potentially be due to a warm bias in the THIR BTs. Figure 4.12 shows that the relative IRIS-THIR bias is not constant with scene temperature, before or after the bias correction. Over half of cases are concentrated in the narrow temperature range where the IRIS BT is between 290-296 K. Here, the relative IRIS-THIR bias strongly reduces with increasing scene temperature and this feature remains after bias correction. This behaviour may be an artefact of our methodology or may be indicative of a real systematic effect in IRIS and/or THIR.

The IRIS - THIR BT differences were then plotted as a function of time to study the stability of the mean difference after bias correction (Figure 4.13). Most cases are concentrated in the first half of the IRIS period, so the standard errors in the BT difference tend to increase near the end of 1970 with much greater fluctuations in the mean differences. Linear least squares regression was applied to the binned mean differences to test for any overall trend. This showed that there was no significant drift

in the relative bias at the 5% level.

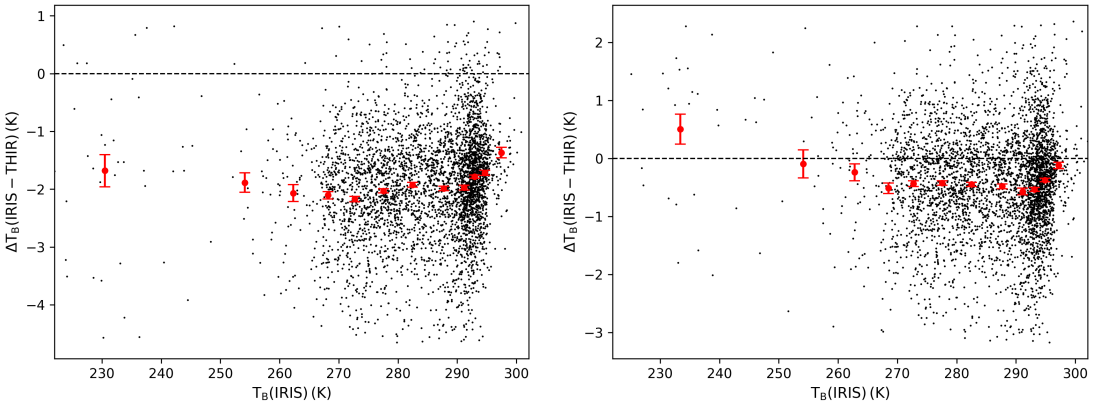



Figure 4.12: Scatter plot of IRIS minus THIR window BT versus IRIS window BT before (left) and after (right) bias-correcting IRIS. The red circles and error bars indicate the mean differences and ±1 standard error for each $T_B$ (IRIS) bin.

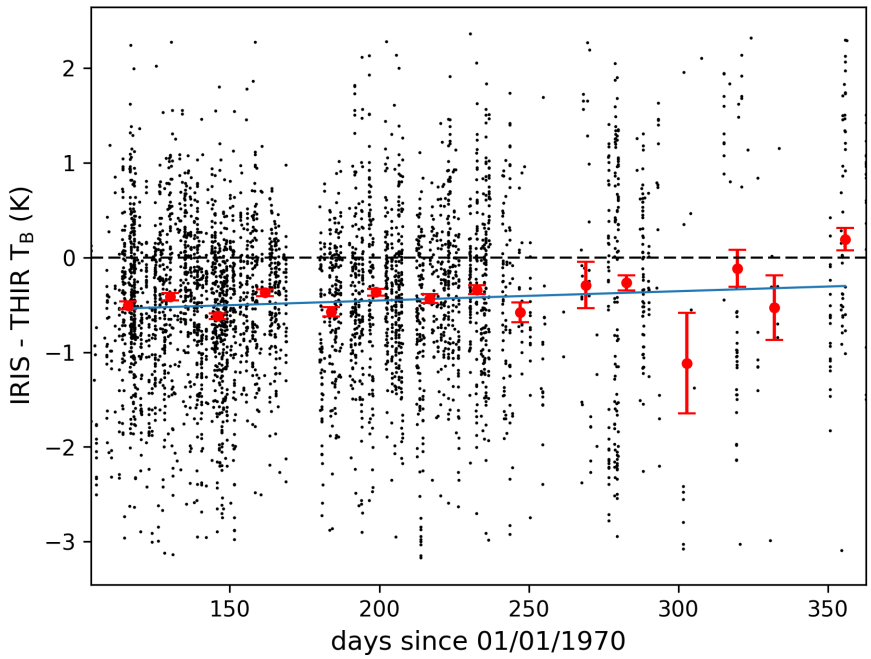



Figure 4.13: IRIS - THIR BT differences for the set of homogeneous cases over the course of 1970 after IRIS bias correction. The red circles and error bars indicate the mean differences and ±1 standard error for each bin (roughly 17 day intervals). The blue line indicates the linear regression as applied to the mean differences.

Orbital variation in the IRIS - THIR BT differences was also studied. For the set

of homogeneous cases, the orbit angles relative to the ascending node were computed. Figure 4.14 shows how the IRIS - THIR BT differences vary with the orbit angle. The

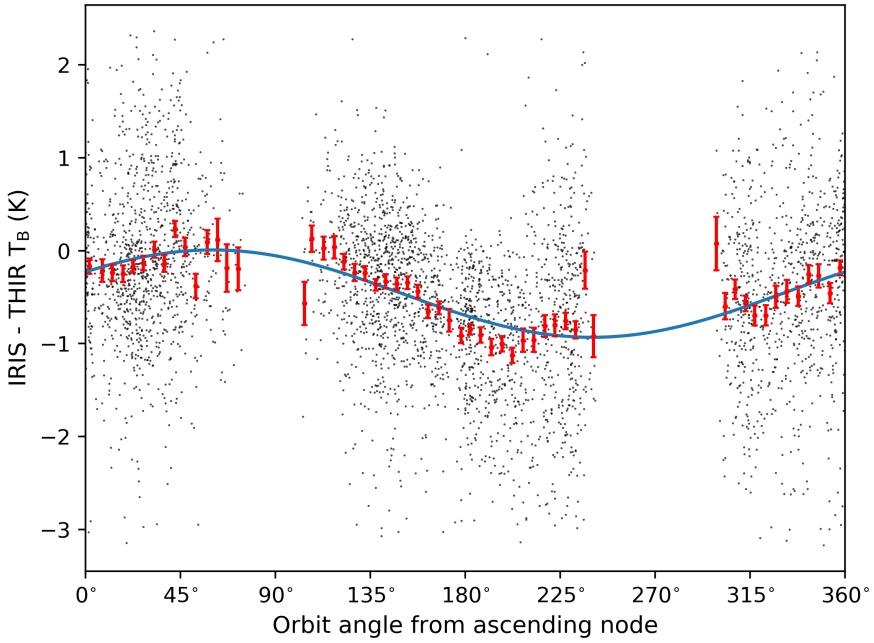



Figure 4.14: IRIS - THIR BT differences versus orbit angle relative to the ascending node, after applying the bias model to IRIS. Also plotted are the binned mean differences and standard errors (red markers) and least-squares sine-fit (blue curve).

differences can be modelled fairly well by a sine function with amplitude 0.47 K and phase 30°. The sine function was fitted to the data with the constraint that the period is 360° (1 orbit). At some orbit angles, the differences deviate from the fitted sine function. This could be the result of IRIS and THIR having separate orbital biases with different phases and amplitudes. These orbital biases may be caused by orbital changes in temperature of the instrument components. To understand if this is a cause of the orbital bias, temperatures of various components of IRIS and THIR were analysed. Figure 4.15 shows how the instrument temperatures vary with orbit angle. The THIR cell and reference temperatures appear to have an orbitally varying cycle with a phase roughly the same as in Figure 4.14. The amplitude of the cycle is greater for the cell temperature ($\approx 0.5$ K) than the reference temperatures ($\approx 0.3$ K).

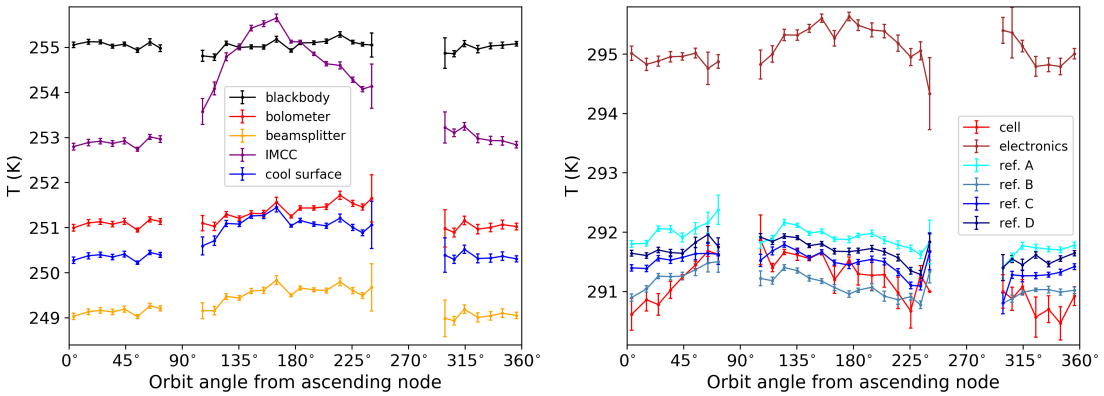



Figure 4.15: IRIS (left) and THIR (right) binned mean instrument temperatures and standard errors versus orbit angle relative to the ascending node. The IRIS blackbody temperature is shown here with 30 K subtracted from it, to aid clarity.

The pointwise correlations between THIR instrument temperatures and the IRIS-THIR BT differences vary between -0.10 and -0.32. This suggests that the variations in THIR instrument temperatures do not explain the IRIS-THIR orbital bias. Except for the blackbody, the IRIS instrument temperatures appear to have orbitally varying cycles, but these are out of phase with that shown in Figure 4.14. The correlations between IRIS instrument temperatures and the IRIS-THIR BT differences are all less than 0.1 in magnitude, so the variations in IRIS instrument temperatures do not explain the IRIS-THIR orbital bias either.

In summary, the IRIS-THIR comparisons show that there may be remaining bias in the IRIS spectra after being corrected with the preliminary bias model. This remaining bias may vary with parameters such as scene temperature and orbit angle. However, we are unable to attribute the relative IRIS-THIR bias to IRIS or THIR, so these results were not used to refine the bias model.

## 4.4 Chapter summary

This chapter has illustrated the challenges involved in re-calibrating a satellite sensor from 1970. Comparing IRIS with more recent hyperspectral sensors and exploiting

NWP data has provided information to help constrain the IRIS bias. These comparisons have shown that the bias varies spectrally with a prominent step change in the ozone absorption band. This step change appears to weaken at lower scene temperatures.

*Bantges et al.* (2016) hypothesised that the bias is due to error in the spectral emissivity applied to the on-board blackbody. This implies that the bias is constant in terms of radiance and therefore depends on scene temperature according to Planck's law. We have derived an estimate of the spectral bias from comparison of IRIS observations and NWP-based RTTOV simulations in clear-sky conditions. This comparison also revealed two important instrument effects not represented in the simulations. Empirical corrections were applied to the simulations to account for these effects. Our spectral bias estimate agrees closely with the IRIS-IMG and IRIS-IASI BT differences from *Bantges et al.* (2016) for most channels relevant to SST retrieval.

The THIR provides the only available source of data for direct validation of the bias-corrected IRIS observations. Comparison of IRIS and THIR window channel BTs showed the presence of an orbitally varying systematic effect. Temperatures of various components of IRIS and THIR were analysed in relation to this, but this did not help explain the cause of this effect. Therefore we were not able to attribute it to either IRIS or THIR.

Although we have not fully quantified the bias, our bias correction will result in improved SST estimates relative to previous studies where no attempt was made at bias correction.

# CHAPTER 5

# DEVELOPING AN SST RETRIEVAL METHOD FOR IRIS

## 5.1 Introduction

Having assessed the quality of the IRIS dataset and bias-corrected it using our best estimate of the spectral bias, we can now develop a method of retrieving SST from IRIS. Section 5.2 outlines the SST retrieval theory. Section 5.3 then explains the retrieval setup, discussing how each of the elements needed to retrieve SST are obtained and how we propose to retrieve SST from IRIS scenes with partial cloud cover. Before applying these methods to retrieve SST from IRIS, they are first tested using hyperspectral radiance data from IASI. This is presented in section 5.4, which explains how 'IRIS-like' IASI spectra can be produced and IRIS-like SSTs retrieved from them. In-situ SST measurements from drifting buoys are used to validate the IRIS-like SSTs and provide uncertainty estimates, which can be applied to the SSTs retrieved from IRIS in chapter 6.

## 5.2 Theory

The concepts of the forward and inverse problems were introduced in chapter 2. The SST and other atmospheric parameters form the state vector. The forward model $\mathbf{F}(\mathbf{x})$ describes the physics which maps the state vector $\mathbf{x}$ into the measurement vector $\mathbf{y}$, such that $\mathbf{y} = \mathbf{F}(\mathbf{x}) + \epsilon$, where $\epsilon$ is the measurement error (*Rodgers*, 2000). Prior information about the state, for example from NWP data, can be used to constrain the solution (i.e. the retrieved state). This is an important element of the SST retrieval method for IRIS. Uncertainty in the prior information can be quantified by a probability density function (PDF) over the state space, P($\mathbf{x}$). The retrieval aims to find the conditional PDF of $\mathbf{x}$ given $\mathbf{y}$, P($\mathbf{x}|\mathbf{y}$). Bayes' theorem relates the various PDFs

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})} \tag{5.1}$$

where P($\mathbf{y}|\mathbf{x}$) is the conditional PDF of $\mathbf{y}$ given $\mathbf{x}$ and P($\mathbf{y}$) is the prior PDF of $\mathbf{y}$. We try to solve equation 5.1 first by linearising the forward model about the prior state (i.e. $\mathbf{F}(\mathbf{x}) = \mathbf{K}\mathbf{x}$), and by assuming that the PDFs in equation 5.1 are Gaussian. A state from P($\mathbf{x}|\mathbf{y}$) must be chosen as the solution to the retrieval problem. It is logical to choose either the state that maximizes P($\mathbf{x}|\mathbf{y}$), known as the maximum a posteriori (MAP) solution, or the state that is the expected value of P($\mathbf{x}|\mathbf{y}$). Since we assume Gaussian PDFs, these are both the same and the solution will be termed $\hat{\mathbf{x}}$. Not all parameters used for forward model simulation are needed in the retrieval. The retrieval can be simplified by including only variables that significantly affect the SST in a reduced version of $\mathbf{x}$, which will be termed $\mathbf{z}$. Starting from equation 5.1 and using these assumptions, an expression can be derived for $\hat{\mathbf{z}}$ (*Rodgers*, 2000).

$$\hat{\mathbf{z}} = \mathbf{z}_a + \mathbf{S}_a\mathbf{K}^T(\mathbf{K}\mathbf{S}_a\mathbf{K}^T + \mathbf{S}_\epsilon)^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x}_a) \tag{5.2}$$

where $\mathbf{z}_a$ is the a priori value of the reduced state vector $\mathbf{z}$ and $\mathbf{S_a}$ is its covariance matrix. $\mathbf{Kx}_a$ is the forward model (RTTOV) simulation of the a priori value of the full state vector $\mathbf{x}_a$ and $\mathbf{S}_\epsilon$ is the covariance matrix of $\mathbf{y}$ - $\mathbf{Kx}_a$. Equation 5.2 forms the basis of the IRIS SST retrieval method as $\hat{\mathbf{z}}$ contains the retrieved (skin) SST along with other retrieved state variables. The covariance of the retrieved state is given in equation 5.3:

$$\hat{\mathbf{S}} = \mathbf{S}_a - \mathbf{S}_a\mathbf{K}^T(\mathbf{K}\mathbf{S}_a\mathbf{K}^T + \mathbf{S}_\epsilon)^{-1}\mathbf{K}\mathbf{S}_a \tag{5.3}$$

Various diagnostics can be used to inspect and analyse the retrieved state. The averaging kernel matrix ($\mathbf{A}$) describes the sensitivity of the retrieved state to the true state. It is defined as

$$\mathbf{A} = \frac{\delta\hat{\mathbf{z}}}{\delta\mathbf{z}} \tag{5.4}$$

and calculated from equation 5.5.

$$\mathbf{A} = \mathbf{S}_a\mathbf{K}^T(\mathbf{K}\mathbf{S}_a\mathbf{K}^T + \mathbf{S}_\epsilon)^{-1}\mathbf{K} \tag{5.5}$$

Values closer to one indicate that the influence of the a priori state on the retrieved state is very small. $\chi^2$ tests can be used to check if the differences between $\mathbf{z_a}$ and $\hat{\mathbf{z}}$ or $\mathbf{y}$, $\mathbf{Kx_a}$ and $\mathbf{K}\hat{\mathbf{x}}$ are consistent with the associated covariances. If it is assumed that the difference $\mathbf{d}$ between two of these quantities is Gaussian with zero mean and covariance $\mathbf{S}_d$ then

$$\chi^2 = \mathbf{d}^T{\mathbf{S}_d}^{-1}\mathbf{d} \tag{5.6}$$

$\chi^2$ tests provide a means of identifying cases where the retrieval has failed to work as expected. Analysis of the difference between these quantities (e.g. $\mathbf{K}\hat{\mathbf{x}}$ - $\mathbf{y}$) is another useful type of diagnostic.

## 5.3 Retrieval setup

This section will discuss how each of the terms in equation 5.2 are obtained and how sets of channels can be selected for retrieving SST. The success of the retrieval relies partly upon reasonably accurate prior state information to help constrain the solution and to provide a suitable linearisation point for generating $\mathbf{K}$ and $\mathbf{Kx_a}$. Cloud obscures the surface and reduces the sensitivity of the measured BT to the SST. This suggests that more accurate SSTs can be retrieved from clear-sky scenes. We therefore chose to retrieve SST from clear-sky scenes to begin with. Very few IRIS scenes can be classified as fully clear-sky, due to the footprint size. This was demonstrated when selecting a set of clear-sky observations using IDCS data (§4.2.3). This means that for the large majority of IRIS observations, the retrieval method must be adapted to retrieve SST in the presence of some cloud. This is explained in §5.3.3.

### 5.3.1 $\mathbf{z}_a$ and $\mathbf{Kx}_a$

SST, atmospheric temperature and water vapour are included in $\mathbf{z}_a$. The atmospheric profiles used to generate $\mathbf{Kx}_a$ reach an altitude of 80 km (0.01 hPa). Channels whose weighting functions peak this high in the atmosphere are not relevant to SST retrieval, so the temperature and water vapour profiles in $\mathbf{z_a}$ were restricted to heights of about 200 hPa (12 km) in clear-sky conditions and 100 hPa (16 km) in partly cloudy conditions. $\mathbf{Kx}_a$ was generated from an ensemble of ERA-20C and CERA-20C profiles as described in §4.2, including the effect of aerosols.

### 5.3.2 a priori covariances

#### 5.3.2.1 $\mathbf{S}_a$

There are a few options for estimating the covariance matrix $\mathbf{S}_a$. For each IRIS observation, the covariance of the ensemble of $\mathbf{z_a}$ obtained as described in §4.2 can be computed. Alternatively, we can specify a more simple form for $\mathbf{S}_a$, which assumes that

atmospheric temperature and humidity are each correlated over characteristic length scales. It was found that the structure of the covariance matrix obtained with the former approach is much more complex, but it is not clear how much of this structure is physical as it is calculated from a very small number of ensemble members. It was therefore decided to use the latter approach over the former. The latter approach was used by *Irion et al.* (2018) when retrieving temperature, water vapour and cloud properties from the Atmospheric Infrared Sounder (AIRS). Off-diagonal elements of $\mathbf{S}_a$ are computed using equation 5.7.

$$\sigma_{ij}^2 = \sigma_i \sigma_j \exp\left(\frac{|h_i - h_j|}{l}\right) \tag{5.7}$$

where $\sigma_{ij}^2$ is the off-diagonal covariance for profile levels $i$ and $j$, $\sigma_i$ and $\sigma_j$ are the square roots of the on-diagonal covariances, $h_i$ and $h_j$ are the altitudes of the profile levels and $l$ is the assumed length scale. *Irion et al.* (2018) assume length scales of 0.5 km for both temperature and humidity. However, the covariances derived from the former (NWP ensemble) approach suggest a greater length scale. From this, we assumed a length scale of 2 km. For the values of the covariances along the diagonal, initially these were set to the values in *Irion et al.* (2018) which are constant with altitude. In this method the covariances between temperature and humidity are not estimated (i.e. they are assumed to be uncorrelated). $\mathbf{S}_a$ also contains the a priori variance in the skin SST. This was set to a large constant value of $(4\,\mathrm{K})^2$, to ensure a high level of sensitivity of the retrieved SSTs to the observations.

#### 5.3.2.2 $\mathbf{S}_\epsilon$

There are various potential sources of error that contribute to $\mathbf{S}_\epsilon$. Radiometric noise (detector noise) contributes mostly random error to each IRIS channel. Some spectral artefacts remain after bias correction and spectral smoothing. These contribute spectrally correlated errors to $\mathbf{S}_\epsilon$. Forward model error also contributes spectrally cor-

related errors to $\mathbf{S}_\epsilon$. The contributions of each error source to $\mathbf{S}_\epsilon$ will be explained in turn.

As described in §4.2.4.2, the IRIS SRF was corrected by widening it by a factor ($f$), determined by finding the factor that best reduces spectral artefacts in the bias estimate. This factor was found to vary spectrally from 1.20 at $720\,\mathrm{cm}^{-1}$ to 1.32 at $1210\,\mathrm{cm}^{-1}$. As an estimate of the uncertainty in $f$, an ensemble of spectral bias estimates was computed by incrementing $f$ by 0.01 from $f$ - 0.05 to $f + 0.05$. This ensemble was calculated for each IRIS spectrum of interest (cases for which SST will be retrieved) and from it, the error covariance ($\mathbf{S}_{\mathrm{SRF}}$) was calculated.

Part of the uncertainty in the bias estimate is derived from uncertainty in the selection of IRIS observations used to derive the bias (§4.2.3). Three possible sets of observations are obtained depending if the normal distribution in Figure 4.5 is centred at 0.9 K, 1.0 K or 1.1 K. The spectral radiance bias was calculated for each set of observations. For each radiance bias estimate, the spectral BT bias was calculated for each IRIS spectrum of interest and the error covariance ($\mathbf{S}_{\mathrm{fit}}$) was calculated from the three BT bias spectra. There is additional uncertainty due to error from estimating the bias from the median of a finite set of IRIS observations. Assuming a large sample and normal distribution, the standard error in the median is equal to the standard error in the mean multiplied by a factor of about 1.25 (*Evans*, 1942). The standard error is spectrally correlated, so the associated error covariance matrix ($\mathbf{S}_{\mathrm{sampling}}$) was determined empirically. 1000 random samples of 10 IRIS spectra were selected from the set of IRIS spectra used to determine the bias. 1000 bias estimates were computed from the median of each sample spectra. The covariance of these bias estimates was computed and converted to a correlation matrix. Assuming that the correlations are independent of the sample size, this correlation matrix was used to compute $\mathbf{S}_{\mathrm{sampling}}$.

An estimate of the average spectral radiometric noise was derived from Figure 38 of *Poli and Brunel* (2016). This was converted to brightness temperature for each observation. It is likely that the noise is correlated between nearby channels, due to

channel overlap. A correlation length scale of 2 channels was assumed when computing the error covariance $\mathbf{S}_{\text{noise}}$ due to radiometric noise.

IASI data can be used to estimate the covariance due to forward model error $\mathbf{S}_{\text{model}}$. Forward model error arises because no RTM is able to fully model the complex physics involved in simulating the ToA radiance. It also incorporates error in the specification of trace gases such as $CO_2$ and $O_3$. It was estimated by taking the mean difference of a set of clear-sky observed and simulated IASI spectra, degraded to the IRIS spectral resolution. The mean difference spectrum was split up into three spectral regions ($CO_2$ wing, long wavelength and short wavelength windows). The standard deviation of channel BTs for each spectral region was used to construct the diagonal of $\mathbf{S}_{\text{model}}$. For the off-diagonal terms, a correlation channel length scale was computed for each spectral region. Error due to potential cloud contamination was also simulated, by producing a set of simulations for different maximum cloud fractions (CF), between 0 and 1.5% in intervals of 0.5%. The covariance of the mean difference spectra for each value of allowed maximum CF was taken to produce the error covariance matrix due to cloud contamination, $\mathbf{S}_{\text{cloud}}$.

The final source of uncertainty contributing to $\mathbf{S}_{\epsilon}$ is due to aerosols. Due to lack of confidence in the exact magnitude of the effect of aerosols on the simulated BTs, the uncertainty in the spectral differences were assumed to be equal to half of the differences. To produce the associated error covariance matrix $\mathbf{S}_{\text{aerosol}}$, inter-channel correlations were assumed to be one between all channels.

To produce $\mathbf{S}_{\epsilon}$, all of the aforementioned error covariance matrices are summed, as we assume that the sources of error are independent. Additionally, $0.01\,\text{K}^2$ is added to each element of $\mathbf{S}_{\epsilon}$, to take into account additional uncertainty due to potential fluctuations in the magnitude of the spectral bias. Figure 5.1 shows each of the component of $\mathbf{S}_{\epsilon}$ for a typical IRIS spectrum. Despite many of the sources of uncertainty contributing to $\mathbf{S}_{\epsilon}$ being highly spectrally correlated, it is highly diagonal, with the main contribution coming from $\mathbf{S}_{\text{noise}}$.

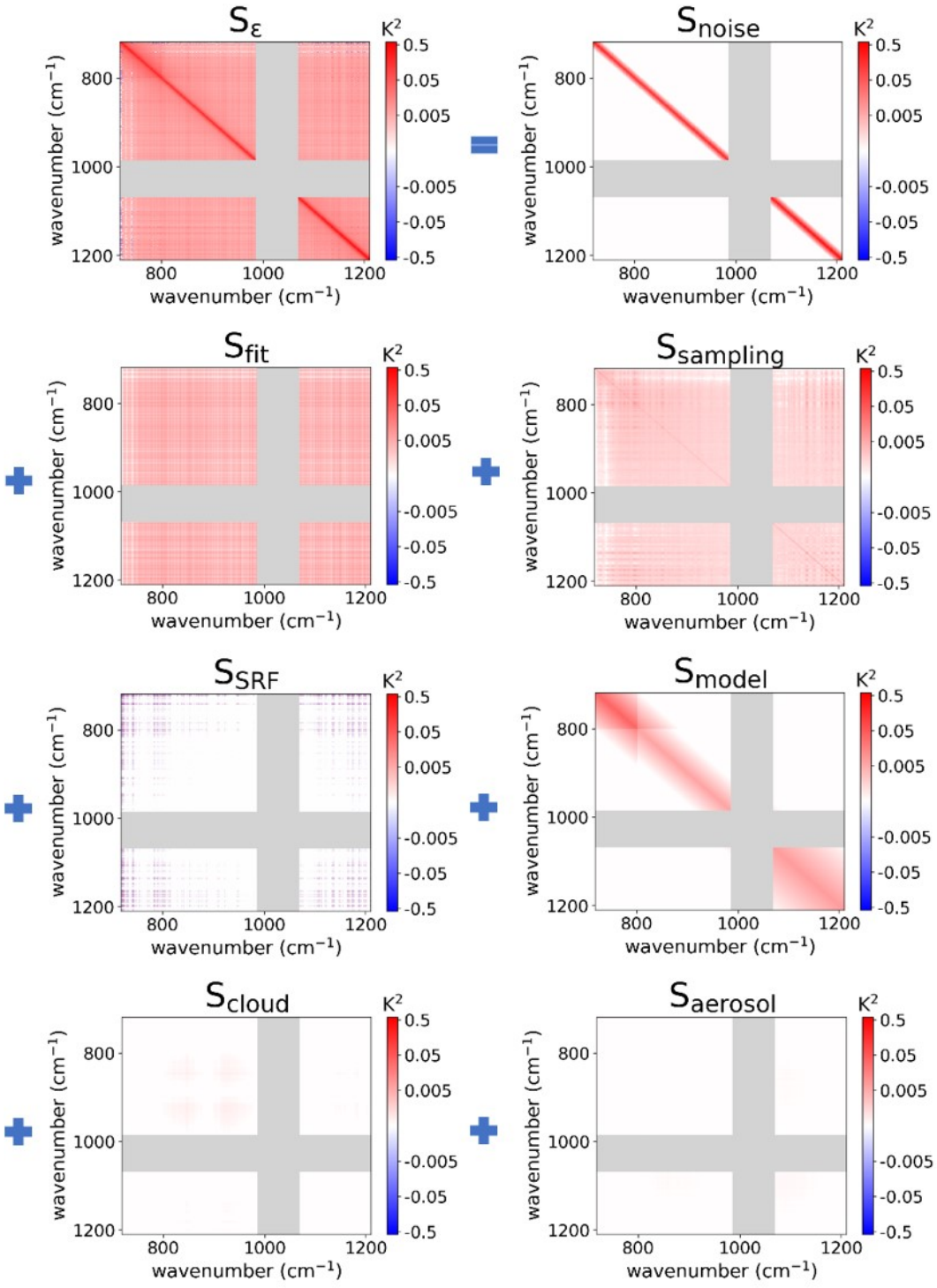



Figure 5.1: Components contributing to the total measurement error covariance matrix $\mathbf{S}_\epsilon$, calculated for a typical IRIS spectrum. Note that an additional $0.01\,\mathrm{K}^2$ is added to each element of $\mathbf{S}_\epsilon$ to account for potential fluctuations in the magnitude of the spectral bias. Channels in the ozone absorption region are not used in the retrieval, so are shaded grey.

### 5.3.3 Retrieving SST from partly cloudy scenes

The SST retrieval method must be adapted in order to retrieve SST for observations affected by cloud. This involves adding variables describing cloud properties to $\mathbf{z}$. A priori estimates of total cloud fraction and cloud height were obtained from ERA-20C and CERA-20C in the same way as the other state variables. Initially, just the CF was allowed to vary in the retrieval, with the cloud height fixed at the a priori value. This means that only CF was added to $\mathbf{z}_a$. An estimate of the a priori uncertainty in the CF was made by computing the standard deviation of the differences between the a priori CF and the 'observed' CF from the IASI cloud mask (*August et al.*, 2012) for a set of IRIS-like observations during 2011. As will be shown in §5.4.2.2, errors in the a priori CF are typically large. The retrieval algorithm can be iterated to help ensure that the CF converges to its true value. As the retrieved SST is highly sensitive to the retrieved CF, this ensures that a more accurate SST is retrieved. Equation 5.8 shows how the retrieved state for the (i+1)$^{\text{th}}$ iteration is obtained using the retrieved state from the previous iteration and the a priori state.

$$\mathbf{z}_{i+1} = \mathbf{z}_a + \mathbf{S}_a\mathbf{K}_i^T(\mathbf{K}_i\mathbf{S}_a\mathbf{K}_i^T + \mathbf{S}_\epsilon)^{-1}(\mathbf{y} - \mathbf{K}_i\mathbf{x}_i + \mathbf{K}_i(\mathbf{z}_i - \mathbf{z}_a)) \tag{5.8}$$

The iteration is started with $\mathbf{z}_0 = \mathbf{z}_a$. The retrieved CF must be kept within the range 0-1. To ensure this, a logit transformation was performed:

$$\text{logit}(\text{CF}) = \log\left(\frac{\text{CF}}{1 - \text{CF}}\right) \tag{5.9}$$

which maps the CF from $[0, 1]$ to [-$\infty$, +$\infty$]. To generate cloud-affected RTTOV simulations, a simple cloud scheme was used which assumes black, opaque clouds at a single level. The CF and cloud top pressure corresponding to this level are specified in $\mathbf{x}_a$. The retrieval method for partly cloudy scenes is tested on IASI data in §5.4.2.2 and §5.4.3.2.

#### 5.3.3.1 Cloud detection

The higher the CF, the more difficult it is to retrieve the SST. This is largely due to reduced sensitivity of the measured BT to changes in the SST. For typical mid-level cloud at an altitude of 3 km, an error in the retrieved CF of 0.01 (1 %) is expected to result in an error in the retrieved SST of the order of 0.2 K. The absolute errors in the a priori CF increase as the CF itself increases. For these reasons, it is important to restrict SST retrieval to IRIS cases with a relatively low CF. A threshold-based approach was used to select IRIS observations with an expected CF of 20 % or below. Most cloud over the ocean occurs at an altitude below 3 km (700 hPa) (*Wu et al.*, 2011). A threshold for minimum window channel BT was calculated from equation 5.10.

$$T_{\text{thresh}} = 0.8\,SST + 0.2\,T_{700\,\text{hPa}} + c_a \tag{5.10}$$

This represents the approximate window BT for a scene with 20 % CF and a cloud top pressure of 700 hPa. The constant $c_a$ is an offset due to atmospheric attenuation and is estimated empirically from IASI data (§5.4.2.2). For any IRIS observations with window BT $> T_{\text{thresh}}$, the CF is likely to be less than 20 %. The mean BT for the ten most transparent channels was used to represent the window BT. IASI data and 'observed' CF values are used to test the effectiveness of this method in §5.4.2.2.

### 5.3.4 Channel selection

A subset of IRIS channels must be chosen for SST retrieval. In principle, if the forward model is unbiased, the uncertainty in the retrieved SST is minimised by using all IRIS channels. However, small biases in the forward model are expected, which can accumulate the more channels are used. Additionally it is more computationally efficient to use a smaller set of channels. Ideally, the chosen set of channels minimises the retrieval uncertainty. Different sets of channels are selected for retrieving SST in clear-sky conditions and partial cloud.

#### 5.3.4.1 Clear-sky

Generally speaking, clear-sky retrieval of SST requires only window channels. However, we can allow channels to be selected outside of the window regions if it results in lower retrieval uncertainty. The method of choosing the optimum set of channels is described by the following steps:

1. Select a suitable starting channel
2. Carry out retrieval with this channel and a second channel and compute the uncertainty in the retrieved SST ($\sigma_{\mathrm{S\hat{S}T}}$)
3. Repeat for different choices of second channel
4. Select the channel that minimises $\sigma_{\mathrm{S\hat{S}T}}$
5. Repeat steps 2-4 to find the third channel that minimises $\sigma_{\mathrm{S\hat{S}T}}$
6. Repeat steps 2-5 until a suitable number of channels have been chosen

$\sigma_{\mathrm{S\hat{S}T}}$ is obtained from $\hat{\mathbf{S}}$, calculated from equation 5.3. A highly transparent window channel at $960\,\mathrm{cm}^{-1}$ is used as the starting channel. The set of channels which we choose from in step 3 are all of the channels in the wavenumber intervals 718-985 and 1070-1210 $\mathrm{cm}^{-1}$. This includes channels in one wing of the $CO_2$ absorption band, and window channels either side of the $O_3$ absorption band. We have found that the IRIS spectral resolution varies from about $3.4\,\mathrm{cm}^{-1}$ to $3.7\,\mathrm{cm}^{-1}$ (§4.2.4.2). Since the channel separation is about $1.4\,\mathrm{cm}^{-1}$, there is considerable overlap between neighbouring channels and channels separated by one or two. For this reason we apply the constraint that channels must be chosen such that the minimum channel separation in the chosen set of channels is three.

Figure 5.2 shows the central wavenumbers of the first 20 channels selected using this method, based on the set of clear-sky IRIS observations. Roughly equal numbers of channels are selected in each of the window regions either side of the $O_3$ absorption

band. Some of these channels are highly transparent and some are more affected by water vapour. Only one channel is in the $CO_2$ wing. This is unsurprising as the $CO_2$ channels have weighting functions that peak high in the atmosphere and so have very low sensitivity to SST. Figure 5.3 shows that the median uncertainty in the retrieved

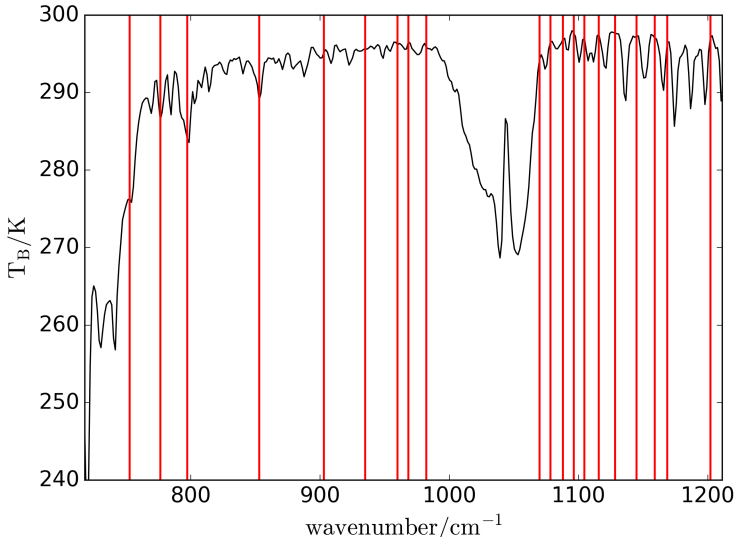



Figure 5.2: Central wavenumbers of the first 20 selected clear-sky IRIS channels, superimposed on a typical IRIS spectrum.

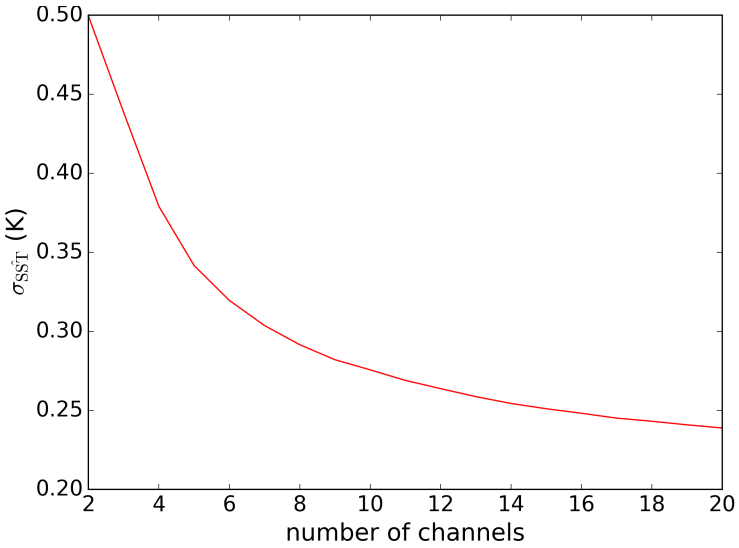



Figure 5.3: The median uncertainty in retrieved SST from the set of clear-sky IRIS cases, as a function of the number of channels used in the retrieval.

SST $\sigma_{\hat{\mathrm{SST}}}$ reduces as more channels are used in the retrieval. It reduces more slowly as more channels are used, suggesting that 20 channels is a suitable number for retrieving SST. In reality, $\sigma_{\hat{\mathrm{SST}}}$ does not continue to reduce due to forward model biases.

#### 5.3.4.2 Partial cloud

A set of channels was selected that minimises the uncertainty in the retrieved CF ($\sigma_{\hat{\mathrm{CF}}}$) as well as $\sigma_{\hat{\mathrm{SST}}}$. These channels were selected based on a set of partly cloudy IRIS-like IASI observations. The process of obtaining IRIS-like IASI observations is explained in §5.4.1. One approach is to select channels that minimise the weighted sum of the median $\sigma_{\hat{\mathrm{CF}}}$ ($\sigma_{\tilde{\mathrm{CF}}}$) and median $\sigma_{\hat{\mathrm{SST}}}$ ($\sigma_{\tilde{\mathrm{SST}}}$), $\sigma_{\tilde{\mathrm{tot}}}$.

$$\sigma_{\tilde{\mathrm{tot}}} = \sqrt{(\mathrm{w}\sigma_{\tilde{\mathrm{SST}}})^2 + \sigma^2_{\tilde{\mathrm{CF}}}} \quad \text{where w} = \frac{\bar{\sigma_{\mathrm{CF}}}}{\bar{\sigma_{\mathrm{SST}}}} \tag{5.11}$$

and $\bar{\sigma_{\mathrm{CF}}}$ ($\bar{\sigma_{\mathrm{SST}}}$) is the mean of $\sigma_{\hat{\mathrm{CF}}}$ ($\sigma_{\hat{\mathrm{SST}}}$) for the set of IRIS-like IASI cases. $\sigma_{\hat{\mathrm{SST}}}$ was weighted to ensure its mean value is equal to the mean of $\sigma_{\hat{\mathrm{CF}}}$. An alternative approach is to select two sets of ten channels that separately minimise $\sigma_{\tilde{\mathrm{CF}}}$ and $\sigma_{\tilde{\mathrm{SST}}}$. Figure 5.4 compares the wavenumbers of the channels selected using each of these approaches.

Only one of the 'simultaneous' channels and two of the 'separate' channels are in the $CO_2$ wing. The remaining channels are in the window regions, with varying amounts of water vapour absorption. $\sigma_{\tilde{\mathrm{SST}}}$ is 2.65 K for the 'simultaneous' channel set and 2.64 K for the 'separate' channel set. The values of $\sigma_{\tilde{\mathrm{CF}}}$ are 0.114 and 0.126 respectively. This suggests that there may be a small advantage in using the 'simultaneous' channel set if equal weight is given to the SST and CF. The effect of the choice of channel set on the retrieved SST and CF will be tested in §5.4.

## 5.4 Testing with IASI

Given the absence of validation data for the IRIS period, some evidence that SST can successfully be retrieved from IRIS is required. If modern hyperspectral measurements

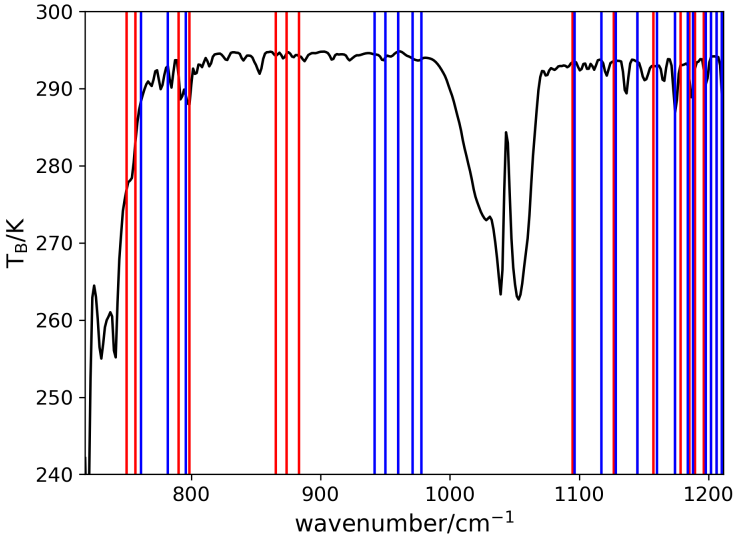



Figure 5.4: Central wavenumbers of the first 20 selected IRIS channels that simultaneously (blue) and separately (red) minimise the uncertainty in the retrieved SST and CF, superimposed on a typical IRIS-like spectrum.

from IASI are used to simulate IRIS-like 'pseudo-observations', then the SSTs retrieved using the IRIS retrieval method can be validated against high quality in-situ SST data from drifting buoys. This should give an indication of the expected IRIS performance. IASI has much higher spatial and spectral resolution than IRIS, so the IASI spectra must be degraded and spatially averaged to match the spatial and spectral resolution of IRIS. The CF values provided for each IASI pixel can be used to select IASI cases with different amounts of cloud for testing the retrieval of SST in clear-sky and partly cloudy conditions.

### 5.4.1 Generating IRIS-like observations

One year of IASI L1c radiance data was obtained (*EUMETSAT*, 2011), choosing the year 2011 as it had the same ENSO phase as 1970 (moderate La Niña). IASI produces 30 fields of regard (FOR) per scanline. Each FOR is an array of 2 x 2 circular pixels with diameter 12 km. To approximate the IRIS spatial resolution, 16 nadir pixels were averaged to produce a single IRIS-like pixel (100 x 100 km). The 16-pixel average

spectrum was then smoothed to match the IRIS spectral resolution. Since IASI has a much narrower SRF than IRIS (FWHM = 0.5 $\mathrm{cm}^{-1}$), the approximation was made that each IASI channel represents a single wavelength at the channel centre. Equation 5.12 describes how IRIS-like radiances are computed from IASI radiances.

$$L_{\text{IRIS-like}} = \frac{\sum_i L_i SRF_i}{\sum_i SRF_i} \tag{5.12}$$

where $L_i$ is the IASI radiance for the $i^{\text{th}}$ IASI channel and $SRF_i$ is the corrected IRIS SRF interpolated to the $i^{\text{th}}$ IASI channel wavenumber. As with the original IRIS data, it was necessary to remove IRIS-like observations over land and sea-ice. The land removal method is as described in §3.3. For sea-ice removal, the same dataset was used as for IRIS, using the daily values of sea-ice concentration during 2011. After this process, about 900,000 IRIS-like IASI spectra remained.

#### 5.4.1.1 Simulating IRIS-like errors

To make these observations more IRIS-like, we can simulate IRIS-like errors, which can be added to the observations. IRIS-like $\mathbf{S}_\epsilon$ were generated as described in §5.3.3.2. From $\mathbf{S}_\epsilon$, the corresponding errors were simulated using a Cholesky decomposition. First, $\mathbf{S}_\epsilon$ were converted to the corresponding correlation matrices $\mathbf{R}_\epsilon$. The elements $R_{ij}$ of $\mathbf{R}_\epsilon$ are calculated from the elements $S_{ij}$ of $\mathbf{S}_\epsilon$ using equation 5.13:

$$R_{ij} = \frac{S_{ij}}{\sqrt{S_{ii} S_{jj}}} \tag{5.13}$$

The Cholesky decomposition of $\mathbf{R}_\epsilon$ is the lower triangular matrix $\mathbf{L}$, such that $\mathbf{L}^{\mathrm{T}}\mathbf{L} = \mathbf{R}_\epsilon$. The matrix of correlated errors $\mathbf{E}_{\text{cor}} = \mathbf{E}_{\text{uncor}}\mathbf{L}$, where $\mathbf{E}_{\text{uncor}}$ represents a matrix of uncorrelated errors. The uncorrelated errors are simulated by selecting random samples from the normal distribution N(0, $S_{ii}$).

### 5.4.2 Retrieving IRIS-like SSTs

#### 5.4.2.1 Clear-sky

A subset of IRIS-like IASI spectra with CF $\leq$ 2% (derived from the IASI cloud mask) were selected to test the 'clear-sky' retrieval method. As the aim of retrieving IRIS-like SSTs is to validate them using drifting buoy data, only cases collocated to within 50 km and 3 hours of quality controlled drifting buoy SSTs were selected to test the retrieval. This is further discussed in §5.4.3. To produce $\mathbf{x}_a$, ERA-Interim was used because ERA-20C and CERA-20C do not cover the year 2011. ERA-Interim has 60 model levels compared to 91 for (C)ERA-20C and has slightly higher spatial resolution at 0.75 degrees. Otherwise, the method of obtaining the state vectors is as described in §4.2.1 for ERA-20C. RTTOV was used to simulate the clear-sky IRIS-like spectra, as well as the corresponding Jacobian matrices. SSTs were retrieved from these cases using the same 20 IRIS channels shown in Figure 5.2. The distribution of retrieved minus a priori skin SST is shown in Figure 5.5. There is a negative median difference

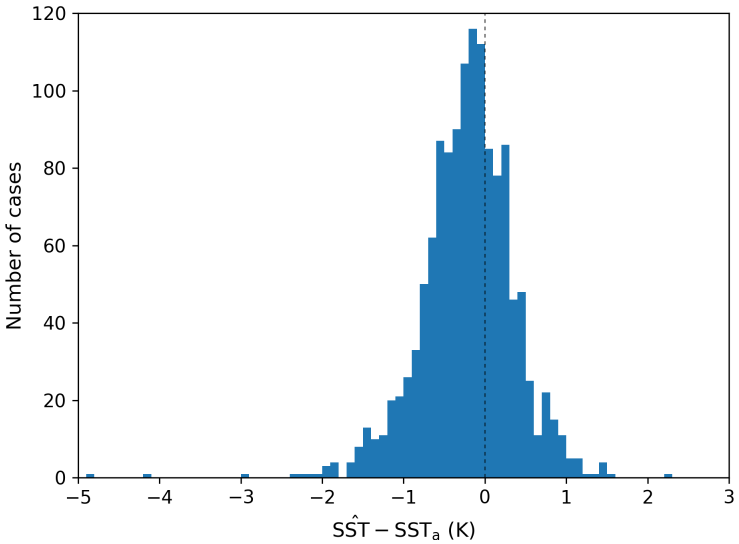



Figure 5.5: Distribution of retrieved minus a priori skin SST for the set of clear-sky ($\leq$ 2% CF) IRIS-like IASI cases.

of -0.21 K. This may be due to the expected small amount of cloud contamination in

the observed spectra. The averaging kernel of the retrieved SST is very close to 1 for all cases (mean of 0.992). This indicates that the influence of the a priori SST on the retrieved SST is very small. Figure 5.6 shows the difference between the retrieved and a priori temperature and water vapour (q) profiles. The retrieved temperature is on average up to 0.35 K warmer than the a priori temperature at 600-650 hPa. There is very little difference on average between the retrieved and a priori humidity.

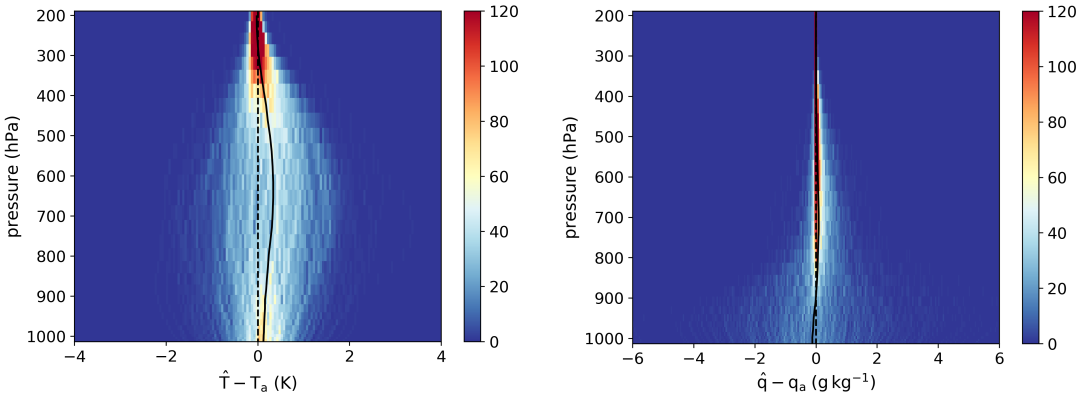



Figure 5.6: 2D histogram of retrieved minus a priori temperature (left) and water vapour (q) (right) profiles for the set of clear-sky IRIS-like IASI retrievals. The median is indicated by the solid black line.

Simulated IRIS-like spectra $\mathbf{K}\hat{\mathbf{x}}$ can be generated from $\hat{\mathbf{x}}$, produced by combining the retrieved state vector $\hat{\mathbf{z}}$ with the a priori values of the remaining parameters in $\mathbf{x}_a$ which are not retrieved. $\mathbf{K}\hat{\mathbf{x}}$ can then be compared with the observed spectra $\mathbf{y}$. If the simulated and observed spectra agree to the degree expected from their uncertainties, this indicates that the retrieved state is consistent with the observations and so the retrieval method is working successfully. Figure 5.7 shows the differences between $\mathbf{K}\hat{\mathbf{x}}$ and $\mathbf{y}$. For most channels, the median difference is below 0.2 K. The difference exceeds 0.3 K for some channels in the $CO_2$ wing, which may be due to errors in the retrieved temperature at altitudes where the weighting functions for these channels peak. Figure 5.8 shows that the distribution of $\chi^2(\mathbf{K}\hat{\mathbf{x}} - \mathbf{y})$ is similar to the expected distribution. This indicates that the uncertainty estimates used to construct the a priori covariance matrices ($\mathbf{S}_a$ and $\mathbf{S}_\epsilon$) are realistic. A threshold for $\chi^2$ was set at the first

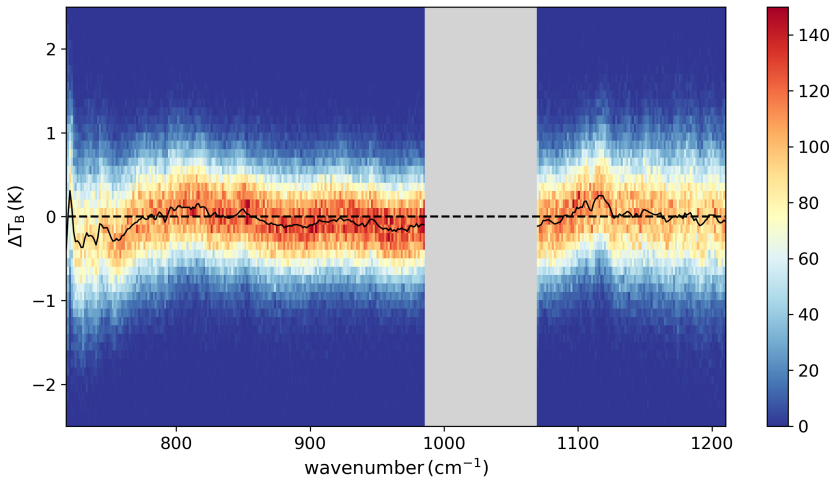



Figure 5.7: 2D histogram of $\mathbf{K}\hat{\mathbf{x}} - \mathbf{y}$ for the set of clear-sky IRIS-like IASI retrievals. The median is indicated by the solid black line. The ozone absorption region is not relevant for SST retrieval, so is shaded grey.

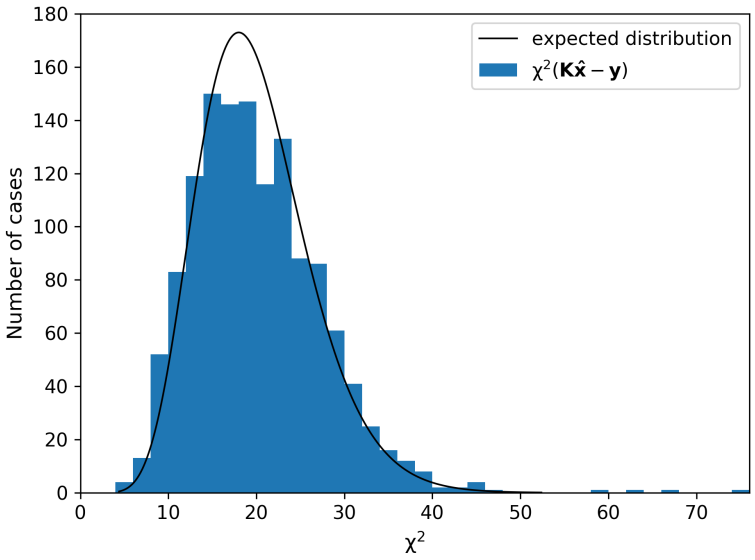



Figure 5.8: $\chi^2(\mathbf{K}\hat{\mathbf{x}} - \mathbf{y})$ for the set of clear-sky IRIS-like IASI retrievals. The black line indicates the expected distribution of $\chi^2$.

bin where the observed number of cases is at least twice the expected number of cases. Cases with $\chi^2$ values greater than this threshold were removed from the set of retrieved SSTs before validation. This test failed to remove some cases where the retrieved - a priori SST is many standard deviations from the median, seen in Figure 5.5. The true

SST uncertainty varies spatially and temporally, so some cases are more likely to be several sigma from the median than others. To partially allow for this, a relatively tolerant threshold, given the sample size, of ±4-sigma from the median was used for removal of outliers.

#### 5.4.2.2 Partial cloud

The method used to obtain a set of IRIS observations with an expected CF of 20 % or below was tested on the set of all-sky IRIS-like IASI observations. A priori values of total CF were taken from ERA-Interim. The constant $c_a$ in equation 5.10 was varied in order to determine a suitable threshold for the minimum window channel BT. Figure 5.9 shows the cumulative distribution of 'observed' CF derived from the IASI cloud mask for four different values of $c_a$, compared to the cumulative distribution for all cases. A

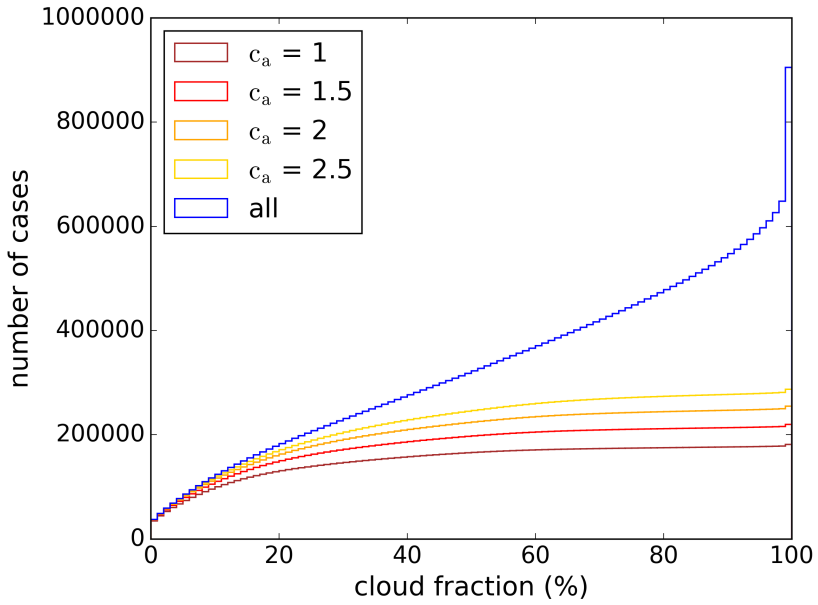



Figure 5.9: Cumulative distribution of 'observed' CF for different values of $c_a$ and for all cases of the set of all-sky IRIS-like IASI observations. The 'observed' CFs are derived from the IASI cloud mask.

higher value of $c_a$ results in a higher proportion of cases with CF $< 20\,\%$ being retained, but also results in a higher proportion of unwanted cases with CF $> 20\,\%$. A value of

2 K was chosen to make a compromise between these two factors. The BT threshold was applied to the set of all-sky IRIS-like IASI observations to obtain a set of cases with CF mostly below 20 %. Figure 5.10 compares the 'observed' (IASI) CFs to the a priori (ERA-Interim) CFs for this set of cases. There is a tendency for ERA-Interim

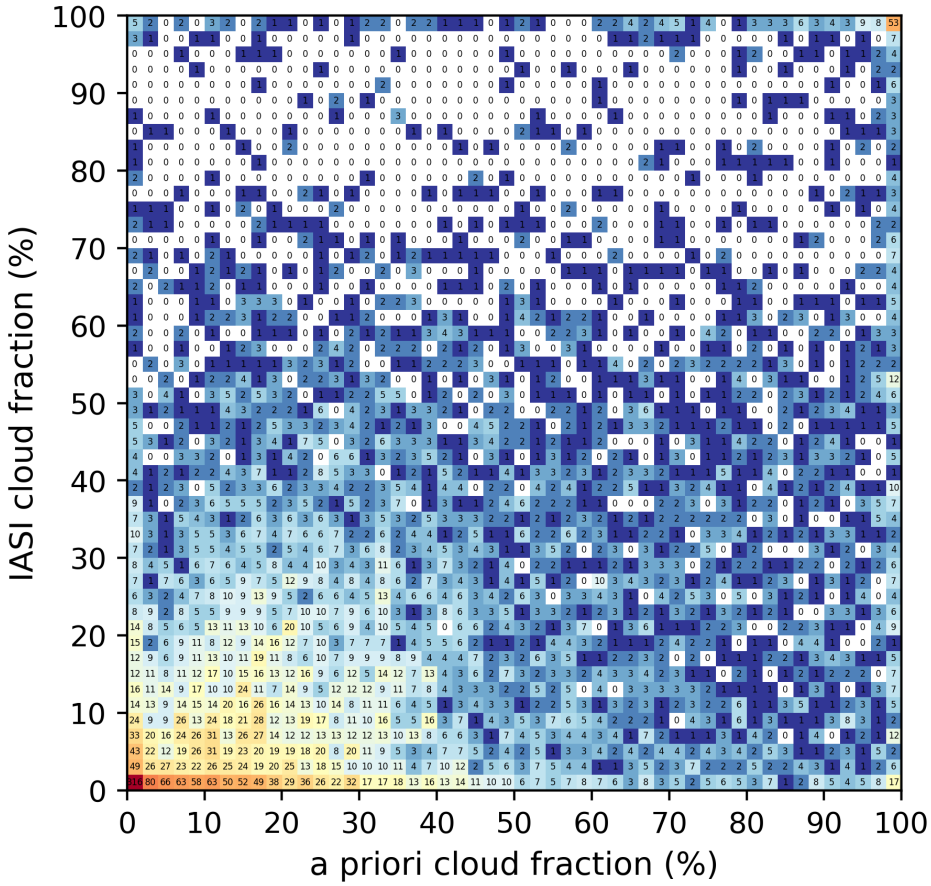



Figure 5.10: 2D histogram of IASI CF vs ERA-Interim CF corresponding to the set of IRIS-like IASI cases with $\leq$ 20% expected CF. Values in bins indicate frequencies.

to overestimate the CF relative to the IASI CF; the median CF is 25% compared to 15% for the IASI CFs. The Pearson correlation coefficient ($r$) is 0.41. The a priori CF uncertainty was calculated from the standard deviation of the errors in the a priori CFs relative to the IASI CFs, where the CFs are expressed as logit(CF).

Using the set of 'simultaneous' channels from §5.3.4.2, SST and CF were retrieved for the first iteration of equation 5.8 for this set of spectra. Figure 5.11 shows the

distribution of retrieved minus a priori SST. The median difference is -0.26 K. There

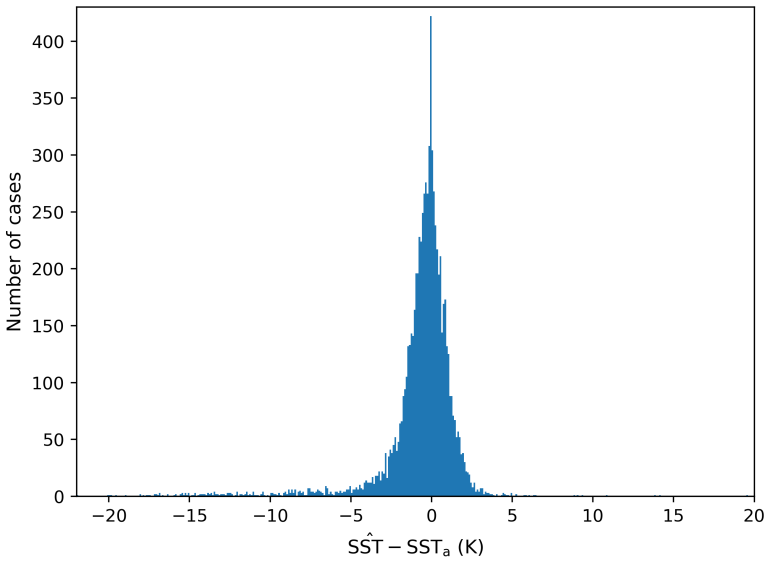



Figure 5.11: Distribution of retrieved minus a priori skin SST after one iteration for the set of IRIS-like IASI retrievals with $\leq 20\%$ expected CF.

is a visible cold tail in the distribution, potentially due to biases in the retrieved CF. Figure 5.12 shows the histogram of retrieved versus 'observed' CF. $r = 0.54$, indicating better agreement between the retrieved and 'observed' CF, relative to the a priori CF. The retrieved CFs tend to slightly underestimate the CF, in contrast to the a priori CFs. The sign of the bias in the retrieved CFs is consistent with the cold tail in Figure 5.11, since the two factors compensate for one another. Figure 5.13 shows the difference between the retrieved and a priori T and $\ln q$ profiles. The retrieved temperature is on average up to 0.2 K lower than the a priori temperature, whilst there is very little difference between the average retrieved and a priori water vapour.

The bias in the retrieved CFs and cold tail in Figure 5.11 are evidence that the retrieval has not yet converged after just one iteration. The retrieval was iterated twice more and this was also repeated for the set of 'separate' channels to check the sensitivity of the retrieval to the choice of channels. Figure 5.14 shows that iterating the retrieval twice more removes the cold tail seen in Figure 5.11. The choice of channel set has little effect on the distribution of retrieved minus a priori SST, although the

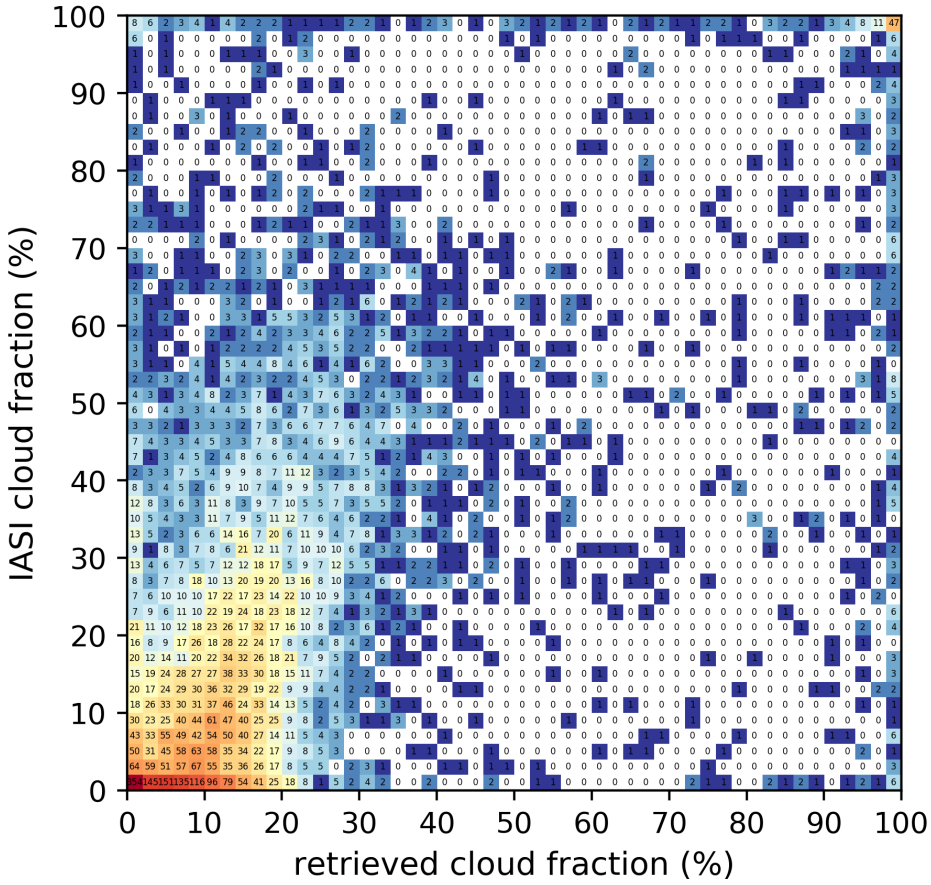



Figure 5.12: 2D histogram of IASI CF vs retrieved CF after one iteration for the set of IRIS-like IASI retrievals with $\leq 20\%$ expected CF. Values in bins indicate frequencies.

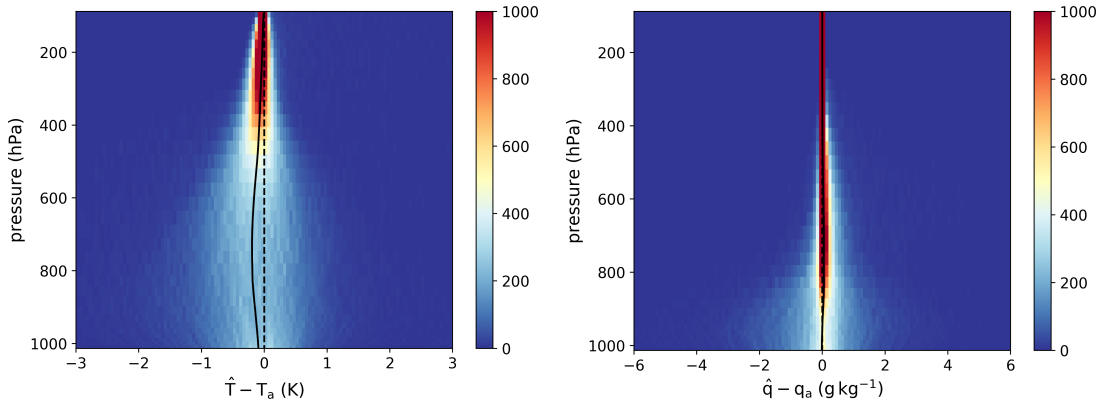



Figure 5.13: 2D histogram of retrieved minus a priori T (left) and ln q (right) profiles for the set of IRIS-like IASI retrievals with $\leq 20\%$ expected CF. The median is indicated by the black line.

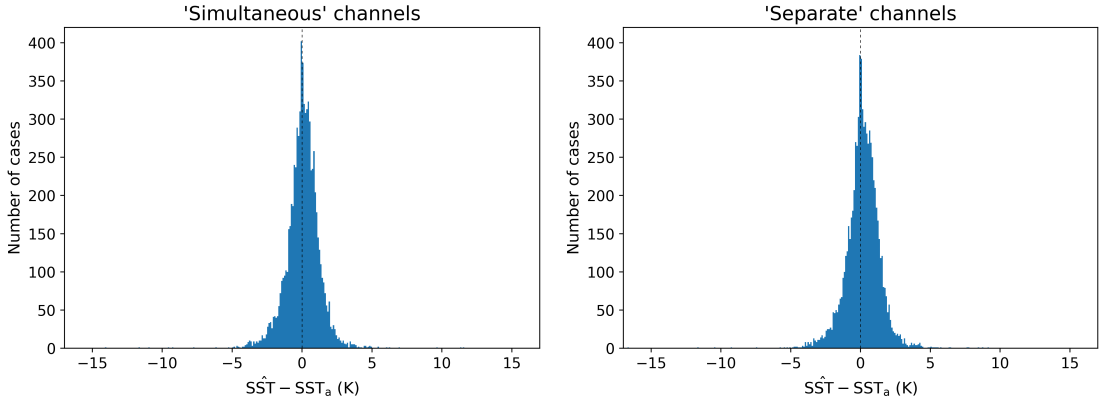



Figure 5.14: Distribution of retrieved minus a priori SST after 3 iterations for the set of of IRIS-like IASI retrievals with $\leq 20\%$ expected CF using the sets of 'simultaneous' and 'separate' channels.

SSTs retrieved using the 'separate' channels are about 0.1 K higher on average than using the 'simultaneous' channels. Figure 5.15 compares the distributions of retrieved minus 'observed' CF for each channel set.

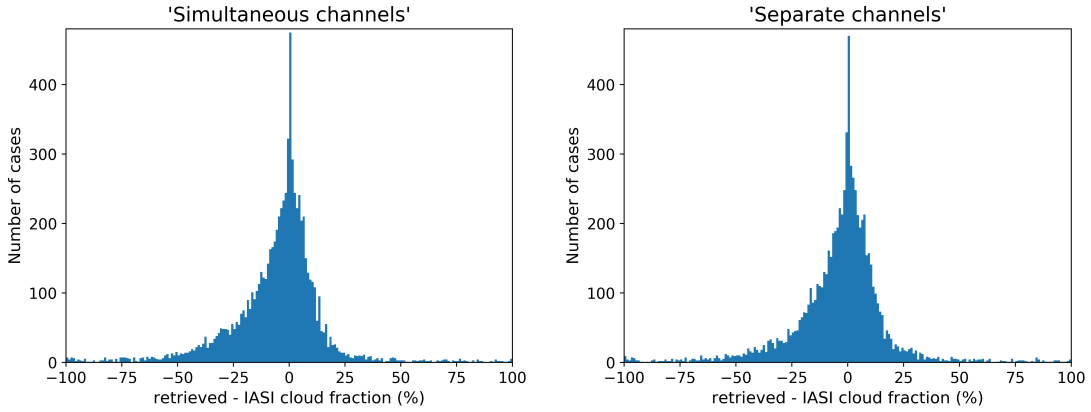



Figure 5.15: Distribution of retrieved minus IASI CF after 3 iterations for the set of IRIS-like IASI retrievals with $\leq 20\%$ expected CF using the sets of 'simultaneous' and 'separate' channels.

The distribution is more negatively skewed for the 'simultaneous' channels, although the standard deviations are about the same. These results do not give a clear indication of which set of channels is more suitable for SST retrieval from IRIS, so it was decided that based on the slightly lower value of $\sigma_{\tilde{\mathrm{CF}}}$ found in §5.3.4.2, the set of 'simultaneous' channels should be used to retrieve SST from IRIS.

As with the clear-sky IRIS-like IASI SSTs, a chi-squared test on $\mathbf{K}\hat{\mathbf{x}}-\mathbf{y}$ was used to remove cases where the retrieval failed to perform as expected. Unlike for the clear-sky retrievals, the averaging kernel of the SST is not always close to one due to the effects of including CF in the retrieval. As a final diagnostic, Figure 5.16 compares the SST and CF averaging kernels for this set of cases. For most cases, $\mathbf{A}$(SST) and $\mathbf{A}$(CF) are

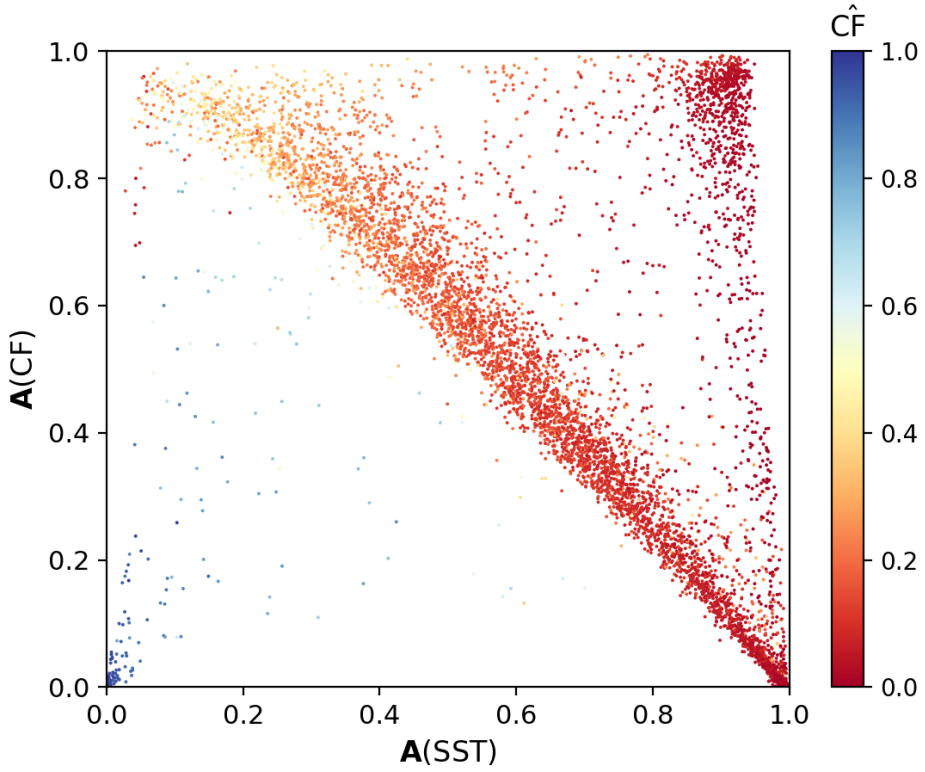



Figure 5.16: Comparison of SST and CF averaging kernels for the set of IRIS-like IASI retrievals with $\leq$ 20% expected CF. Data points are coloured according to retrieved CF.

approximately inversely proportional, because for cases with higher retrieved CF, the retrieval generally becomes more sensitive to CF and less sensitive to SST. There are a set of cases with very high $\mathbf{A}$(SST) and $\mathbf{A}$(CF) with retrieved CF close to zero. A small set of cases can also be seen with near zero $\mathbf{A}$(SST) and $\mathbf{A}$(CF) with retrieved CF close to one. It might be expected that more accurate SSTs are retrieved when $\mathbf{A}$(SST) and $\mathbf{A}$(CF) are both high. However, only 11% of cases have $\mathbf{A}$(SST) and $\mathbf{A}$(CF) both $>0.7$,

so it was decided not to filter cases according to the averaging kernels. In the next section, the retrieved clear-sky and partly cloudy SSTs will be validated against drifting buoy SST measurements to make a better assessment of their quality.

### 5.4.3 Validating the IRIS-like IASI SSTs

The drifting buoy data was obtained from ICOADS Release 3.0 (*Freeman et al.*, 2017). A variety of quality control flags accompany the SSTs, enabling selection of high quality measurements. For example, if the drifting buoy reports the same position for over 21 days or has a 7-day average speed over $2\,\mathrm{m\,s^{-1}}$, the data was discarded. Additionally, IRIS-like IASI observations significantly affected by aerosols are discarded in the same way as for IRIS. Where multiple drifting buoy observations are collocated to a single IRIS-like IASI observation, the mean of the drifting buoy SSTs was taken. The retrieved IRIS-like IASI SSTs represent skin SSTs, so were converted to 'bulk' SSTs by adding 0.17 K (*Donlon et al.*, 2002). This correction represents the mean difference between skin and bulk SSTs for wind speeds above $6\,\mathrm{m\,s^{-1}}$ in both day and nighttime conditions. At lower wind speeds, the difference is much more variable during the daytime due to stratification of the upper ocean layers.

#### 5.4.3.1 Clear-sky results

Figure 5.17 compares the distributions of retrieved - drifting buoy SST for daytime and nighttime cases. For both sets of cases there are small cold biases for completely clear scenes (table 5.1). These biases increase slightly for cases with some cloud contamination ($\mathrm{CF} > 0\%$), as expected. The IASI cloud mask is not perfect, so there may be some cloud contamination even in the 0% CF cases which could explain these small biases. The differences between the biases and standard deviations for daytime and nighttime are fairly small. IASI has a sun-synchronous mid-morning orbit, with observations at around 9.30 am and 9.30 pm local solar time. The strength of thermal stratification is about the same at these two times of day on average, which may explain the small

differences. IRIS has a noon orbit, so the difference in the skin-bulk SST offset is likely to be much greater between day-time and night-time cases. Estimates of the

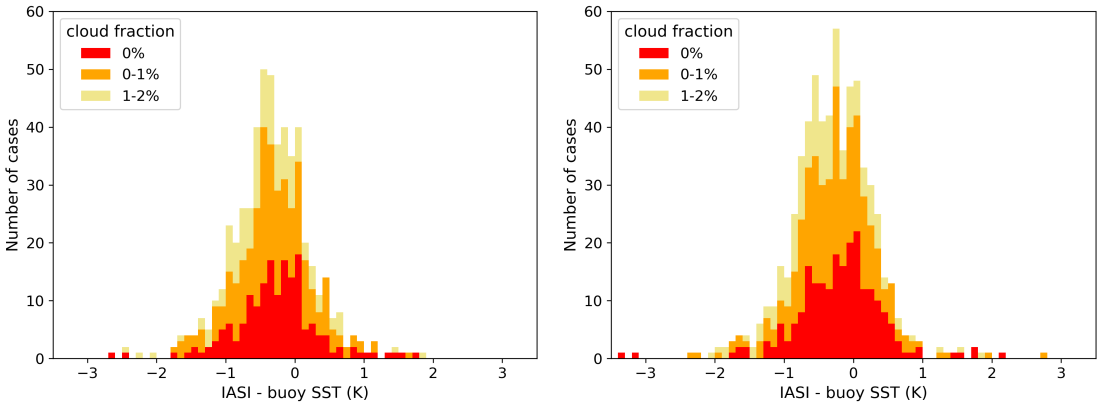



Figure 5.17: Stacked histograms showing the distribution of retrieved minus drifting buoy SST for the set of clear-sky IRIS-like IASI daytime cases (left) and nighttime cases (right), stratified by CF.

| CF (%) | Daytime | | | Nighttime | | |
|---|---|---|---|---|---|---|
| | Bias (K) | St. dev. (K) | $\sigma_{SST}$ (K) | Bias (K) | St. dev. (K) | $\sigma_{SST}$ (K) |
| 0 | -0.24 | 0.62 | 0.46 | -0.16 | 0.64 | 0.49 |
| 0-1 | -0.35 | 0.56 | 0.39 | -0.27 | 0.59 | 0.42 |
| 1-2 | -0.43 | 0.59 | 0.42 | -0.44 | 0.67 | 0.53 |

Table 5.1: IRIS-like IASI - buoy statistics and estimated SST uncertainties ($\sigma_{SST}$) for daytime and nighttime cases. The bias is computed from the median of the IRIS-like IASI - buoy SSTs.

uncertainties in the retrieved SSTs can be obtained by subtracting the contributions due to collocation error and drifting buoy measurement error from the standard deviation of the retrieved minus buoy SSTs. SST CCI L4 data (*Merchant et al.*, 2014) was used to estimate the contribution from collocation error. Drifting buoy measurement uncertainty was assumed to be 0.2 K (*Kennedy*, 2014). The estimated SST uncertainties in table 5.1 suggest that it should be possible to retrieve SSTs from IRIS with an uncertainty of 0.5 K or below in clear-sky conditions.

To put these results into context, we tested the retrieval with raw IASI spectra. For each group of 16 IASI pixels obtained as in §5.4.1, the spectrum corresponding to one

pixel is selected. The IRIS channel selection method was adapted to select a set of 20 IASI channels, choosing from the same spectral intervals as for IRIS. The measurement error covariance was constructed by summing $\mathbf{S}_{noise}$, $\mathbf{S}_{model}$, $\mathbf{S}_{cld}$ and $\mathbf{S}_{aerosol}$. $\mathbf{S}_{noise}$ is the covariance of the IASI radiometric noise, which is much smaller in magnitude than the IRIS radiometric noise. $\mathbf{S}_{cld}$ was computed from the covariance of the mean differences in the observed and simulated raw IASI spectra for the set of cases with 0, 0-1 and 0-2% CF. $\mathbf{S}_{model}$ and $\mathbf{S}_{aerosol}$ were computed as described in §5.3.2.2. SSTs were retrieved from the set of raw IASI spectra and the same chi-squared test on $\mathbf{K}\hat{\mathbf{x}} - \mathbf{y}$ and standard deviation test on the retrieved - a priori SST were applied. It was noted that the chi-squared distribution of $\mathbf{K}\hat{\mathbf{x}} - \mathbf{y}$ had a smaller median value than the expected distribution. This suggests that the uncertainty estimates for the observations and/or simulations are too high. Figure 5.18 shows the distributions of retrieved - drifting buoy SST for the sets of daytime and nighttime IASI cases with a CF of 0, 1 and 2%.

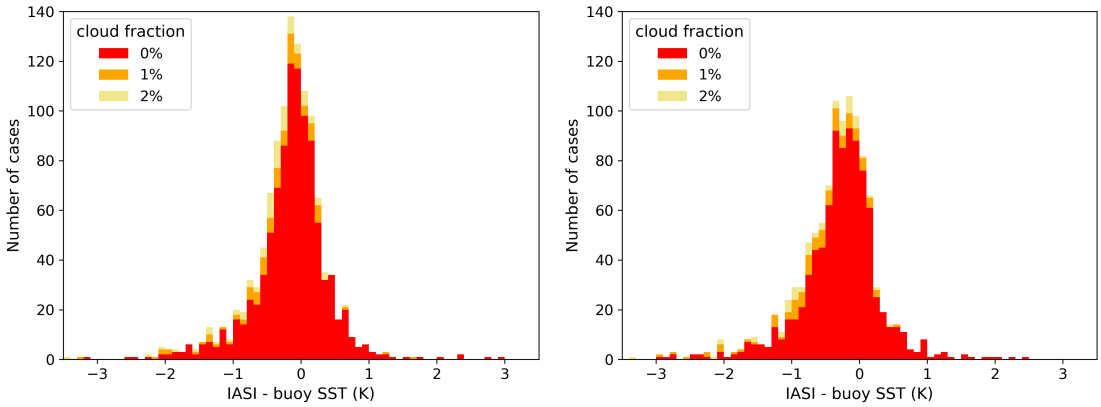



Figure 5.18: Stacked histograms showing the distribution of retrieved minus drifting buoy SST for the set of clear-sky raw IASI daytime cases (left) and nighttime cases (right), stratified by CF. Note the different colour scale to Figure 5.17, as the IASI pixel CF is given to a precision of only 1%.

The vast majority of cases have 0% CF, due to the small size of a single IASI pixel in comparison to the IRIS footprint. For the 0% CF cases, the biases are similar in magnitude to the biases in the IRIS-like IASI SSTs. However, the bias in the raw IASI SSTs is greater for the nighttime cases whereas the bias in the IRIS-like IASI

SSTs is greater for the daytime cases. The estimated SST uncertainties are similar to those for the IRIS-like IASI cases. This suggests that the quality of SSTs that can be retrieved from IRIS in clear-sky conditions should be similar to the quality of SSTs retrieved from IASI, although the IRIS SSTs have much lower spatial resolution. This is despite the IRIS measurement error being much higher and more correlated than for IASI. For the nighttime IASI cases, these results can be compared to *O'Carroll et al.*

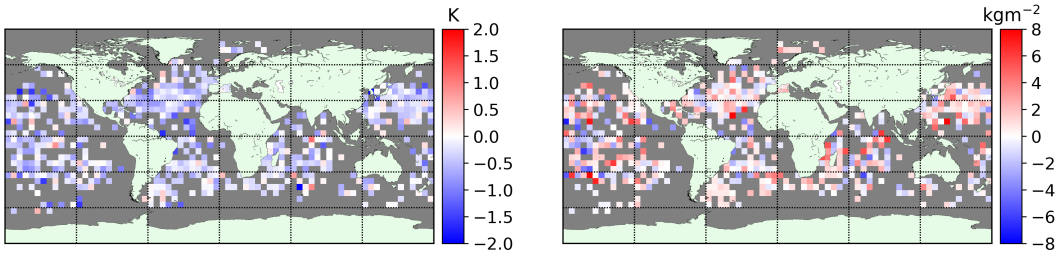
| CF (%) | Daytime | | | Nighttime | | |
|---|---|---|---|---|---|---|
| | Bias (K) | St. dev. (K) | $\sigma_{\mathrm{SST}}$ (K) | Bias (K) | St. dev. (K) | $\sigma_{\mathrm{SST}}$ (K) |
| 0 | -0.10 | 0.56 | 0.37 | -0.23 | 0.58 | 0.41 |
| 1 | -0.32 | 0.70 | 0.56 | -0.59 | 0.69 | 0.55 |
| 2 | -0.34 | 0.63 | 0.47 | -0.52 | 0.67 | 0.52 |

Table 5.2: Raw IASI - buoy statistics and estimated SST uncertainties ($\sigma_{\mathrm{SST}}$) for daytime and nighttime cases. The bias is computed from the median of IASI - buoy SSTs.

(2012), who also validated IASI SSTs against drifting buoy SSTs during the period October 2010 to July 2011. They determined a mean bias of -0.33 K for a set of 2715 collocations. For collocations deemed likely to be 100% cloud free, the bias in *O'Carroll et al.* (2012) reduced slightly to -0.25 K. They also validated IASI SSTs against AVHRR SSTs, revealing a mean bias of -0.35 K. The standard deviation of 0.58 K that we have obtained for the cases with 0% CF is much higher than the value of 0.40 K obtained in *O'Carroll et al.* (2012). A possible reason for this is the difference in the distance threshold used for collocating IASI and drifting buoy observations. In *O'Carroll et al.* (2012), the distance threshold was such that the drifting buoy observation was always in the central half of the IASI Instantaneous Field of View (IFOV). This is a much stricter threshold than the 50 km we use and means that collocation error is minimal, so the standard deviation is much closer to the SST uncertainty.

To understand the spatial distribution of the SST errors, they were gridded at 5° by 5° resolution over the whole of 2011. In clear-sky conditions, error in the retrieved water vapour profile is expected to be an important driver of error in the retrieved

SST. If the retrieved total column water vapour (TCWV) is too high, this results in increased atmospheric absorption which is compensated for by a warm bias in the SST. Figure 5.19 shows the spatially gridded SST errors and retrieved - a priori TCWV.




(a) (b)

Figure 5.19: Error in retrieved SST (a) and retrieved - a priori TCWV (b) for the set of clear-sky IRIS-like IASI cases. The estimated SST collocation errors were removed in (a).

The cool bias in the SSTs is fairly uniform whilst the retrieved - a priori TCWV is positive in most regions and not so uniform. There is a weak positive correlation of 0.27 between the two variables. This suggests that the TCWV differences are only able to explain a small fraction (7 %) of the variability in the SST errors.

#### 5.4.3.2 Partial cloud results

Figure 5.20 shows the distribution of retrieved - drifting buoy SST for the set of partly cloudy IRIS-like IASI cases and table 5.3 presents the corresponding validation statistics. The bias is -0.49 K for the clearest cases (CF $\leq$ 1 %), but changes sign to 0.28 K for cases with CF $\geq$ 15 %. SST uncertainties were estimated for each CF stratum in the same way as for the clear-sky retrievals. For the clearest strata, the uncertainties are roughly twice those found for the clear-sky retrievals and increase slightly at higher retrieved CF. This suggests that the addition of CF as an extra variable to retrieve results in a large increase in the retrieved SST uncertainty, regardless of the retrieved CF.

Figure 5.21 shows the relationship between the errors in retrieved CF and SST,

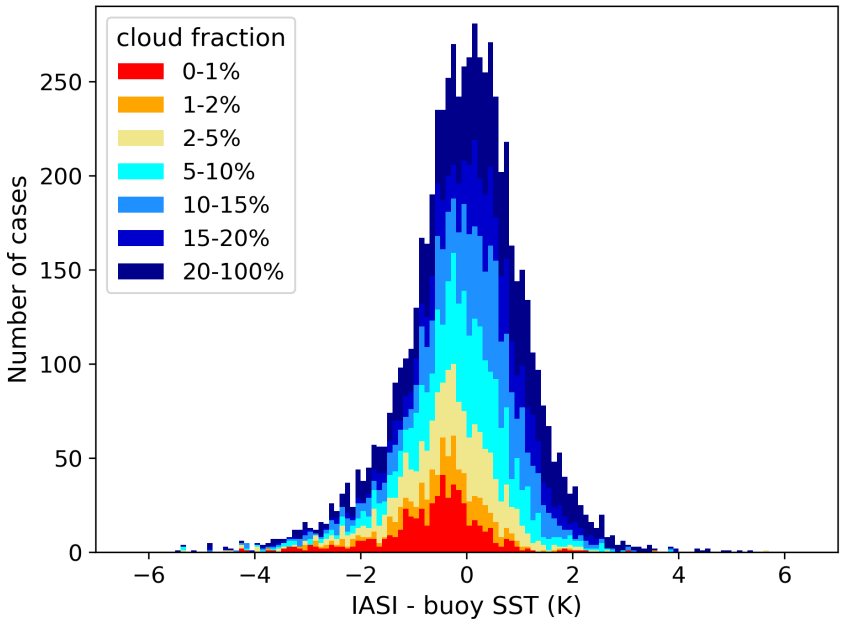



Figure 5.20: Stacked histograms showing the distribution of retrieved minus drifting buoy SST for the set of partly cloudy IRIS-like IASI cases, stratified by retrieved CF.

| CF (%) | Bias (K) | St. dev. (K) | SST uncertainty (K) |
|---|---|---|---|
| 0-1 | -0.49 | 0.98 | 0.89 |
| 1-2 | -0.35 | 0.92 | 0.82 |
| 2-5 | -0.25 | 1.04 | 0.95 |
| 5-10 | -0.02 | 1.02 | 0.93 |
| 10-15 | 0.19 | 1.08 | 1.00 |
| 15-20 | 0.28 | 1.08 | 0.99 |
| 20-100 | 0.28 | 1.32 | 1.25 |

Table 5.3: IASI - buoy statistics for the set of partly cloudy IRIS-like IASI cases. The bias is computed from the median of the IASI - buoy SSTs.

showing that in most cases, if the retrieved CF is overestimated, the retrieved SST is also overestimated. Spearman's correlation coefficient is 0.47, suggesting that CF errors are not the only important driver of SST errors. To better understand the SST and CF biases, the errors in the retrieved SSTs and CFs were gridded at 5° by 5° resolution over the whole of 2011. Figure 5.22 shows that the SST and CF errors are both spatially correlated in many regions.

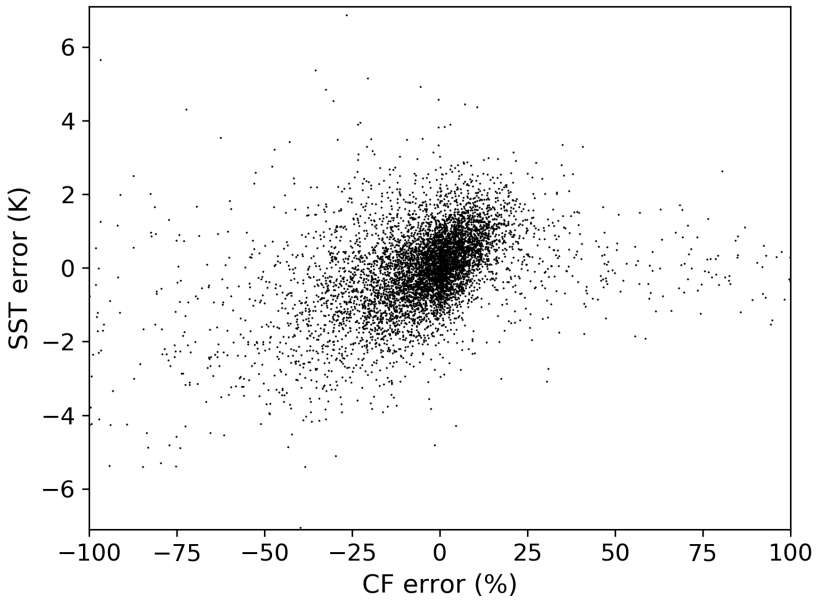
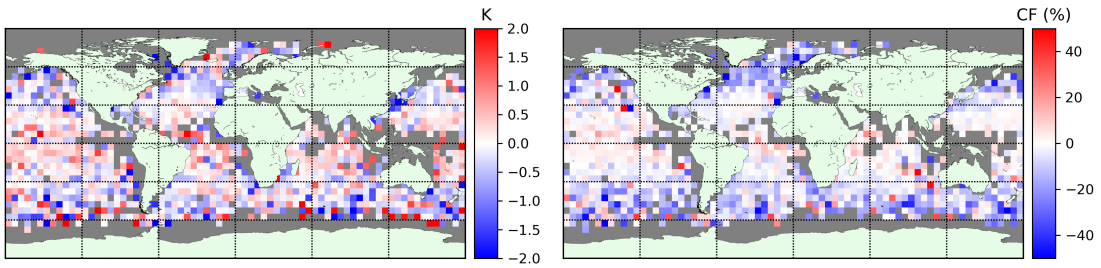



Figure 5.21: Scatter plot of errors in retrieved CF and SST for the set of partly cloudy IRIS-like IASI cases.




(a) (b)

Figure 5.22: Errors in retrieved SST (a) and CF (b) for the set of partly cloudy IRIS-like IASI cases. The estimated SST collocation errors were removed in (a).

The errors vary latitudinally, with the retrieved CFs tending to be too low at high latitudes and about right at low latitudes. This pattern is similar for the retrieved SSTs, as they tend to be slightly too low at high latitudes and slightly too high at low latitudes. The latitudinal variation in the SST/CF error may be explained by variables that vary with latitude, such as orbit angle and temperature. In §4.3, a possible orbitally varying bias in the IRIS BTs was found. This bias should not be present in the IRIS-like IASI BTs, suggesting that the latitudinal variation is explained by other

variables. We studied the relationship between the SST errors and the retrieved SST, CF and TCWV, all of which vary with latitude. Figure 5.23 shows the dependencies of the SST error on these three variables.

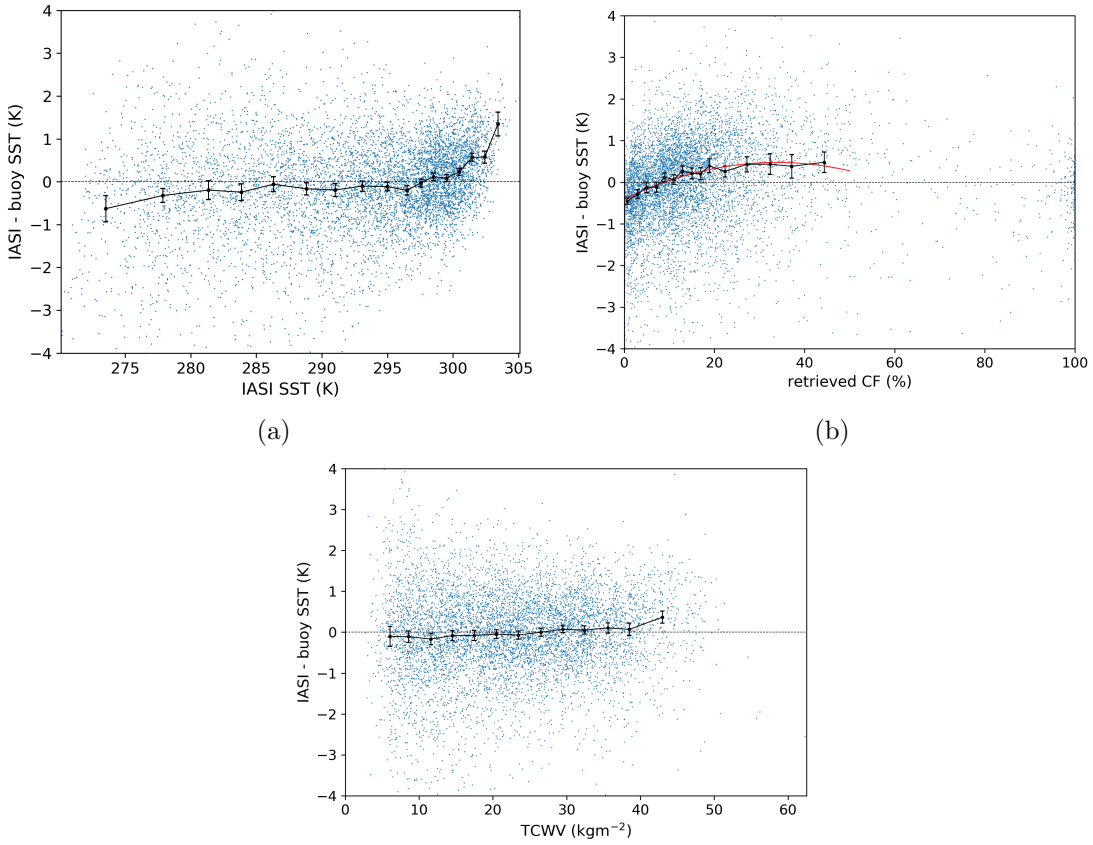



Figure 5.23: Dependency of retrieved SST error on retrieved SST (a), retrieved CF (b) and TCWV (c). Data points are in blue. The black circles and bars indicate the median and $\pm 2\times$ standard error, plotted for bins of width depending on the density of data points. The red curve in (b) indicates the quadratic least-squares fit.

The SST error switches from negative to positive with increasing SST and CF, but the trend is non-linear in both cases. The SST bias has a weaker dependence on TCWV. The relationship between the SST bias and these three predictors was quantified using multiple linear regression. As there are relatively few cases with CF $> 50\,\%$ and the SSTs for these cases have higher uncertainties, it was decided to discard these cases. It was assumed that the relationship between the SST bias and the three predictors

can be modelled as a polynomial function. We started with single linear terms in each of the predictors and obtained the regression statistics. If the coefficients of the terms are highly statistically significant (coefficients differ from zero by more than three sigma), they are retained. Higher order terms and cross terms were added progressively, retaining them if they are highly statistically significant and increase $R^2$ (explained variance) by at least 0.001. After this process, the following function was found to represent the relationship between the SST bias (B) and three predictors:

$$\begin{aligned} B = &- 0.5172 + 0.1561T - 0.0118T^2 + 0.0002T^3 + 0.0226C - 0.1050W \\ &+ 0.0022TC - 0.0014TW + 0.0040CW \end{aligned} \tag{5.14}$$

where T, W and C represent the retrieved SST (°C), TCWV ($kgm^{-2}$) and CF (%). The $R^2$ value was 0.244, indicating that only a quarter of the total variance is explained by the retrieved SST, CF and TCWV. It is likely that a large fraction of the total variance is due to random errors. If this is the case, the remaining variance due to correlated/structured errors may be largely explainable by these three predictors.

This bias correction model was applied to the IRIS-like IASI SSTs and the errors in the bias-corrected SSTs were gridded as before (Figure 5.24). The warm bias in

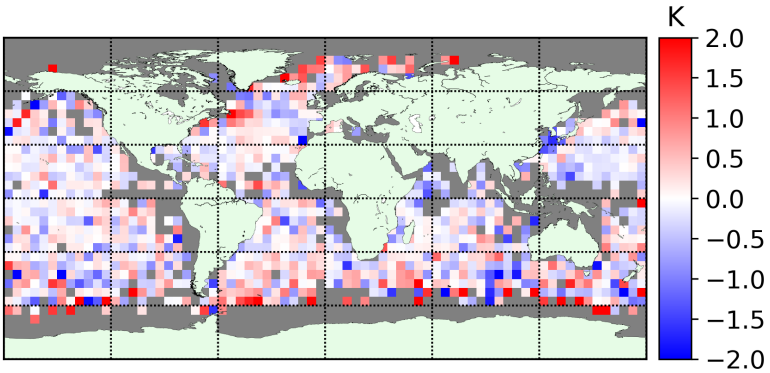



Figure 5.24: Errors in retrieved SST for the set of partly cloudy IRIS-like IASI cases after applying the bias correction model.

the SSTs at low latitudes appears to have reduced after bias correction, but the cold

bias at high latitudes has been overcorrected, resulting in a fairly strong warm bias in parts of the North Atlantic and some areas below 30°S. Figure 5.25 shows the latitudinal variation in the mean SST error before and after bias correction. The warm

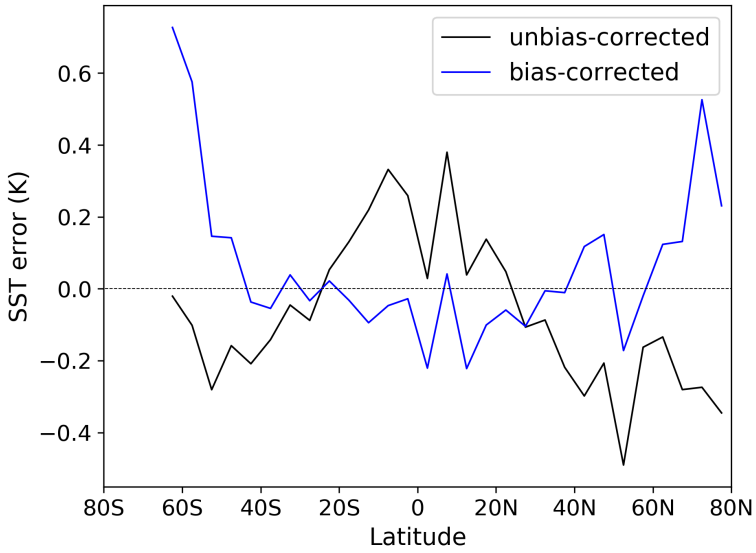



Figure 5.25: SST error variations with latitude for the set of partly cloudy IRIS-like IASI retrievals before and after applying the bias correction model.

bias at high latitudes after bias correction is as strong or stronger than the cold bias at high latitudes before bias correction. There is some reduction in the bias at low latitudes, although it slightly overcorrects the bias. We expect the IRIS SST biases to have similar spatial scales and magnitudes to the IRIS-like IASI SST biases, but the spatiotemporal distribution of the biases may be different due to different synoptic-scale weather patterns in 1970 and 2011. This means that this bias correction model may be less effective in correcting whatever biases are present in the SSTs retrieved from IRIS.

## 5.5 Chapter summary

In this chapter, we have developed an OE-based scheme for retrieving SST from IRIS observations identified as clear-sky and partly cloudy. To test these retrieval methods,

IRIS-like IASI spectra were generated with the same error characteristics as the IRIS spectra. By validating the retrieved IRIS-like IASI SSTs against high-quality drifting buoy SSTs, it was shown that SST can be retrieved to an accuracy of 0.4-0.5 K in clear-sky conditions. However, our results show that even a small amount of cloud contamination ($< 2\%$) in the IRIS field of view can result in a cold bias of up to 0.6 K.

The inclusion of cloud fraction as an additional parameter to be retrieved greatly increases the uncertainty in the SST, even for cases where the retrieved CF is close to zero. This is at least in part due to errors in the retrieved CF, which are correlated with errors in the retrieved SST. The partly cloudy SST errors were found to be spatially correlated in some regions with an overall zonal bias. A regression-based bias model was derived to express the SST error in terms of the retrieved SST, CF and TCWV. This was shown to have limited effectiveness in reducing the SST bias on average, and will therefore not be applied to the SSTs retrieved from partly cloudy IRIS spectra in the next chapter.

# CHAPTER 6

# RETRIEVING SEA SURFACE TEMPERATURE FROM IRIS

## 6.1 Introduction

In the previous chapter, the SST retrieval method for IRIS was developed and tested using IASI data in clear-sky and partly cloudy conditions. Validation of the SSTs retrieved from IRIS-like IASI observations suggested that, in theory, SSTs can be retrieved from IRIS to an accuracy of about 0.4-0.5 K in conditions known a priori to be clear-sky and 0.8-1.0 K in cases where the expected cloud fraction is up to 20 %. However, these uncertainty estimates fail to take into account potential biases in the retrieved SSTs, which can be due to a variety of factors. The SSTs retrieved using the clear-sky retrieval method can be affected by cold biases due to cloud contamination. From analysis of the clear-sky IRIS-like IASI retrieval it was shown that biases in retrieved TCWV may produce biases in retrieved SST, although the correlation is weak. A regression-based bias model was developed for the set of partly cloudy IRIS-like IASI cases, to explain biases in the retrieved SSTs in terms of SST, CF and TCWV. Application of this bias model was found to reduce the SST biases in some regions but increase them in others. Therefore, we do not expect this bias model to improve the quality of the SSTs retrieved from partly cloudy IRIS cases and so it will not be applied

to the IRIS SSTs.

In section 6.2, diagnostics and results are presented for the SSTs retrieved from IRIS using the methodology set out in chapter 5. Results are presented separately for clear-sky and partly cloudy IRIS cases. Section 6.3 then explains how the retrieved SSTs are processed into a monthly, gridded dataset with uncertainty estimates. The issue of error correlations in the IRIS SSTs is also discussed in this section. In section 6.4, comparisons are made between the gridded IRIS SSTs and two historical in-situ based datasets: HadISST2 and HadSST3. This enables us to assess whether the IRIS and in-situ SSTs agree, taking into account their respective uncertainties. Finally in section 6.5, the clear-sky and partly cloudy gridded IRIS SSTs are compared as a check for consistency between the two different approaches to retrieving SST from IRIS.

## 6.2 Retrieving SST from IRIS

### 6.2.1 Clear-sky

The set of clear-sky IRIS cases previously identified from IDCS data is likely to be a small fraction of the total number of clear-sky cases. This is largely a consequence of geolocation error in the IDCS data. We can try to identify additional clear-sky cases by starting with the set of cases where the CERA-20C ensemble mean CF is below 1 %. Many of these cases may have a CF above 1 % in reality, but the SSTs retrieved from these cases can be filtered to remove cloud-contaminated cases. To improve the quality of the a priori simulations, we discarded cases where the CERA-20C ensemble SST standard deviation was above 0.5 K. SSTs were retrieved for the remaining cases, and the same chi-squared test on $\mathbf{K}\hat{\mathbf{x}} - \mathbf{y}$ and standard deviation based test used to filter the IRIS-like IASI SSTs were applied to the set of IRIS SSTs. A chi-squared test on $\mathbf{y} - \mathbf{K}\mathbf{x_a}$ was also applied to remove cases where there are particularly large differences between the observations and a priori simulations.

It was found that the chi-squared distribution for $\mathbf{y} - \mathbf{K}\mathbf{x_a}$ was shifted to higher val-

ues relative to the expected distribution. This indicates that the uncertainty estimates in the measurements and/or the forward simulations are too low. The chi-squared distribution for $\mathbf{K}\hat{\mathbf{x}} - \mathbf{y}$ is close to the expected distribution, so this is evidence that the CERA-20C SST uncertainties are likely to be too low. To compensate for this, we adjusted the method of selecting the chi-squared threshold for $\mathbf{y} - \mathbf{K}\mathbf{x}_\mathbf{a}$. The observed chi-squared distribution was binned relatively coarsely (bin width = 5) and the difference was taken between the central value of the modal bin and the median of the expected distribution. The expected distribution was shifted by this difference and the threshold was selected as done previously.

The methods of selecting the CERA-20C-derived and IDCS-derived sets of clear-sky IRIS cases are very different, so they were analysed separately. The locations of the observations for each set of cases are shown in Figure 6.1. The spatial distribution is very different for each set of cases: the CERA-20C-derived cases are concentrated in the Mediterranean and close to coastlines whereas the IDCS-derived cases are more spread out over each ocean region.

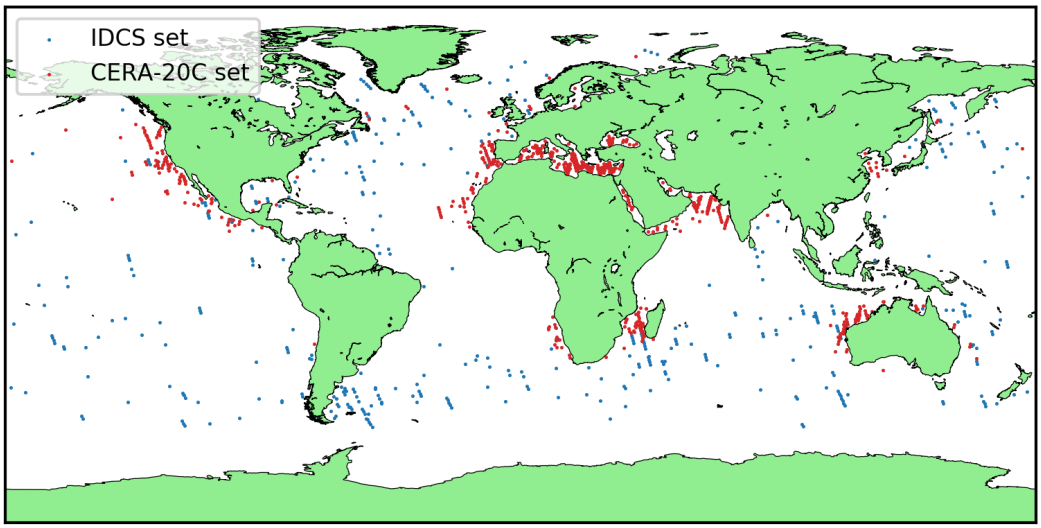



Figure 6.1: Locations of IRIS observations for the sets of CERA-20C and IDCS-derived clear-sky cases.

Figure 6.2 shows the distribution of retrieved - a priori SST for these two sets of cases, showing the distribution for the CERA-20C cases before and after application of the QC tests. For the CERA-20C cases, the distribution pre-QC is relatively symmetrical without a cold tail. This suggests that there is not a high level of cloud contamination in this set of cases. The standard deviation reduces from 1.22 K to 0.66 K post-QC with the median remaining at about +0.1 K. The post-QC standard deviation is close to the clear-sky IRIS-like IASI - drifting buoy SST standard deviation and is lower than the standard deviation for the IDCS cases (0.82 K). The higher standard deviation for the IDCS cases may be because an upper limit was not imposed on the a priori SST uncertainty when selecting these cases. It also may be because stronger filtering was applied to the CERA-20C cases compared to the IDCS cases. The mean averaging kernel of the SST is 0.997 for the CERA-20C cases and 0.995 for the IDCS cases. Both of these values are very close to one, indicating that the retrieval has very low sensitivity to the a priori SST.

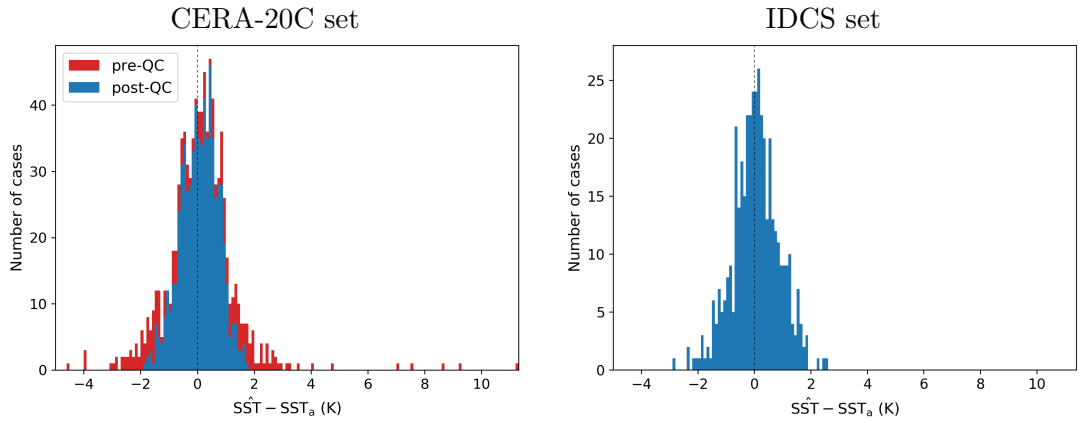



Figure 6.2: Distribution of retrieved minus a priori SST for the set of CERA-20C (left) and IDCS (right) derived clear-sky IRIS cases. The distribution for the CERA-20C cases is shown before and after application of QC tests.

Figure 6.3 shows that there are clear differences between the median retrieved and a priori temperature and water vapour profiles at lower levels for the CERA-20C cases but not the IDCS cases. For the CERA-20C cases, the median retrieved temperature and water vapour are up to 0.6 K lower and 1 g/kg higher than the a priori temperature

and water vapour. This acts to reduce the BTs simulated from the retrieved state ($\mathbf{K}\hat{\mathbf{x}}$)

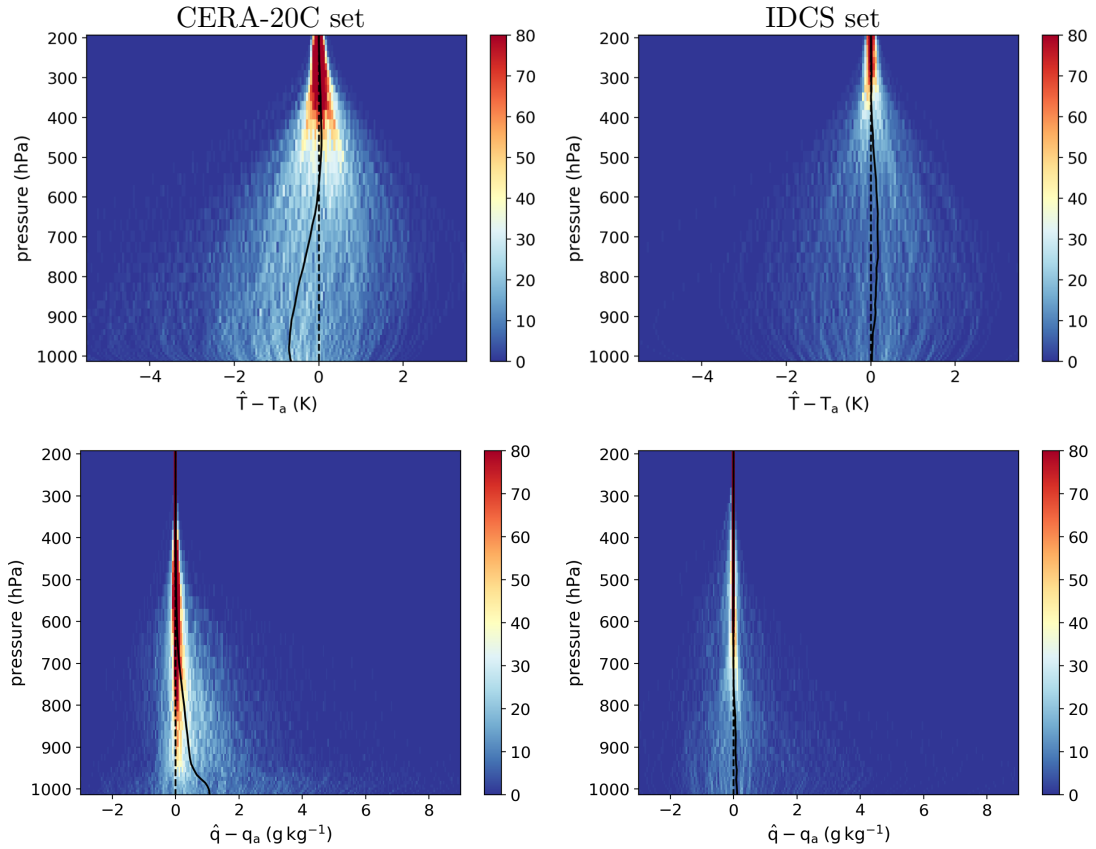



Figure 6.3: 2D histograms of retrieved minus a priori temperature and water vapour (q) profiles for the set of CERA-20C (left panels) and IDCS (right panels) derived clear-sky IRIS cases. The median difference in each case is indicated by the solid black line.

relative to the a priori state ($\mathbf{Kx_a}$). These differences appear to have very little effect on the retrieved SSTs, since the median retrieved minus a priori SST is only +0.1 K. Figure 6.4 shows density plots of $\mathbf{K}\hat{\mathbf{x}} - \mathbf{y}$ for each set of clear-sky cases. For the set of CERA-20C-derived cases, there is a negative bias of up to 0.3 K in the $CO_2$ channels ($< 760\,\mathrm{cm}^{-1}$) and higher wavenumber window channels. This spectrally varying bias is not seen for the set of IDCS-derived cases, as the bias correction was estimated using this set of cases. The bias in $\mathbf{K}\hat{\mathbf{x}} - \mathbf{y}$ for the CERA-20C-derived cases indicates that the bias correction applied to the IRIS observations may not be appropriate in all cases.

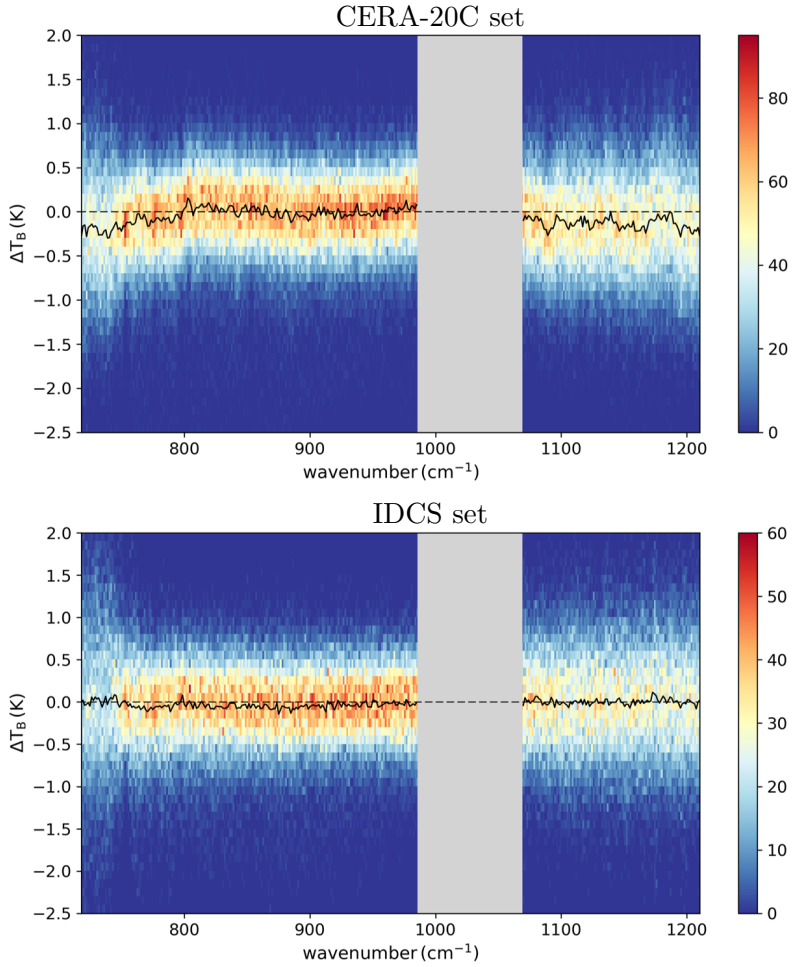



Figure 6.4: 2D histograms of $\mathbf{K}\hat{\mathbf{x}} - \mathbf{y}$ for the sets of CERA-20C and IDCS-derived clear-sky IRIS cases. The median difference in each case is indicated by the solid black line.

The spatial distribution of each set of cases is very different and the scene temperature is 6 K warmer on average for the CERA-20C-derived cases. From chapter 4, there is evidence that the IRIS bias may vary with scene temperature, so this could be a factor contributing to the spectral bias in $\mathbf{K}\hat{\mathbf{x}} - \mathbf{y}$ for this set of cases. A second difference between the two sets of cases is that the IDCS-derived cases are daytime only; chapter 4 provides evidence that the IRIS bias may vary diurnally. It is unlikely that the bias in $\mathbf{K}\hat{\mathbf{x}} - \mathbf{y}$ is due to cloud contamination as the spectral signature of cloud is fairly constant across the IR window channels for water clouds and greatly varies in the 800-1000 $\mathrm{cm}^{-1}$

for ice clouds (*Huang et al.*, 2004).

In summary, the diagnostics presented in this section suggest that the retrieval is working successfully for the two sets of clear-sky IRIS cases. However, as seen in chapter 4 there are potentially remaining spectral biases in the IRIS observations which propagate into the retrieved SSTs. In §6.3, the clear-sky IRIS SSTs will be gridded into a monthly dataset and compared with the in-situ based SST datasets.

### 6.2.2 Partial cloud

SSTs were retrieved from the set of IRIS observations with an expected CF of 20 % or below, selected using the threshold method in §5.3.3.1. As with the IRIS-like IASI cases, the retrieval was iterated three times. For the a priori CF uncertainty, we used the value calculated from the set of partly cloudy IRIS-like IASI cases. This may be an underestimate of the true CF uncertainty, as the errors in NWP CFs from 1970 are likely to be greater than in 2011. The same chi-squared and standard deviation tests applied to the IRIS-like IASI cases were applied to this set of IRIS cases. Figure 6.5 shows the distribution of retrieved minus a priori SST, stratified by retrieved CF. Table 6.1 shows that there is very little relative bias between the retrieved and a priori SSTs for the clearest cases. The standard deviations of the differences are higher than for

| Cloud fraction (%) | Median difference (K) | St. dev. (K) |
|---|---|---|
| 0-1 | -0.07 | 1.28 |
| 1-2 | 0.03 | 1.33 |
| 2-5 | 0.35 | 1.30 |
| 5-10 | 0.33 | 1.27 |
| 10-15 | 0.31 | 1.33 |
| 15-20 | 0.31 | 1.30 |
| 20-100 | 0.21 | 1.33 |

Table 6.1: Retrieved minus a priori SST statistics for the set of partly cloudy IRIS cases.

the IRIS-like IASI - drifting buoy differences, but this is expected as the uncertainties in the NWP SSTs in 1970 are likely to be much higher than modern drifting buoy

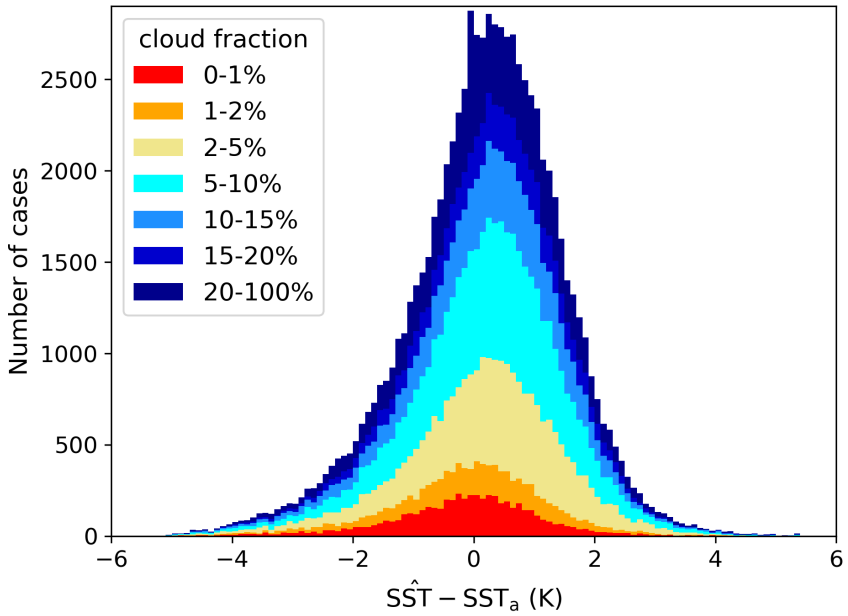



Figure 6.5: Stacked histograms showing the distribution of retrieved minus a priori SST for the set of partly cloudy IRIS cases, stratified by retrieved cloud fraction.

uncertainties.

Figure 6.6 shows that the median differences between the retrieved and a priori $T$ and $\ln q$ profiles are very small at all levels. However, the standard deviations of the

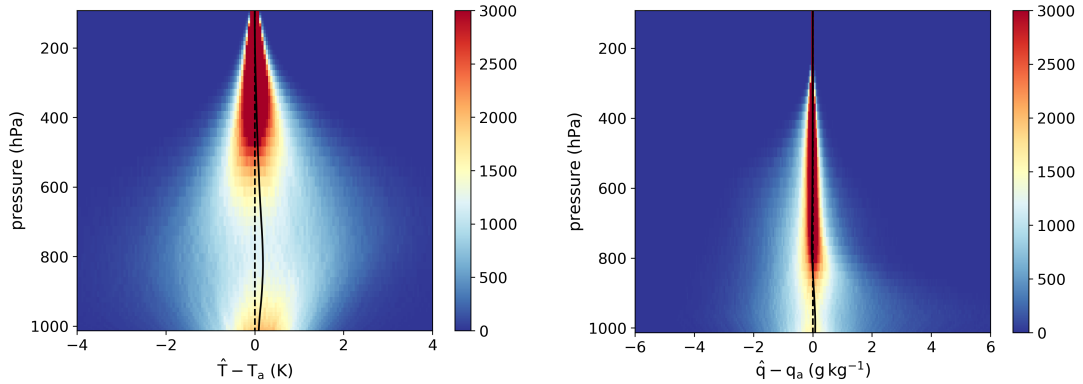



Figure 6.6: 2D histogram of retrieved minus a priori $T$ (left) and $q$ (right) profiles for the set of partly cloudy IRIS cases. The median difference is indicated by the solid black line.

differences are much higher than for the partly cloudy IRIS-like IASI retrievals. This

may be explained by the NWP $T$ and $q$ profiles being less accurate in 1970 than in 2011.

Figure 6.7 shows the distribution of $\mathbf{K}\hat{\mathbf{x}} - \mathbf{y}$. The median difference is below 0.2 K

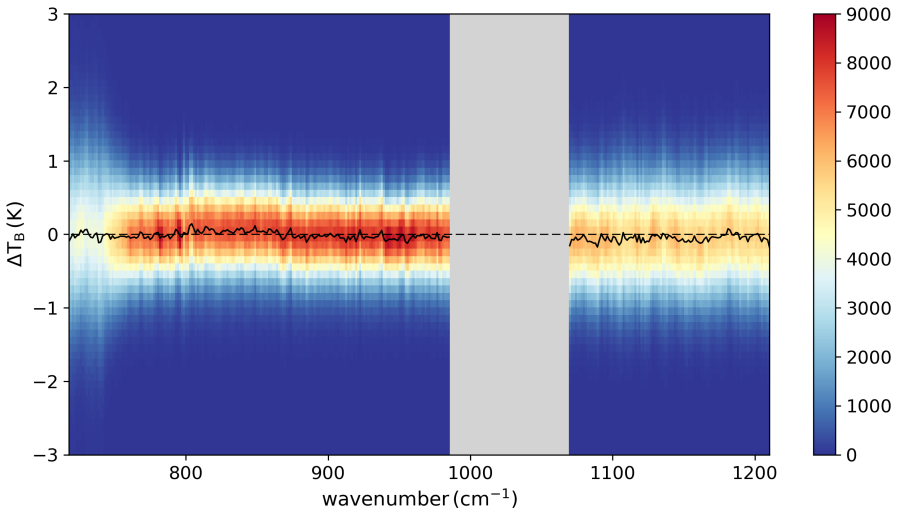



Figure 6.7: 2D histogram of $\mathbf{K}\hat{\mathbf{x}} - \mathbf{y}$ for the set of partly cloudy IRIS cases. The median difference is indicated by the solid black line.

for all channels and the standard deviation is 0.4-0.5 K for the window channels between 810-985 $\mathrm{cm}^{-1}$, increasing to 0.5-0.65 K for the window channels over 1070 $\mathrm{cm}^{-1}$. This shows that the retrieved state is highly consistent with the observations as there is very little relative bias between them.

Figure 6.8 compares the distributions of retrieved and a priori CF. The median retrieved CF is only 8 %, compared to 13 % for the a priori CFs. In §5.4.2.2 it was noted that ERA-Interim tended to overestimate the CF relative to the observed IASI CFs. It is therefore unsurprising that the ERA-20C and CERA-20C-derived CFs are much higher on average than the retrieved CFs for the set of IRIS cases. Figure 6.9 compares the SST and CF averaging kernels for this set of retrievals, with the data points coloured by retrieved CF. The distribution of points is very similar to that for the set of partly cloudy IRIS-like IASI retrievals.

The diagnostics presented in this section show that the retrieval is working in a

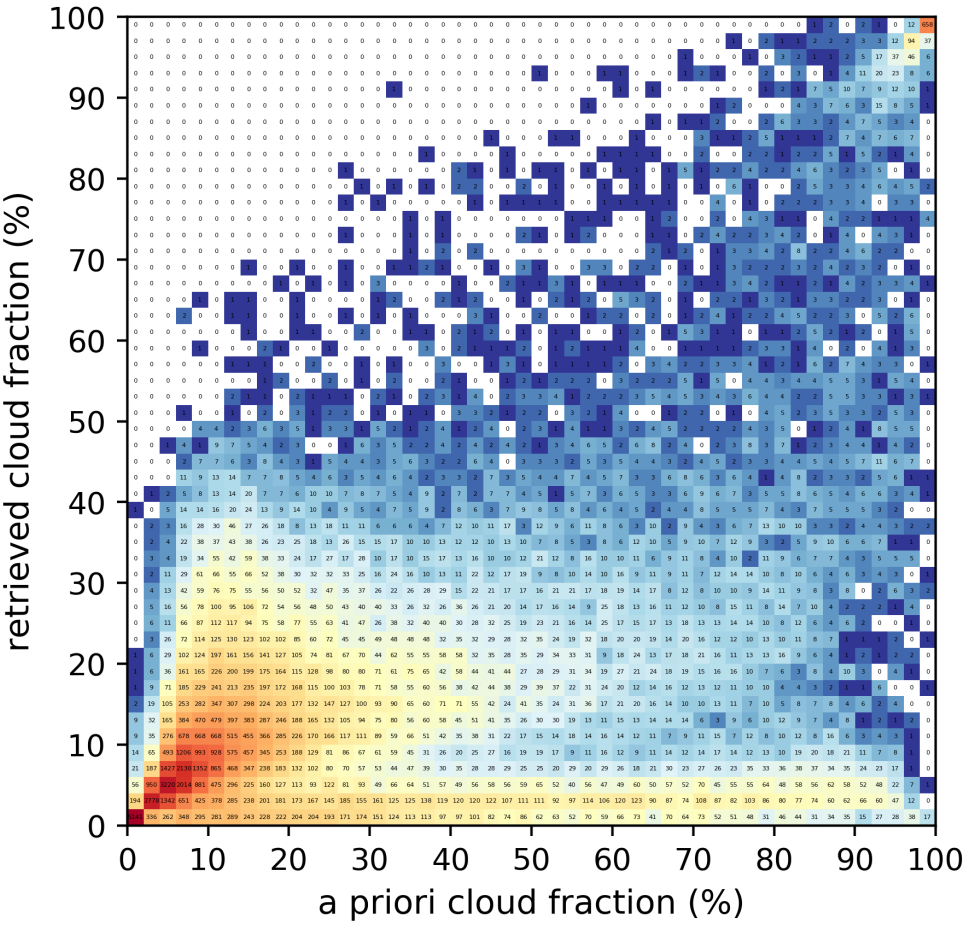



Figure 6.8: 2D histogram of retrieved vs a priori CF for the set of partly cloudy IRIS cases. Values in bins indicate frequencies.

manner consistent with the partly cloudy IRIS-like IASI retrieval. In the following section, the partly cloudy IRIS SSTs will be processed into a gridded, monthly dataset and compared with the in-situ based SST datasets.

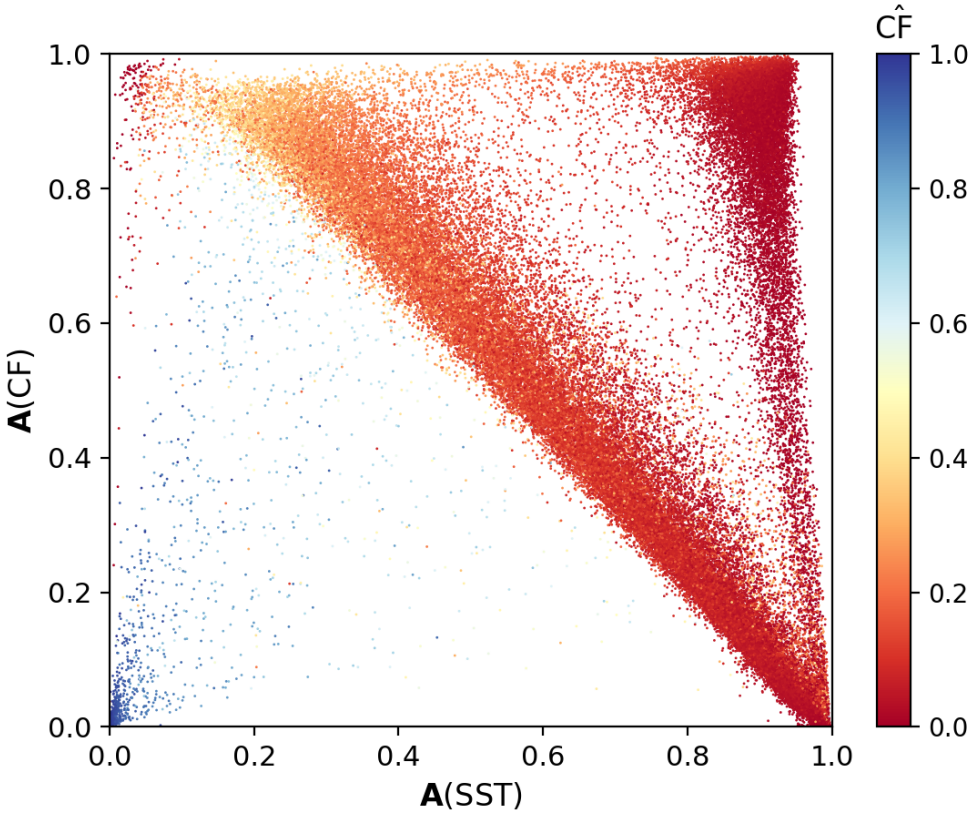



Figure 6.9: Comparison of SST and CF averaging kernels for the set of IRIS retrievals with $\leq 20\%$ expected CF. Data points are coloured according to retrieved CF.

## 6.3 Creating a dataset of IRIS SSTs

We have retrieved two small sets of IRIS SSTs from clear-sky cases and a much larger set of about 70,000 IRIS SSTs from partly cloudy cases. We can now construct a monthly, gridded dataset of these SST estimates. Each set of clear-sky SSTs are gridded separately and not included in the large set of partly cloudy SSTs. This is because the different sets of SSTs have different error characteristics. Additionally, for the set of partly cloudy SSTs, the relatively small number of cases with a retrieved cloud fraction over 50 % are discarded. A small number of IRIS SSTs below -1.8 °C (freezing point of seawater) and above 307 K are also discarded.

A spatial resolution of 5° was chosen for gridding the IRIS SSTs, to ensure that almost all gridboxes contain at least one SST. No attempt was made at spatial infilling,

so gridbox SSTs were calculated only for gridboxes with at least one SST. The dataset is produced as follows:

1. IRIS SSTs are converted to anomalies from a daily climatology. The climatology is produced by the SST CCI project (*Merchant et al.*, 2014): it covers the period 1981-2010 and is gridded at 0.05° resolution.

2. IRIS SST anomalies are gridded at monthly, 5° resolution. For each gridbox, the weighted average of the IRIS SST anomalies is taken, where the weights are equal to the inverse square of the SST uncertainties.

3. The SST climatology is regridded to monthly, 5° resolution. This is added to the gridded IRIS SST anomalies to produce gridded IRIS SSTs.

The uncertainties in the gridded SSTs are calculated according to equation 6.1,

$$\sigma_{\text{grid}} = \sqrt{\sigma^2_{\overline{\text{SST}}} + \sigma^2_{\text{sam}}} \tag{6.1}$$

where $\sigma_{\overline{\text{SST}}}$ is the uncertainty in the gridbox average SST originating from uncertainties in the individual SSTs and $\sigma_{\text{sam}}$ is the sampling uncertainty. If we initially assume that the errors in the IRIS SSTs are uncorrelated, then $\sigma_{\overline{\text{SST}}} = \frac{\sigma_{\text{SST}}}{\sqrt{\text{n}}}$ where n is the number of SST measurements in the gridbox. $\sigma_{\text{SST}}$ estimates are obtained from the validation of IRIS-like IASI against drifting buoy SSTs. For both sets of clear-sky IRIS SSTs, $\sigma_{\text{SST}}$ is a single value obtained from the full set of clear-sky IRIS-like IASI cases (daytime and nighttime combined). For the partly cloudy SSTs, $\sigma_{\text{SST}}$ varies with retrieved CF (§5.4.3.2).

The gridbox sampling error represents the error in the gridbox average SST due to the incomplete spatiotemporal sampling of IRIS SSTs within a gridbox. It is not possible to estimate the sampling error directly, but the sampling uncertainty can be estimated empirically from SST CCI data. For each year of available CCI data (1982-2016), two sets of gridded monthly SST anomalies are calculated. The first set is

calculated with complete spatiotemporal sampling and the second set is calculated with the spatiotemporal sampling of the IRIS SSTs. The difference between the two sets of anomalies provides yearly estimates of the gridbox sampling errors. The standard deviation of the yearly sampling errors forms our estimate of the sampling uncertainty, $\sigma_{\mathrm{sam}}$.

### 6.3.1 Error correlations in the IRIS SSTs

We investigated the assumption that the errors in the IRIS SSTs are uncorrelated. Figure 6.10 shows a map of the retrieved minus a priori SSTs for the set of clear-sky CERA-20C-derived cases, focusing on the Mediterranean region where this set of cases are most heavily concentrated. This figure suggests that there may be some spatial/temporal correlation in the retrieved minus a priori SSTs, because there are clusters of positive/negative SST differences. This could be due to correlation in the IRIS SST errors or the a priori SST errors, or both. Correlation in the IRIS SST errors could be due to a combination of correlation in the forward model errors and correlation in the observed IRIS BT errors.

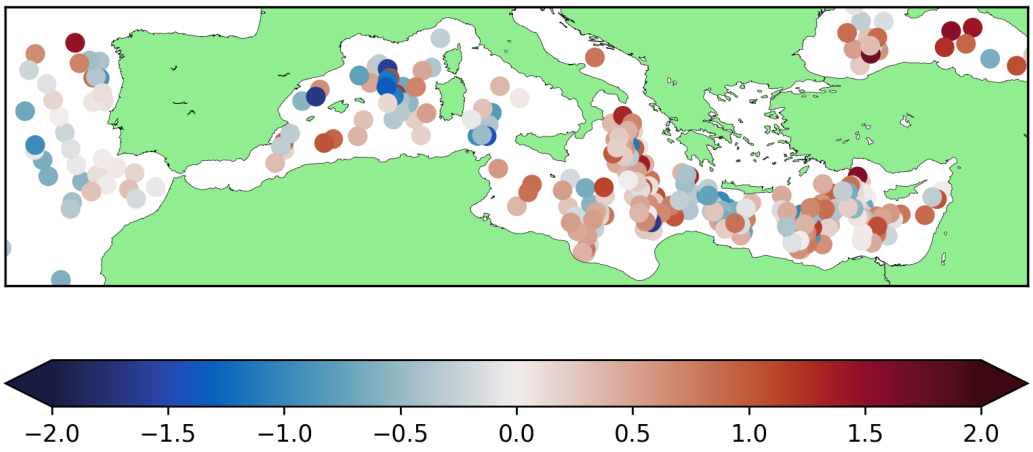



Figure 6.10: Map of retrieved - a priori SSTs for the set of clear-sky CERA-20C-derived cases, zoomed in to the Mediterranean region. The points are scaled approximately to the IRIS footprint size.

We can estimate the degree of spatial/temporal correlation in the IRIS BTs by

comparison with THIR window channel BTs for the set of homogeneous cases, identified in chapter 4. The procedure is to take the difference between all IRIS-THIR BT differences (double difference) and sort by time interval ($\Delta$t). Figure 6.11 shows the BT differences plotted as a function of $\Delta$t, up to $\Delta t = 1$ hour.

The IRIS/THIR observation frequency is 16 seconds, so $\Delta$t is always a multiple of 16 seconds. The standard deviation of the differences is seen to more than double from 0.63 K for $\Delta t = 16$ s to 1.32 K for $\Delta t = 3600$ s (1 hour), although there are considerable fluctuations. From the standard deviations, the correlation between the differences can be estimated as a function of $\Delta$t. Let the IRIS-THIR BT differences for two cases be $\delta_1$ and $\delta_2$. The uncertainty in the double difference ($\delta_{21} = \delta_2 - \delta_1$) can be expressed as:

$$\begin{aligned}\sigma^2(\delta_{21}) &= \left(\frac{\partial\delta_{21}}{\partial\delta_1}\right)^2 \sigma^2(\delta_1) + \left(\frac{\partial\delta_{21}}{\partial\delta_2}\right)^2 \sigma^2(\delta_2) + 2r_{21}\left(\frac{\partial\delta_{21}}{\partial\delta_1}\right)\left(\frac{\partial\delta_{21}}{\partial\delta_2}\right)\sigma(\delta_1)\sigma(\delta_2) \\ &= (2 - 2r_{21})\sigma^2(\delta)\end{aligned} \tag{6.2}$$

where $r_{21}$ is the correlation between $\delta_1$ and $\delta_2$ and $\sigma(\delta_1) = \sigma(\delta_2) = \sigma(\delta)$. From equation 6.2:

$$r_{21} = 1 - \frac{\sigma^2(\delta_{21})}{2\sigma^2(\delta)} \tag{6.3}$$

$\sigma(\delta_{21})$ and $\sigma(\delta)$ are known; $\sigma(\delta_{21})$ varies with time interval as shown in Figure 6.11 and $\sigma(\delta) = 0.89$ K. Figure 6.11 shows that the correlation is 0.75 at $\Delta t = 16$ s, reducing to near zero on average at around $\Delta t = 1500$ s. The correlation rises to nearly 0.5 at around $\Delta t = 2250$ s, although there is considerable noise in the signal. For the purpose of calculating the gridbox SST uncertainty, we are interested only in time intervals up to 112 s (within an orbit) and integer multiples of the orbital period up to one month. The correlations were therefore computed for cases with these time intervals only and also stratified by distance, up to the maximum distance in a 5° gridbox. Figure 6.12 shows a histogram of the correlation, binned by distance and time interval. The intraorbit correlations ($\leq 112$ s) are in the range 0.6 to 0.75 whereas the interorbit correlation is generally weaker than 0.5. As in Figure 6.11 the signal is noisy. We make

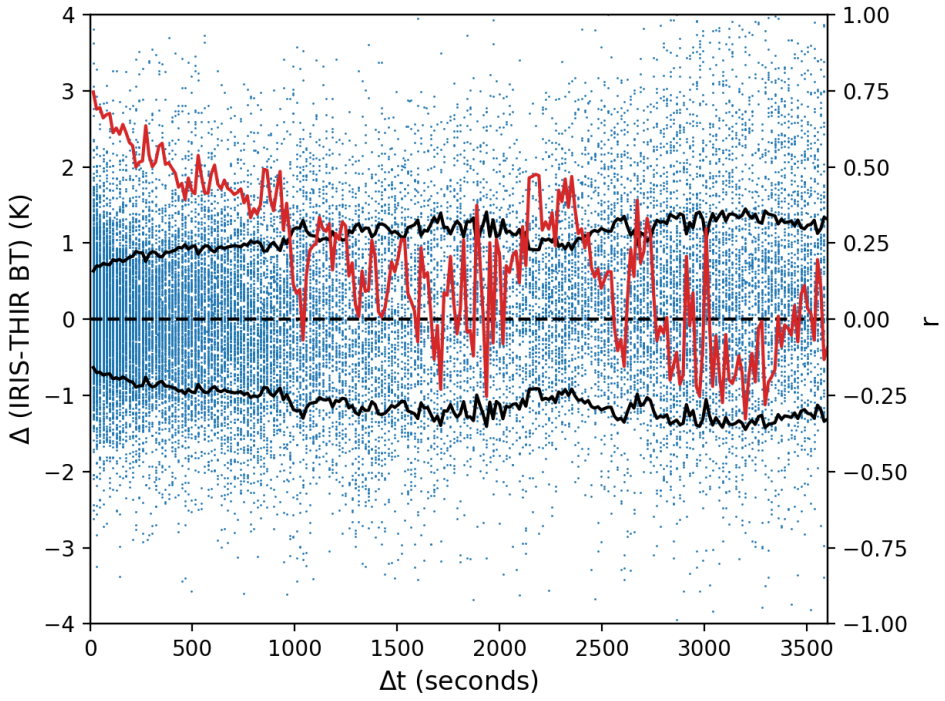



Figure 6.11: Differences between IRIS-THIR BT differences as a function of time interval up to $\Delta t = 3600\,s$ (1 hour). The solid black lines indicate $\pm 1\sigma$ of the differences every 16 seconds and the solid red line indicates the correlation between the differences every 16 seconds.

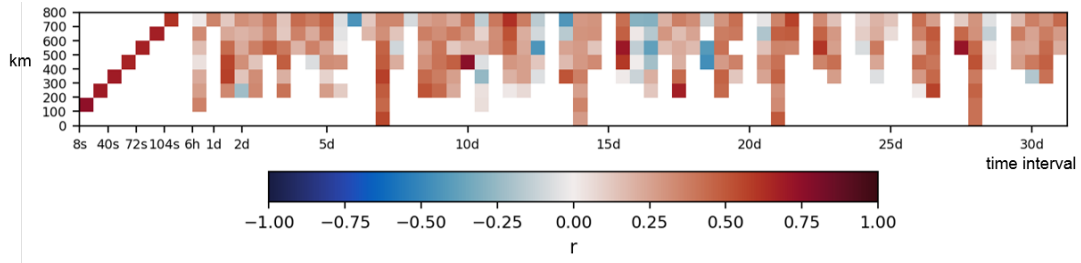



Figure 6.12: 2D histogram of correlation in the IRIS-THIR window BT as a function of time interval and distance within a 5° gridbox. Correlations are shown only for bins with $\geq 30$ cases.

the simplification that only the intraorbit correlation is considered for calculation of the gridbox SST uncertainty. Taking into account the error correlation, the variance in $\sigma_{\overline{\mathrm{SST}}}$ becomes

$$\sigma_{\overline{\mathrm{SST}}}{}^2 = \frac{1}{n^2} \sum_{ij} r_{ij} \sigma_i \sigma_j \tag{6.4}$$

where $n$ is the number of observations in the gridbox.

An obvious caveat of this approach to estimating the correlation in the IRIS BTs comes from using the IRIS-THIR BT as a proxy for the IRIS BT error, given that there may be error correlations in the THIR BTs. As there are no other sources of concurrent data that can be used to estimate the IRIS error correlations, this caveat is unavoidable.

## 6.4 Comparison with historical in-situ SST datasets

In this section, maps of the monthly differences between the gridded IRIS SSTs and historical in-situ based SST datasets are presented. The IRIS SSTs represent skin SSTs, whereas the in-situ datasets represent bulk SSTs, so the IRIS SSTs were first converted to bulk SSTs by adding 0.17 K (*Donlon et al.*, 2002). In order to assess whether the IRIS SSTs and in-situ SSTs agree to within their respective uncertainties, we can compute the normalised distribution of the differences z:

$$z = \frac{\mathrm{SST_{IRIS}} - \mathrm{SST_h}}{\sqrt{\sigma^2_{\mathrm{IRIS}} + \sigma^2_{\mathrm{h}}}} \tag{6.5}$$

where $\mathrm{SST_{IRIS}}$ and $\mathrm{SST_h}$ are the gridded IRIS SSTs and gridded SSTs from the in-situ dataset and $\sigma_{\mathrm{IRIS}}$ and $\sigma_{\mathrm{h}}$ are the respective gridded uncertainty estimates. If there is agreement between the IRIS and in-situ datasets, we expect z to be Gaussian with a standard deviation close to 1. Out of the four SST datasets described in chapter 2, uncertainty estimates could only be obtained for HadISST2 and HadSST3 so the comparisons are carried out for these two datasets only.

The HadSST3 uncertainties are represented in an ensemble of 100 realisations of the dataset. These realisations explore uncertainties due to measurement error, sampling error and bias adjustments. The HadISST2 reconstruction is in the form of a 10-member ensemble, which explores the uncertainty range associated with data biases and analysis uncertainty (*Kennedy et al.*, 2019a). For both in-situ datasets, the ensemble

median was used as the central estimate to compare with the dataset of IRIS SSTs and the ensemble standard deviation was used to represent the uncertainty in the ensemble median.

### 6.4.1 Clear-sky results

#### 6.4.1.1 IDCS-derived cases

Figure 6.13 shows the monthly gridded IRIS-HadISST2 and IRIS-HadSST3 SSTs for the IDCS-derived clear-sky cases. In total, there are 39 fewer filled gridboxes in the IRIS-HadSST3 plot as HadSST3 is not spatially infilled. This means that there are some IRIS SSTs, mainly below 50°S, from the IDCS-derived cases in locations with no in-situ observations. As HadISST2 and HadSST3 both use the same in-situ measurements from version 2.5 of ICOADS, the main driver of differences between them is the difference in spatial infilling. This mainly affects less well-observed regions in the Southern Hemisphere. Differences between HadISST2 and HadSST3 are negligible in most parts of the Northern Hemisphere.

To assess the level of agreement between the IRIS SSTs and in-situ based SST datasets, histograms of z-values were plotted (Figure 6.14). Gaussian curves are also plotted to show the difference between the observed distribution and the distribution expected if there is perfect agreement. The standard deviations of the z-values are 1.35 and 1.44 for IRIS-HadISST2 and IRIS-HadSST3 respectively. The mean z-values are about 0.2 for both reconstructions, meaning that the set of IDCS-derived IRIS SSTs is slightly warmer on average than the reconstructions. These results suggest that, taking each set of cases as a whole, the IRIS and in-situ SSTs do not agree to within their stated uncertainties. To better understand the level of agreement, we can use quantile-quantile (Q-Q) plots to compare the observed distribution of z-values to the distribution expected if there is agreement (Figure 6.15). For both IRIS-HadISST2 and IRIS-HadSST3 the departures in the sample (observed) quantiles are negative on the left side and positive on the right side. This indicates that the tails in the observed

distributions are longer than expected.

The long negative tail may be a sign of some residual cloud contamination in the set of IDCS-derived cases, resulting in a cold bias in the retrieved SSTs in some cases. It is more to difficult to explain the cause of the long positive tail. Cloud contamination can only produce a warm bias in the retrieved SST when there is a temperature inversion, which is usually a rare occurrence.

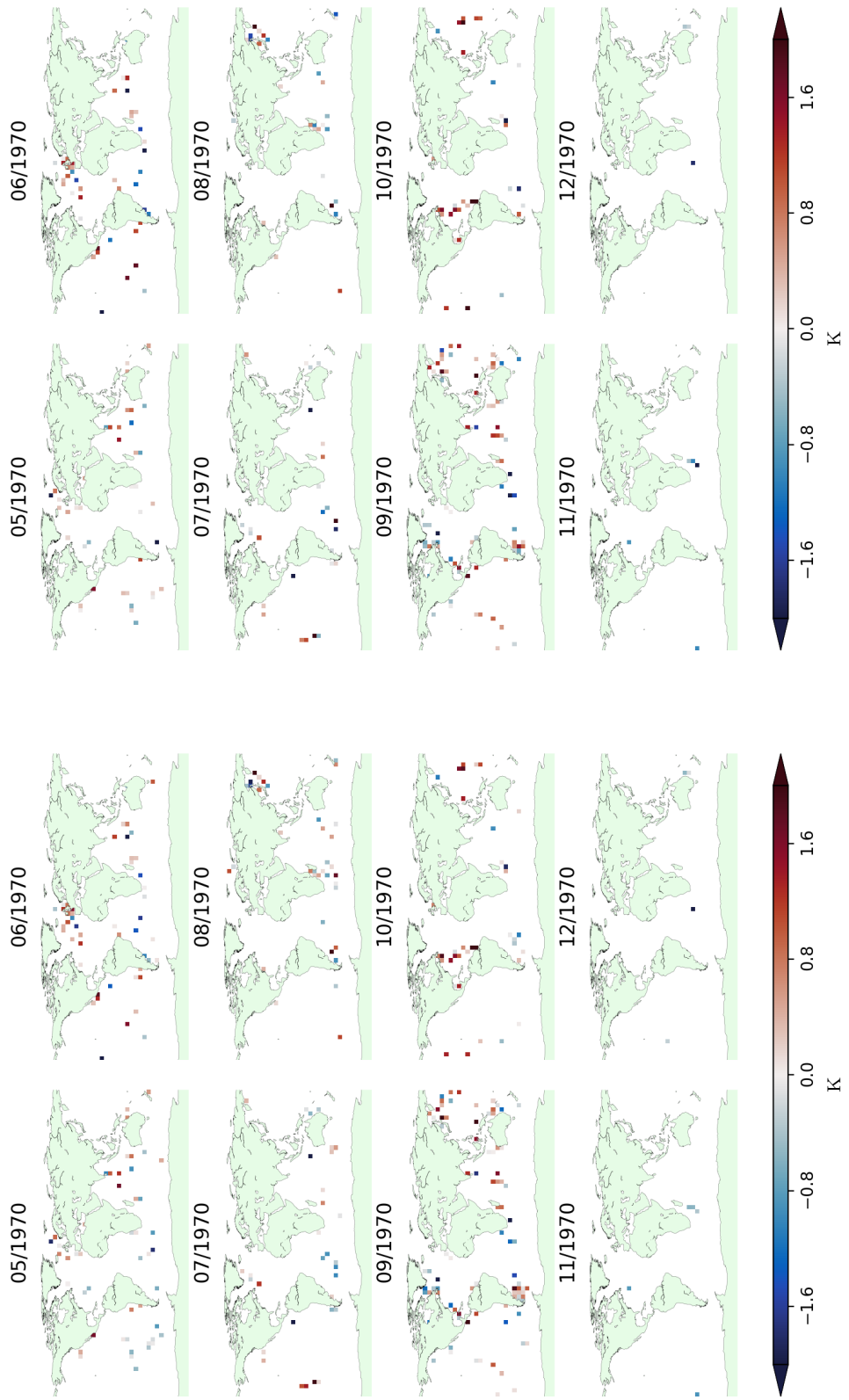



Figure 6.13: Monthly gridded IRIS minus HadISST2 (left) and HadSST3 (right) SST for the set of IDCS-derived clear-sky cases.

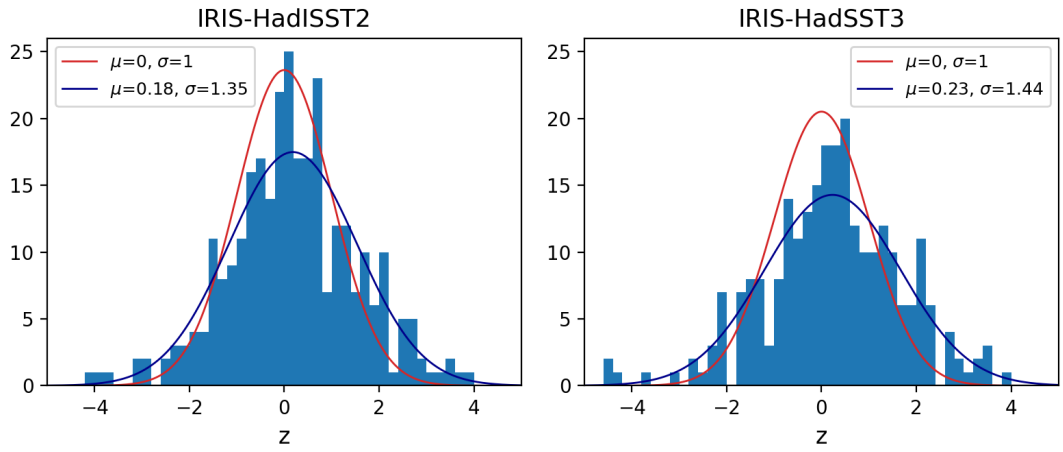



Figure 6.14: Histograms of z-values for the normalised IRIS-HadISST2 and IRIS-HadSST3 SST differences for the set of IDCS-derived clear-sky cases (blue). Gaussian curves are plotted for the case where there is perfect agreement between IRIS and the reconstructions (red) and from the observed data (dark blue).

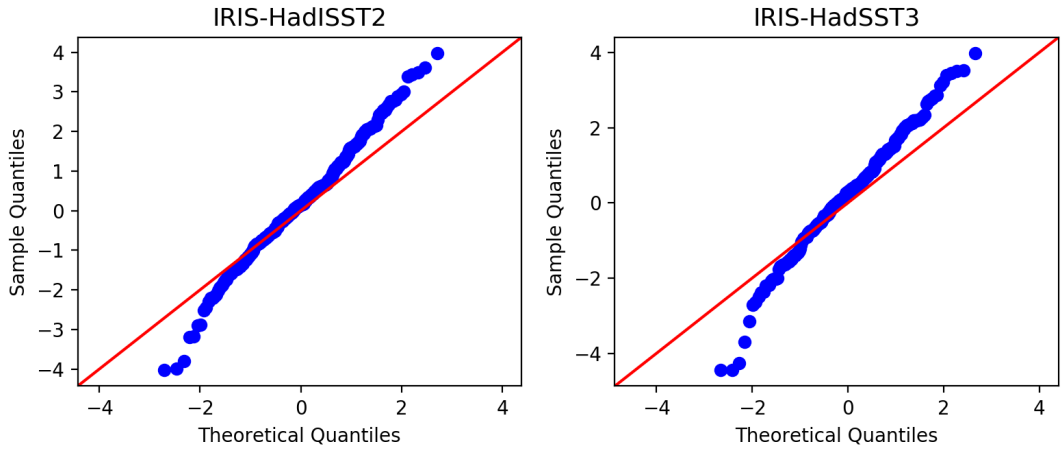



Figure 6.15: Q-Q plots comparing the quantiles of the observed z-values for the IRIS-HadISST2 and IRIS-HadSST3 differences with the theoretical quantiles of the z-values for the distributions expected if the IRIS and in-situ SSTs agree to within their stated uncertainties for the set of IDCS-derived cases (blue dots). The red diagonal line indicates the curve expected if the observed distribution matches the expected distribution.

#### 6.4.1.2 CERA-20C-derived cases

Figure 6.16 shows the monthly gridded IRIS-HadISST2 and IRIS-HadSST3 SSTs for the CERA-20C-derived clear-sky cases. There is only one less filled gridbox in the IRIS-HadSST3 plot, because this set of IRIS SSTs are almost all located in gridboxes

with $\geq 1$ in-situ observation. Figures 6.17 and 6.18 show histograms of the z-values and corresponding Q-Q plots for this set of cases. The mean z-values are slightly less for this set of IRIS SSTs and the standard deviations of the z-values are much closer to one, indicating that the IRIS and in-situ SSTs are more consistent for this set of cases. Figure 6.18 shows that the departures in the sample quantiles are much smaller for the negative quantiles for this set of cases, suggesting that there is likely to be very little cloud contamination in the CERA-20C-derived cases. However, Figure 6.18 shows that there is a small positive skew in the distributions of observed z-values.

In these comparisons, it is important to remember that the daytime IRIS SSTs all correspond to local solar noon when thermal stratification is near its maximum. In low wind speed conditions and high insolation, the skin-bulk difference can exceed +1 K (*Donlon et al.*, 2002); considerably different from the standard value of -0.17 K used to correct the IRIS skin SSTs. This may explain some of the variability seen in Figures 6.13 and 6.16.

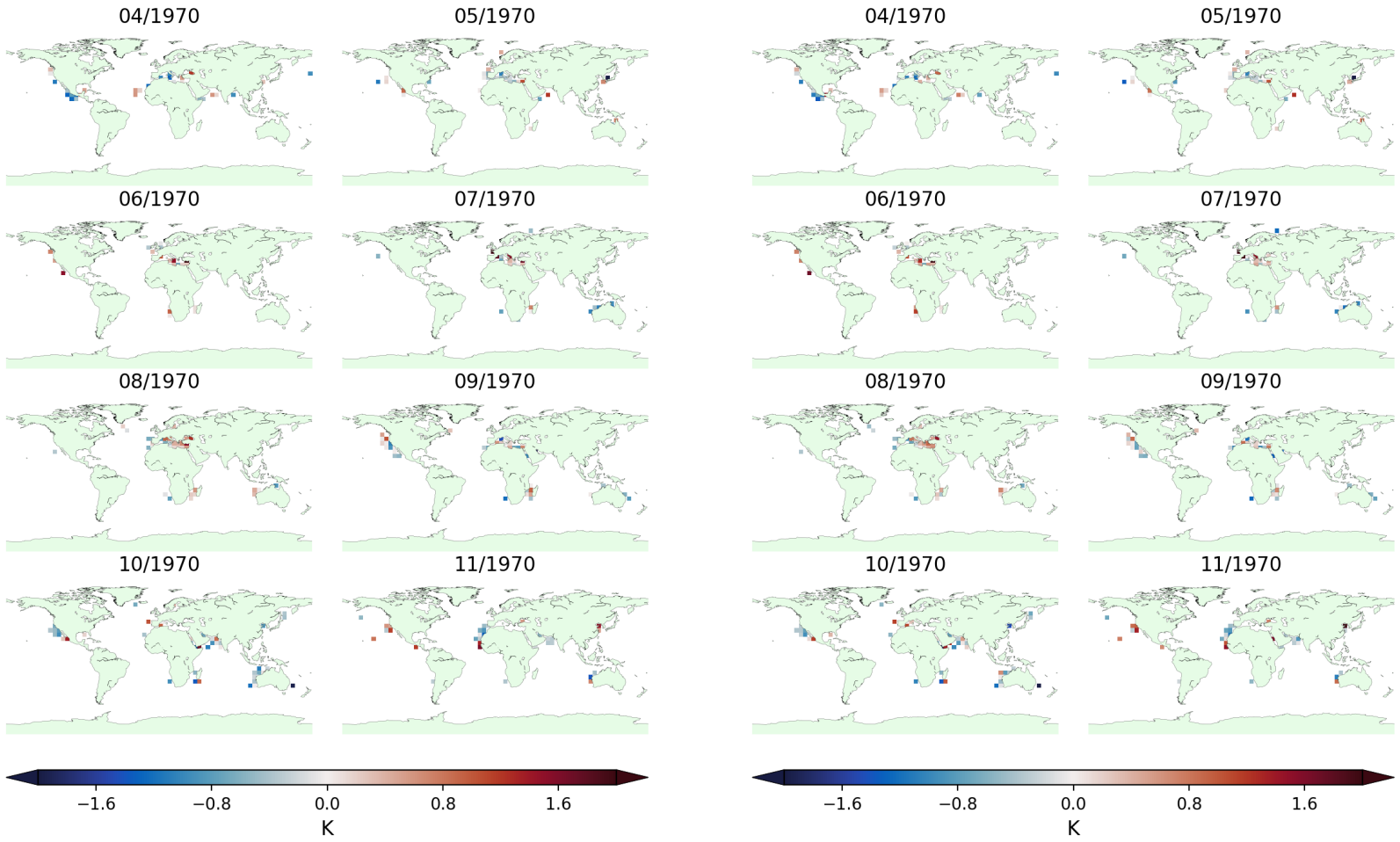



Figure 6.16: Monthly gridded IRIS minus HadISST2 (left) and HadSST3 (right) SST for the set of CERA-20C-derived clear-sky cases.

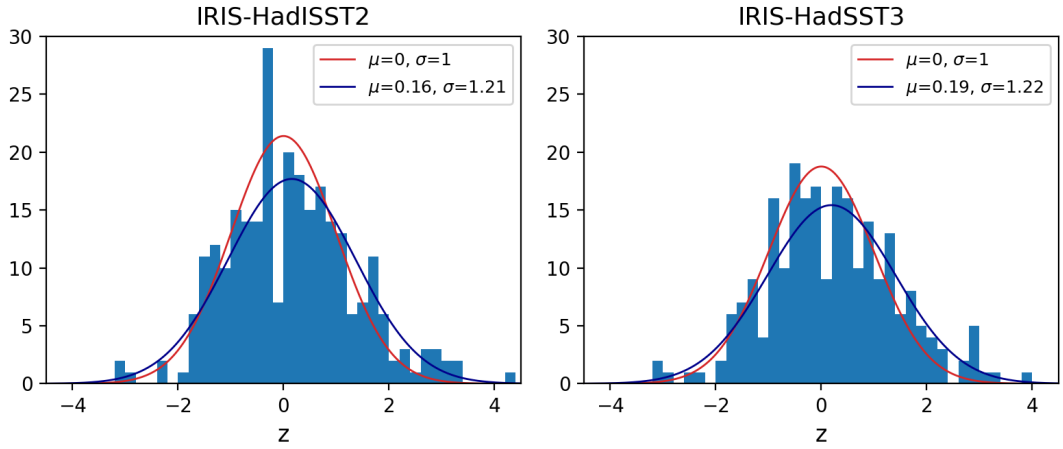



Figure 6.17: Histograms of z-values for the normalised IRIS-HadISST2 and IRIS-HadSST3 SST differences for the set of CERA-20C-derived clear-sky cases (blue). Gaussian curves are plotted for the case where there is perfect agreement between IRIS and the reconstructions (red) and from the observed data (dark blue).

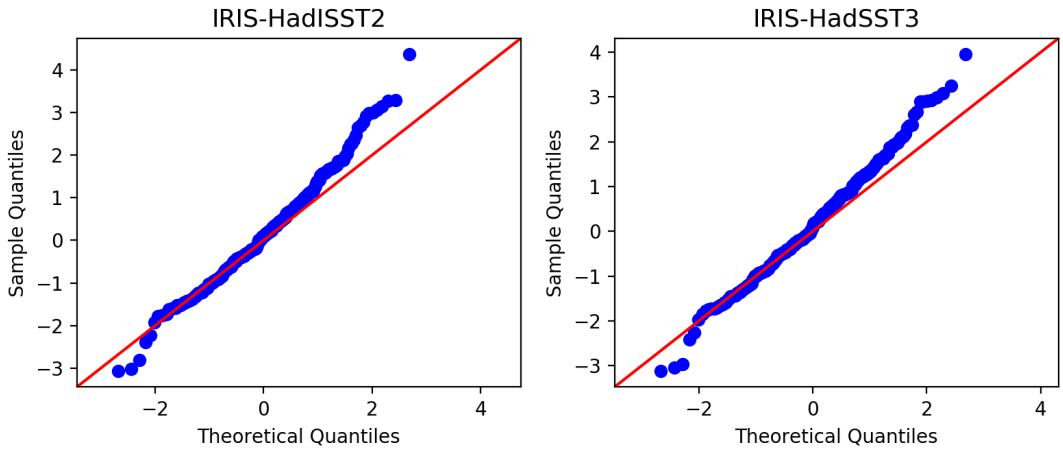



Figure 6.18: Q-Q plots comparing the quantiles of the observed z-values for the IRIS-HadISST2 and IRIS-HadSST3 differences with the theoretical quantiles of the z-values for the distributions expected if the IRIS and in-situ SSTs agree to within their stated uncertainties for the set of CERA-20C-derived cases (blue dots). The red diagonal line indicates the curve expected if the observed distribution matches the expected distribution.

### 6.4.2 Partly cloudy results

Figure 6.19 shows the monthly gridded IRIS minus HadISST2 and HadSST3 SST for the party cloudy cases. This shows a clear zonal bias in the SST differences with the IRIS

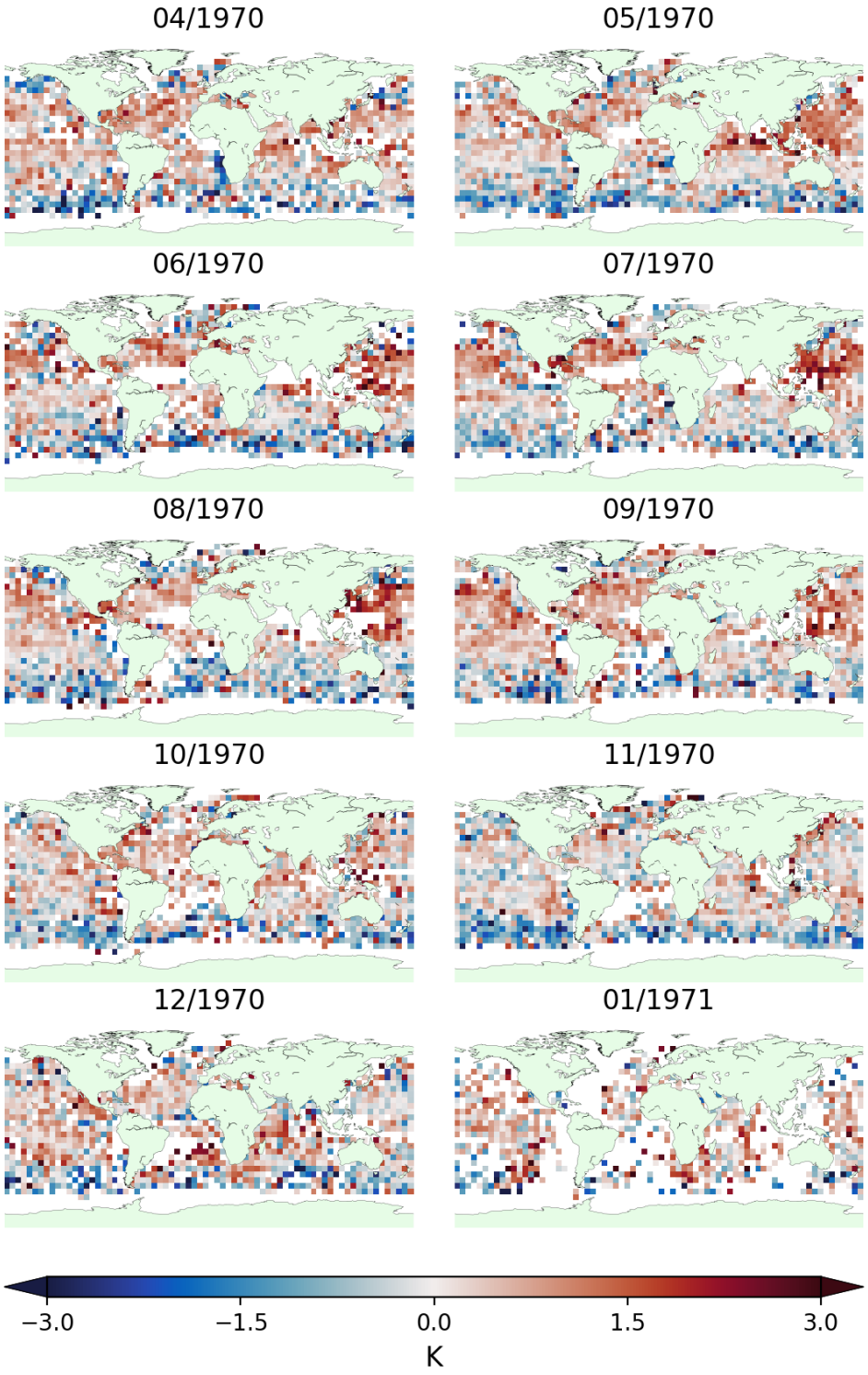
SSTs colder than the in-situ datasets at low latitudes in the Southern Hemisphere and generally warmer than the in-situ datasets in the Northern Hemisphere. Additionally, there are areas with pronounced positive/negative differences in SST of up to 2 or 3 K. There are also large month-to-month changes in the SST differences in some areas. For example, in a region of the western North Pacific, IRIS is 2-3 K warmer than the in-situ datasets in July-August but 0-1 K cooler than the in-situ datasets in November-December. A zonal bias was seen in the IRIS-like IASI SSTs for the set of partly cloudy cases, although it was smaller in magnitude than observed in the IRIS minus in-situ SSTs. A zonal bias in the retrieved IRIS-like IASI CF was found to be one of the main causes of the zonal SST bias, so it is likely that this is also a major cause of the zonal bias in the IRIS minus in-situ SSTs. As with the clear-sky SSTs, Figure 6.19 shows that HadISST2 and HadSST3 agree very closely in most parts of the Northern Hemisphere, where there is better coverage of in-situ measurements.

z-values were calculated to assess the level of agreement between the partly cloudy IRIS SSTs and in-situ based SSTs. Figure 6.20 shows the monthly gridded z-values for IRIS-HadISST2 and IRIS-HadSST3. The distribution of z-values strongly reflects the distribution of SST differences in Figure 6.19. In many regions, there are clusters of gridboxes with z-values greater than $\pm 2$; this indicates a lack of agreement between the IRIS and in-situ based SSTs and suggests that there are inter-gridbox error correlations in the IRIS and/or in-situ based SSTs. If there are zonal/regional biases in the IRIS SSTs rather than the in-situ based SSTs as suggested previously, these error correlations originate in the IRIS SSTs.

The corresponding Q-Q plots are shown for each month in Figure 6.21, for the IRIS-HadISST2 comparison only. This shows that for all ten months, the distributions of observed z-values have longer, broader tails than the expected (theoretical) distributions. There are some differences between the size of the departures in the observed-theoretical quantiles between the months, however, with the monthly standard deviation of z-values varying from 1.14-1.41. It should be noted that the standard deviation of 1.14 in January 1971 is an outlier, relatively speaking, although the IRIS SSTs are much less well-sampled in this month, with many missing gridboxes.

The overall conclusion from the results presented in this section is that there are large zonal/regional biases in the IRIS minus in-situ based SSTs, which are likely to originate largely from the IRIS SSTs. Despite this, there is evidence that there are large regions of ocean, mostly in lower latitudes of the Southern Hemisphere, where the IRIS and in-situ based SSTs are in broad agreement.

We now want to better understand the causes of the biases seen in Figure 6.19. One way of doing this is to perform sensitivity tests, to study the effect of filtering the set of cases used to construct the gridded IRIS SSTs on the gridded IRIS SSTs themselves.


04/1970
05/1970
06/1970
07/1970
08/1970
09/1970
10/1970
11/1970
12/1970
01/1971
−3.0
−1.5
0.0
1.5
3.0
K

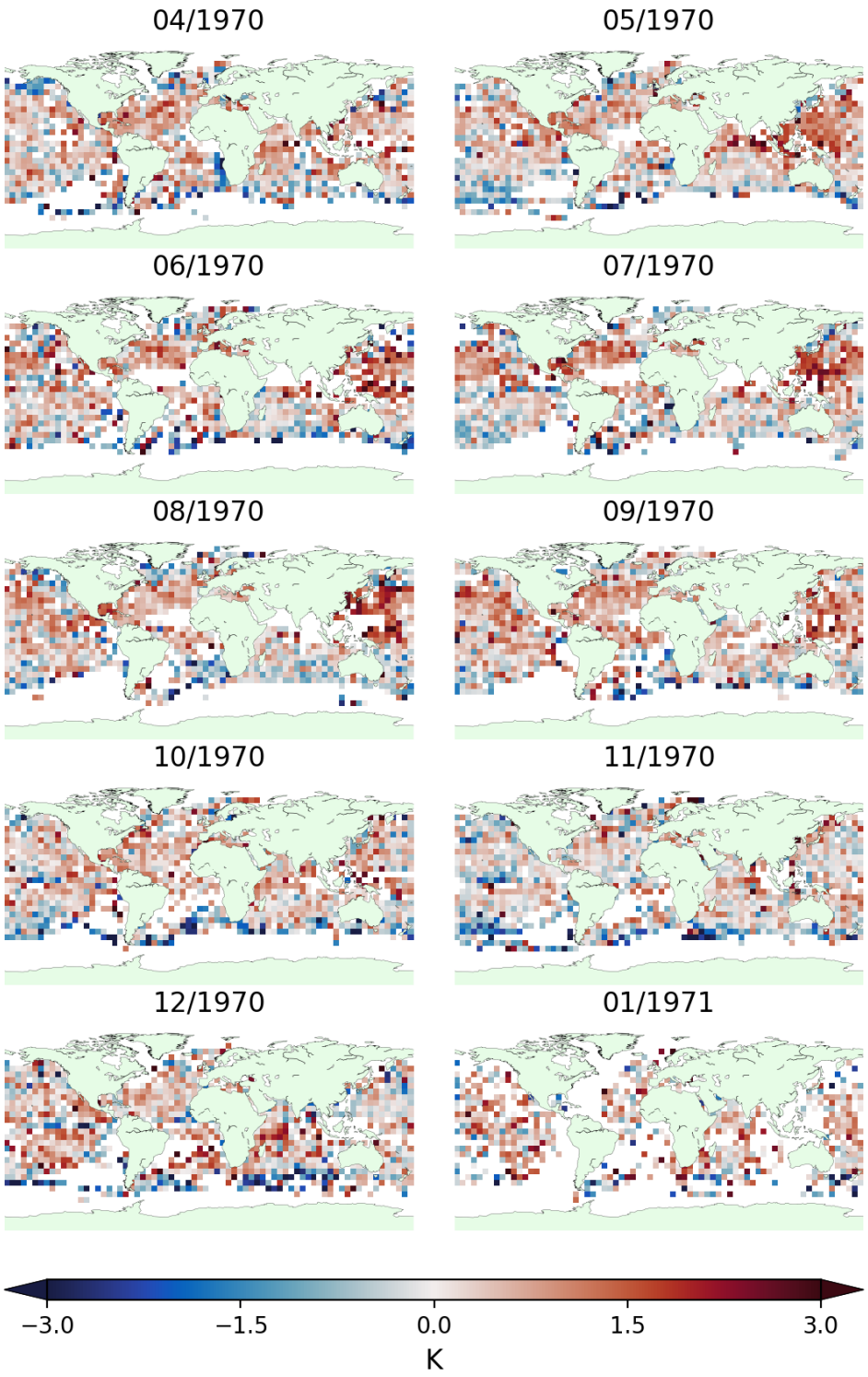
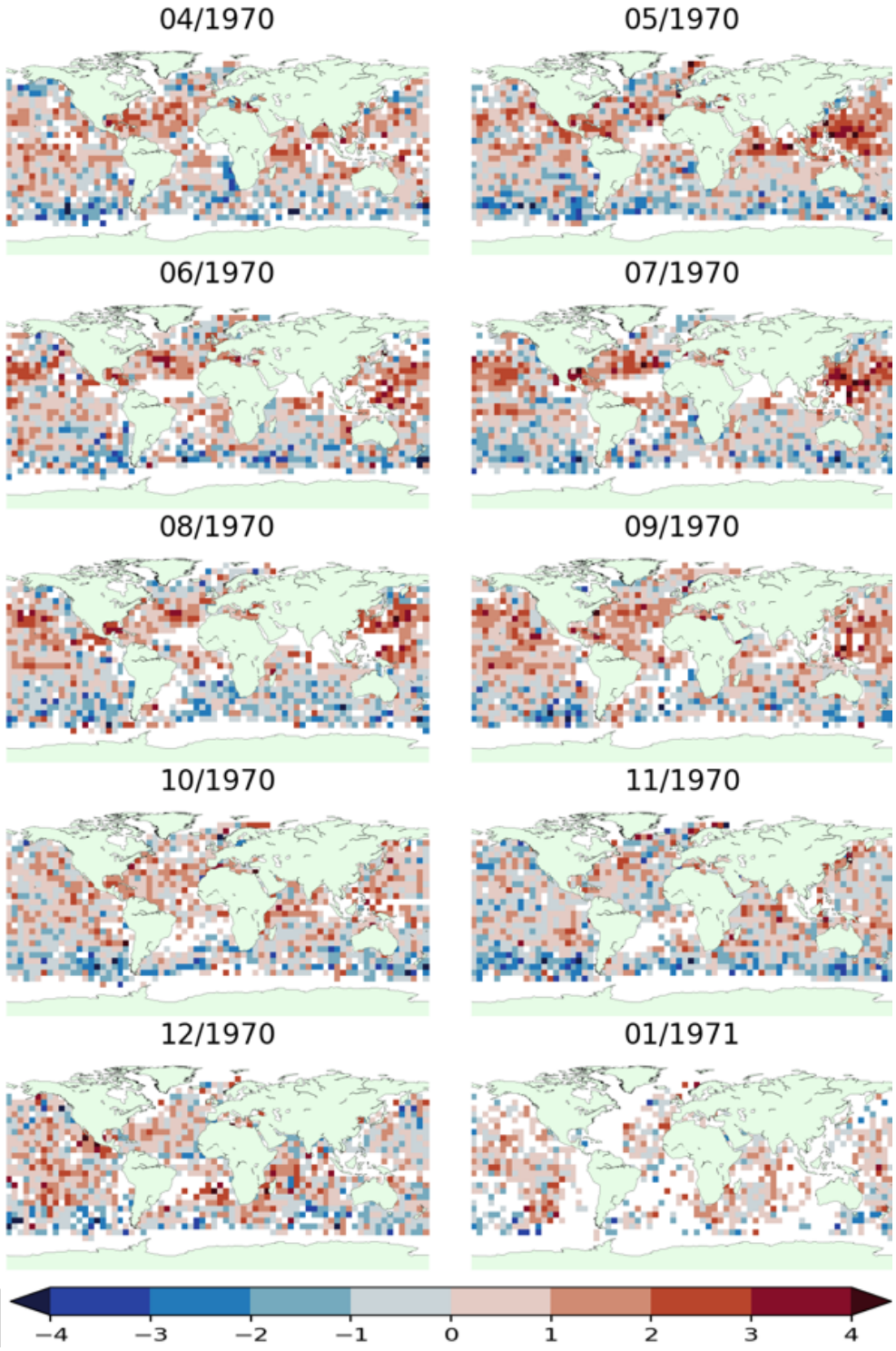



Figure 6.19: Monthly gridded IRIS minus HadISST2 (left) and HadSST3 (right) SST calculated from the partly cloudy IRIS SSTs.


04/1970
05/1970
06/1970
07/1970
08/1970
09/1970
10/1970
11/1970
12/1970
01/1971
−4
−3
−2
−1
0
1
2
3
4

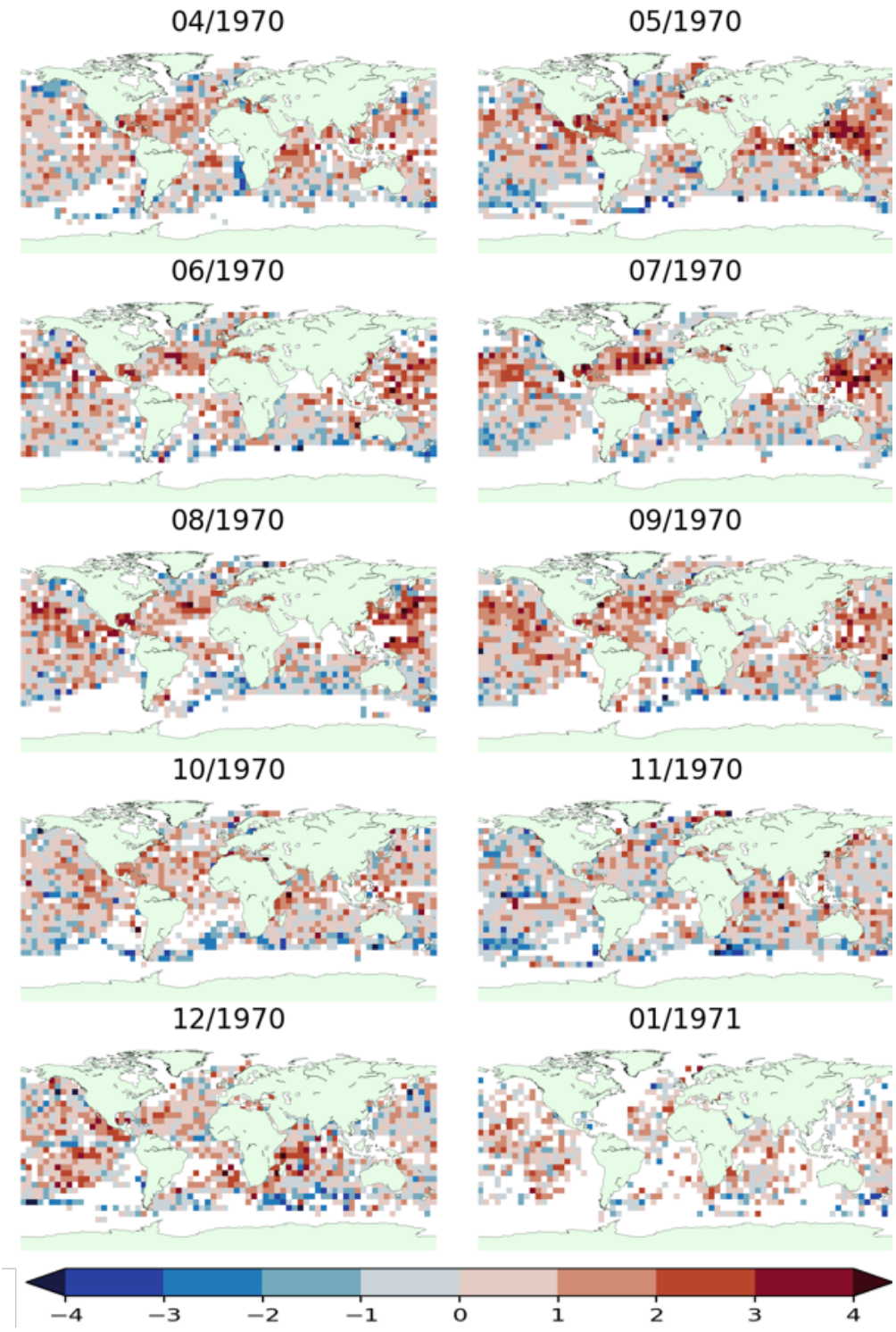



Figure 6.20: Monthly gridded z-values for IRIS minus HadISST2 (left) and IRIS minus HadSST3 (right) calculated for the set of partly cloudy IRIS SSTs.

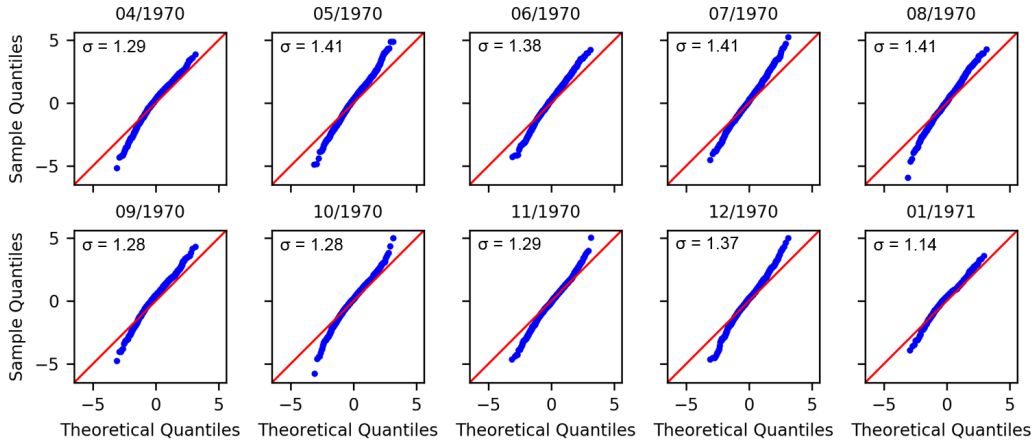
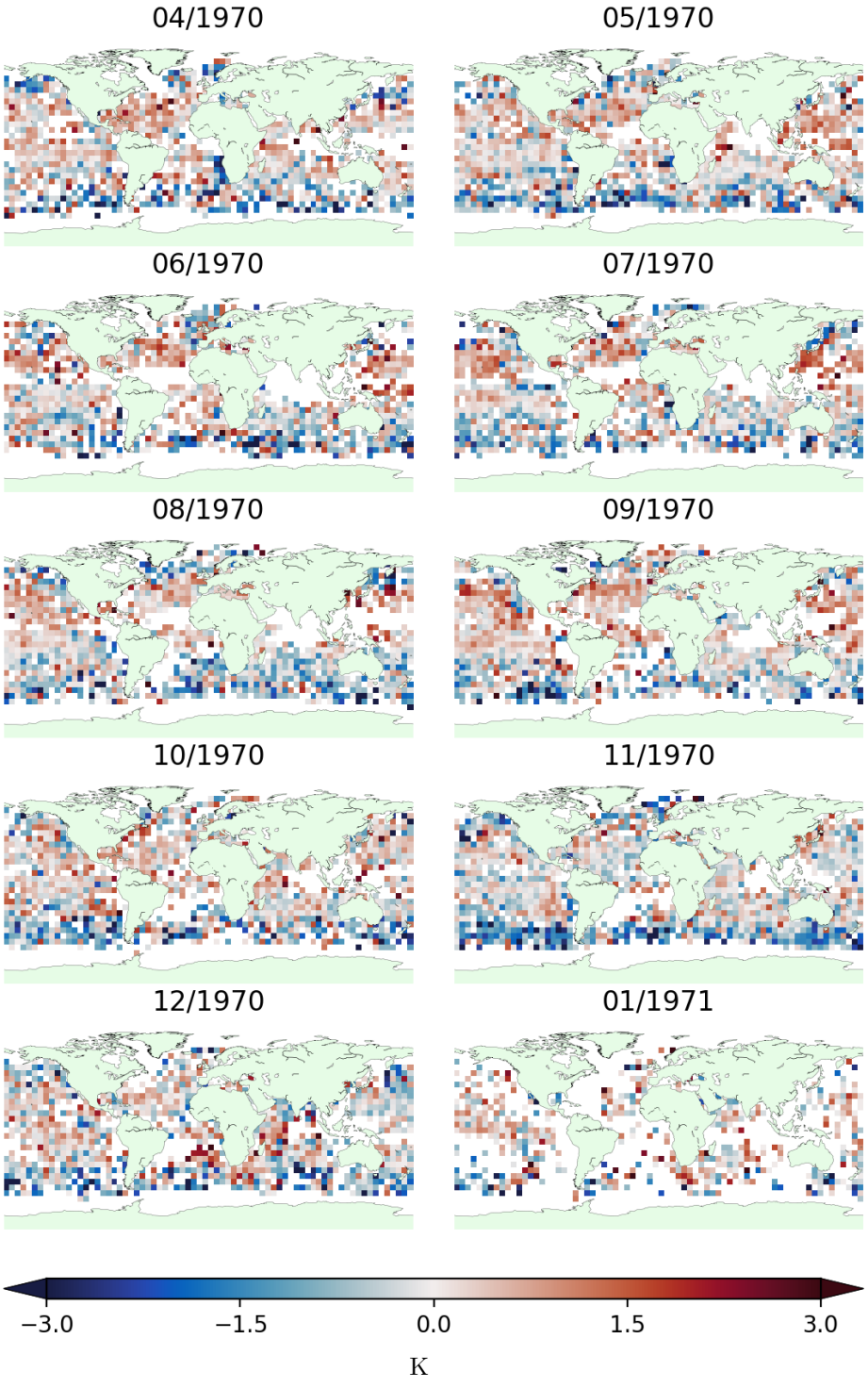



Figure 6.21: Monthly Q-Q plots comparing the quantiles of the observed z-values for the gridded, partly cloudy IRIS-HadISST2 differences with the theoretical quantiles of the z-values for the distributions expected if the IRIS and in-situ SSTs agree to within their stated uncertainties (blue dots). The red diagonal line indicates the curve expected if the observed distribution matches the expected distribution. Standard deviations of observed z-values are also shown.

#### 6.4.2.1 Cloud height

The effect of error in the retrieved CF on error in the retrieved SST is strongly dependent on cloud height. This is because the temperature difference between the cloud top and sea surface is roughly proportional to cloud height up to the tropopause (11 km for the International Standard Atmosphere (ISA)). Cloud height was assumed to be fixed at the NWP values in the retrieval. Errors in these cloud height values may be disguised as errors in the retrieved CFs e.g. a positive error in the cloud height could be disguised as a positive error in the cloud fraction and vice versa. By removing cases with cloud above a certain height when constructing the gridded IRIS SSTs, we might expect the errors in the IRIS SSTs to reduce. Cases with a cloud height above 800 hPa ($\approx$2 km) were discarded, accounting for $\approx$30 % of all cases. Figure 6.22 shows the monthly gridded IRIS-HadISST2 SSTs with these cases removed and also shows the difference between the original gridded IRIS SSTs and the IRIS SSTs with these cases removed.


04/1970
05/1970
06/1970
07/1970
08/1970
09/1970
10/1970
11/1970
12/1970
01/1971
−3.0
−1.5
0.0
1.5
3.0
K

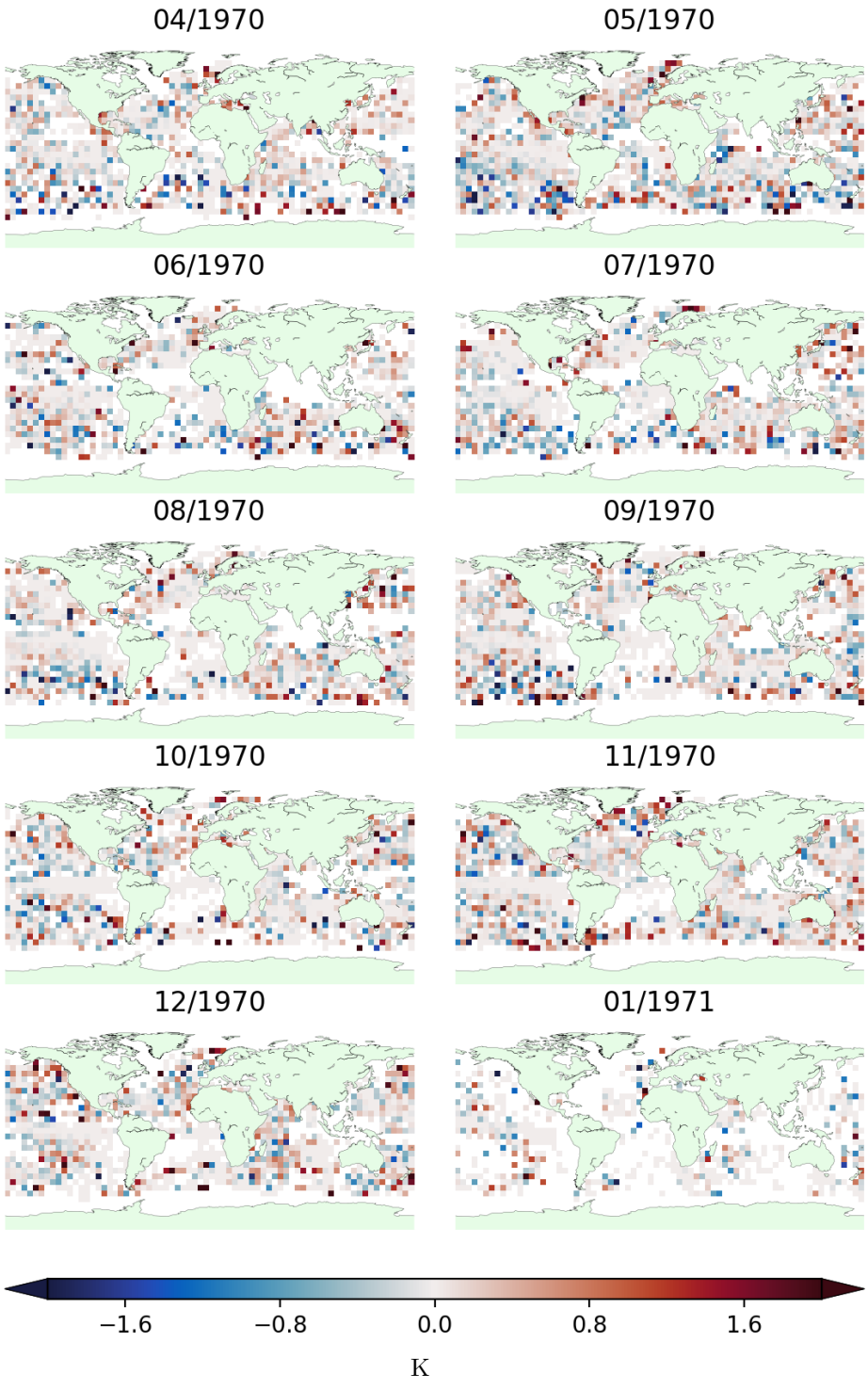



Figure 6.22: (a) Monthly gridded IRIS minus HadISST2 SST for cases where the cloud height is below 800 hPa. (b) original gridded IRIS SST minus (a)

The effect of restricting the IRIS SSTs to cases with cloud height below 800 hPa is relatively small in most areas. The differences are zero in many gridboxes near the equator, simply because all of the cases in these gridboxes have cloud heights below 800 hPa. The differences tend to be greater at higher latitudes, but the gridboxes with greater differences are typically those with a higher fraction of cases with a cloud height above 800 hPa. As the distribution of the differences is also fairly random, this suggests that the differences are largely due to sampling noise rather than a real geophysical signal.

## 6.5 Comparison of clear-sky and partly cloudy IRIS SSTs

It is worth comparing the clear-sky and partly cloudy gridded IRIS SSTs for gridboxes where both are available. This tells us about the level of agreement between the two different approaches to retrieving SST from IRIS. Figures 6.23 and 6.24 show histograms of the z-values and the corresponding Q-Q plots for the IDCS and CERA-20C clear-sky minus partly cloudy gridded SSTs. The standard deviations of the z-values are 1.03

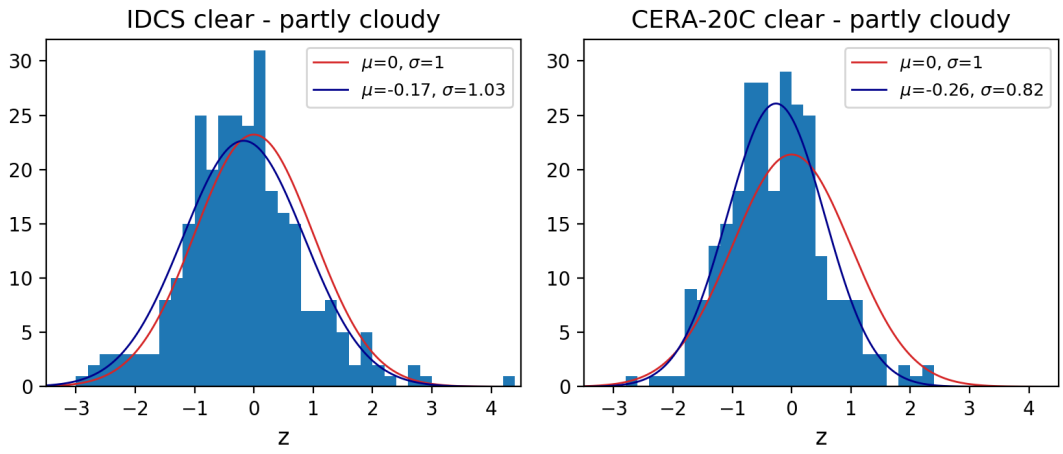



Figure 6.23: Histograms of the z-values calculated from the IDCS (left) and CERA-20C (right) clear-sky minus partly cloudy gridded IRIS SST differences. Gaussian curves are plotted for the case where there is perfect agreement between the clear-sky and partly cloudy IRIS SSTs (red) and from the observed data (dark blue).

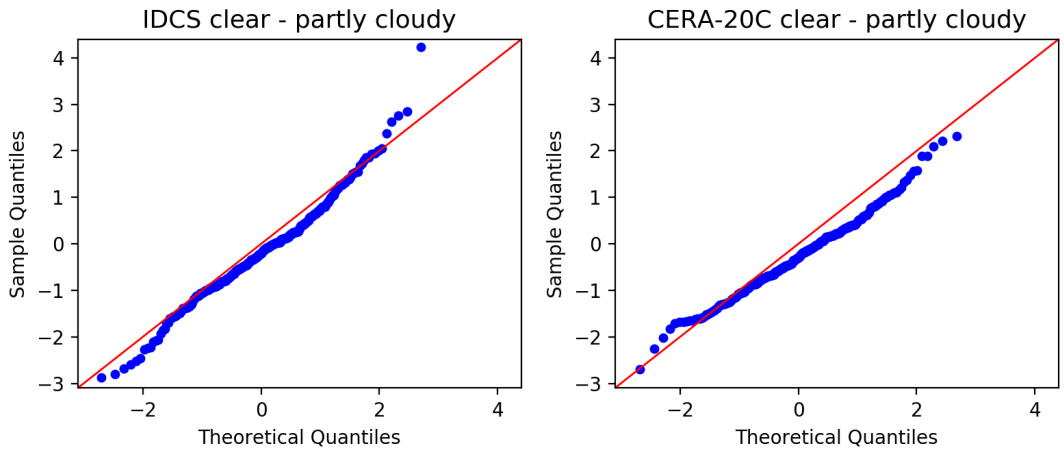



Figure 6.24: Q-Q plots of the z-values calculated from the IDCS (left) and CERA-20C (right) clear-sky minus partly cloudy gridded IRIS SST differences (blue dots). The red diagonal line indicates the curve expected if the observed distribution matches the expected distribution.

and 0.82 respectively, indicating that there is overall agreement between the gridded clear-sky and partly cloudy SSTs, to within their estimated uncertainties. The fact that the standard deviation is much smaller than 1 for the CERA-20C clear-sky minus partly cloudy gridded SSTs suggests that the uncertainty in the gridded CERA-20C-derived SSTs may be overestimated. The Q-Q plot for the IDCS based clear-sky SSTs shows close agreement between the observed and expected distribution of z-values. The Q-Q plot for the CERA-20C based clear-sky SSTs indicates that the tails of the observed distribution are shorter than expected. For both sets of clear-sky cases there is a small negative offset in the observed z-values. This reflects the fact that the IDCS and CERA-20C based clear-sky SSTs are on average 0.17 K and 0.26 K cooler than the partly cloudy SSTs, respectively. This difference may be due to cloud contamination in the clear-sky cases and/or a cloud-related warm bias in the partly cloudy SSTs.

## 6.6 Chapter summary

In this chapter, monthly, gridded SST estimates were obtained from sets of clear-sky and partly cloudy IRIS observations and compared to two of the historical in-situ based SST datasets: HadSST3 and HadISST2.

From the diagnostics presented in §6.2 there were no serious issues with the clear-sky retrievals, although evidence was found for a remaining bias of up to 0.3 K in the CERA-20C-derived set of cases, indicating that the bias correction applied to IRIS may not be appropriate in all cases. For the partly cloudy retrievals, the diagnostics were broadly consistent with those from the partly cloudy IRIS-like IASI retrievals.

In order to determine the level of agreement between the gridded IRIS SSTs and the in-situ based SSTs, normalised differences (z-values) were computing using the uncertainty estimates in the gridded SSTs. The gridded IRIS SST uncertainties were calculated by combining the uncertainties due to measurement and sampling error. The calculation of the measurement uncertainty is complicated by error correlations in the IRIS SSTs. An attempt was made to quantify these correlations by estimating the error correlations in the IRIS BTs. This was done by comparing IRIS and THIR window BTs and computing the correlation in the IRIS-THIR BT as a function of time interval. Correlations of +0.6 to 0.75 were found in the IRIS-THIR BTs for cases in the same orbit and within a 5° by 5° gridbox. A major caveat of this approach is that there may be different error correlations in the THIR BTs. The sampling uncertainty in the gridded IRIS SSTs was estimated empirically from 0.05° SST CCI L4 data covering the period 1982-2016. For the in-situ based SST datasets, the uncertainties are estimated from the ensembles of realisations which explore the range of measurement and sampling uncertainties and also uncertainties associated with bias adjustment.

For the comparison of clear-sky gridded IRIS SSTs with HadSST3 and HadISST2 the standard deviations of the z-values were greater than 1 for both sets of clear-sky cases. This implies that the gridded IRIS SSTs and in-situ based SSTs do not agree

to within their respective uncertainties. It is difficult to attribute the cause of this disagreement to either the IRIS SSTs or the in-situ based SSTs. Q-Q plots showed that the tails of the distributions of z-values are longer than expected, a potential sign of cloud contamination in the sets of clear-sky IRIS observations. It may also be the case that the uncertainties in the gridded IRIS SSTs and/or in-situ based SSTs are underestimated, resulting in a greater spread of z-values.

Large zonal and regional differences of up to $\approx \pm 2$K were seen in the partly cloudy IRIS minus in-situ based SSTs. As a zonal bias was observed in the partly cloudy IRIS-like IASI SSTs, it is likely that there are zonal/regional biases in the partly cloudy IRIS SSTs, causing these differences. This also means that the error correlations in the IRIS SSTs are likely to have much longer length scales than we have assumed in the calculation of the gridded uncertainties, with error correlations likely between gridboxes. The standard deviations of the z-values were greater than 1 for each month, indicating that there is not agreement between the partly cloudy IRIS SSTs and in-situ based SSTs, when taking the differences from each month as a whole. Our results suggest, however, that the IRIS and in-situ based SSTs may be in broad agreement in some ocean regions.

A sensitivity test was carried out to investigate if the biases in the IRIS-HadISST2 SSTs are sensitive to cloud height. It was found that restricting the IRIS SSTs to cases with cloud height below 800 hPa has no real geophysical effect on the relative IRIS-HadISST2 biases.

As a final assessment of the quality of the IRIS SSTs, the clear-sky and partly cloudy gridded IRIS SSTs were compared for gridboxes where both were available. This showed that both sets of clear-sky SSTs agreed with the partly cloudy SSTs to within their estimated uncertainties. However, the clear-sky SSTs were found to be on average about 0.2 K cooler than the partly cloudy SSTs.

The results from this chapter will be discussed further in the next chapter and suggestions will be made of further work that could potentially improve our results.

# CHAPTER 7

# DISCUSSION AND CONCLUSIONS

Satellites have provided a continuous source of SST measurements to supplement in-situ based SST records since 1981 (*O'Carroll et al.*, 2019). The inclusion of satellite-based SSTs is vital for achieving a more robust and globally complete SST record. The derivation of early satellite-based SST datasets is therefore of immense value. The primary aim of this thesis has been to investigate the degree to which useful estimates of SST can be obtained from a satellite sensor, predating the established satellite record, intending to assess the reliability of the historical SST reconstructions in the pre-AVHRR era.

In the literature, there are a few examples of past attempts to retrieve SST from IRIS. These studies are limited in terms of their spatial scale and the methods used for cloud detection and SST retrieval. There have been major improvements in NWP and radiative transfer modelling since these past studies, enabling more accurate simulation of IRIS spectra. Additionally, high resolution measurements from hyperspectral instruments like IASI are available for indirectly testing the SST retrieval scheme.

In chapter 3 we assessed the quality of the IRIS dataset and removed erroneous spectra. Chapter 4 then dealt with a spectral bias discovered in the IRIS measurements. It was crucial to quantify and remove this bias in order to minimise bias in the SSTs

retrieved from IRIS. In chapter 5, we tested an OE-based retrieval scheme on IRIS-like IASI spectra for retrieving SST in clear-sky and partly cloudy conditions. Finally, in chapter 6 we used the retrieval scheme to retrieve SSTs from IRIS spectra known a priori to be cloud free and from a much larger set of potentially cloud-affected spectra with near-global coverage. A gridded, monthly dataset of SST was derived for each set of cases. Comparisons were then made between the dataset of IRIS SSTs and two of the historical in-situ based SST datasets, to assess their level of agreement. In sections 7.1-7.4, we discuss our findings from the previous chapters in turn, starting with chapter 3. We then make suggestions of further work that could build upon this thesis in section 7.5. Finally, section 7.6 summarises our overall conclusions.

## 7.1 Quality control of the IRIS data

In chapter 3, a wide variety of QC tests were applied to remove erroneous IRIS spectra. The first three stages of QC were reproduced from *Poli and Brunel* (2016), but an additional principal component-based test was applied to remove further anomalous spectra. After application of the QC, there were some remaining spectra with unphysically low BTs at wavenumbers beyond 1400 $\mathrm{cm}^{-1}$. This is not an issue for SST retrieval, as SST cannot be retrieved from these very cloudy scenes.

The QC'd spectra were assessed by comparison with all-sky simulations. The IRIS observations were found to be on average 3-4 K colder than the simulations across the window regions. A large part of this bias is probably due to systematic under-representation of cloud in ERA-20C and the other main contributor is the bias in the IRIS observations; the subject of chapter 4.

## 7.2 The IRIS calibration bias

Evidence for a bias in the IRIS spectra was shown in *Bantges et al.* (2016), where the global mean BT spectrum for IRIS was compared with global mean IASI and AIRS

spectra. IRIS was found to have a spectrally varying cold bias of up to 2 K in the 800-1000 $cm^{-1}$ window relative to IASI and AIRS. However, this is not a direct comparison as these three sensors operated in different time periods.

To further quantify the spectral bias, we compared observed and simulated IRIS spectra for clear-sky cases only. This required a method of cloud detection to select clear-sky IRIS observations. To avoid the bias in the IRIS observations from influencing the cloud detection, visible imagery from the IDCS was used. Unfortunately, only a small sample of a few hundred clear-sky observations were identified using this method, due to geolocation error in the IDCS data, which could only be partially corrected. Comparing the observed and simulated spectra for this set of cases led us to discover the inability of the simulations to capture certain instrument effects. These instrument effects result in a spectral shift and a discrepancy in the spectral resolution between the observations and simulations. For the spectral shift we applied the correction determined empirically in *Harries et al.* (2001). We estimated our own empirical correction for the spectral resolution. Having applied these corrections to the simulations, the median observed-simulated radiance difference formed our preliminary estimate of the IRIS bias. Our preliminary bias model makes the assumption that the spectral radiance bias is the same for all IRIS observations, which follows from the hypothesis made by *Bantges et al.* (2016) regarding the origin of the bias.

Our spectral bias estimate was found to agree closely with the IRIS-IMG and IRIS-IASI spectral differences from *Bantges et al.* (2016) in all spectral regions relevant to SST retrieval, except for the $CO_2$ absorption band. The discrepancy in the $CO_2$ band is explained by the increase in global mean $CO_2$ concentration between 1970 (IRIS) and 2012 (IASI) and so is not a deficiency in the simulations.

THIR BTs were used to provide an independent assessment of our bias model in the 11.5 µm window. Figure 4.13 shows that the relative bias in the IRIS-THIR BT reduces from about 2 K to 0.5 K for cases with a scene temperature in the range 270-290 K after applying our preliminary bias correction to IRIS. However, the bias varies

strongly with scene temperature above about 292 K and this variation is not reduced after applying the bias correction. Like IRIS, THIR is an old instrument which may have calibration issues of its own and so it cannot be treated as 'truth' without other evidence. We are therefore unable to attribute the remaining bias to either IRIS or THIR. Further analysis of the IRIS-THIR BTs showed that there is no statistically significant drift in the relative bias during the IRIS period, but the relative bias was found to vary with orbital position. We hypothesised that the orbital bias could be due to orbital variations in the temperatures of the IRIS and/or THIR instrument components. However, analysis of the relationship between the IRIS-THIR BTs and instrument temperatures did not support this hypothesis.

The comparison of IRIS and THIR BTs provides support that our IRIS bias estimate has the correct magnitude, although it does not clearly support the assumption made about the constancy of the bias expressed as radiance. We have found that the bias is likely to be complex, with dependencies on factors such as orbital position. However, as we are unable to attribute these additional bias terms to either IRIS or THIR, these results cannot be used to refine our bias model.

Nonetheless, our work here is original in that we have exploited a combination of forward model simulations and data from other sensors onboard Nimbus-4 to constrain the IRIS bias. In section 7.5 we provide suggestions of how our bias model may be improved.

## 7.3 Developing an SST retrieval method for IRIS

In chapter 5, we developed an OE-based scheme to retrieve SST from IRIS observations identified as clear-sky and partly cloudy. Certain choices had to be made when developing this scheme. For example, two ways of estimating the covariance in the a priori state were considered. The more computationally expensive approach is to calculate the covariance directly for each case from an ensemble of state vectors. The

issue with this approach is that unless a large ensemble is used, there can be non-geophysical structure in the covariances. Instead, we used a second approach adopted by *Irion et al.* (2018) where atmospheric temperature and humidity are assumed to be correlated over characteristic vertical length scales.

There is no direct means of testing the SST retrieval scheme on IRIS because of lack of independent data; other than the in-situ data which we aim to compare with. As a means of testing the retrieval design, we generated IRIS-like IASI spectra, imposing the same error characteristics as IRIS. The IRIS retrieval scheme was used to retrieve SST from these IRIS-like spectra and quality controlled drifting buoy SSTs were used to validate the retrieved SSTs. Our results show that SST can be retrieved to an accuracy of 0.4-0.5 K from spectra known a priori to be cloud free. For comparison, we repeated the validation for SSTs retrieved from raw clear-sky IASI spectra using the same retrieval method. The SSTs were retrieved to a similar level of accuracy as the IRIS-like IASI SSTs, showing that the effect of degrading the spatial and spectral resolution of the observations and adding IRIS-like errors to the observations has little effect on the retrieval accuracy. The estimated clear-sky SST uncertainty of 0.4-0.5 K compares favourably with the standard deviation of 0.40 K in the IASI minus drifting buoy SSTs from *O'Carroll et al.* (2012), where an EOF regression was used to retrieve SST from IASI.

From the validation of the partly cloudy SST retrieval technique, we found that the inclusion of cloud fraction as an additional parameter to be retrieved greatly increases the uncertainty in the retrieved SST, even for cases where the retrieved CF is close to zero. This is because there is uncertainty in the retrieved CF, which translates to additional uncertainty in the SST. The bias in the IRIS-like IASI SSTs varied greatly with retrieved CF, increasing from -0.49 K for the clearest cases to +0.28 K for cases with retrieved CF $\geq$ 20 %. The errors in the retrieved SST and CF were also found to be correlated (r = 0.47). This makes physical sense, because a positive (or negative) error in the SST can be compensated by a positive (or negative) error in the CF. A

zonal pattern was found in the CF errors, with a similar zonal pattern in the SST errors, albeit more noisy. These results suggest that cloud-related SST biases are to be expected when retrieving SST from partly cloudy IRIS observations. The cloud biases are a possible indication that there is insufficient information in the observations to accurately retrieve the atmospheric state and SST simultaneously.

We investigated whether it may be possible to correct these SST biases using a simple multiple regression model to express the relationship between the SST errors and three predictors: retrieved SST, CF and TCWV. The CF error was not used as a predictor because this is unknown when retrieving SST from IRIS. We found that this model is able to explain only 24 % of the total variance in the SST errors. When this model was applied to the IRIS-like IASI SSTs, it was found to overcorrect the biases in most regions. We concluded that the SST biases cannot be corrected using a simple empirical model.

A limitation of validating the retrieval scheme indirectly is that it is difficult to fully simulate the error characteristics of IRIS. In particular, any remaining bias in the observations is not represented in the validation. There are also likely to be differences between the quality of the NWP fields used to derive the a priori state in the IRIS period and the modern validation period. This should have minimal effect on the clear-sky retrievals, which have very low sensitivity to the a priori state. However, it may have a greater effect on the partly cloudy retrievals of SST and CF, which are typically much more sensitive to the a priori SST and CF.

We have used a novel approach of exploiting data from a modern hyperspectral sensor, to test the IRIS retrieval scheme. This has enabled us to understand the quality of the SSTs that can be retrieved from IRIS and provides an insight into the potential sources of error. It also provides a means of validating the method for selecting partly cloud-affected IRIS observations.

## 7.4 Retrieving sea surface temperature from IRIS

The next step is to apply the retrieval scheme formulated in chapter 5 to retrieve SST from IRIS. The clear-sky retrieval technique is applied to two small sets of IRIS spectra identified as being cloud-free. The partly cloudy retrieval technique is applied to a much larger set of IRIS spectra with near-global coverage obtained using the cloud fraction threshold test from chapter 5.

The retrieval diagnostics ($\mathbf{K}\hat{\mathbf{x}} - \mathbf{y}$) indicate potential remaining wavelength dependent bias of up to 0.3 K and 0.2 K in the set of CERA-20C-derived clear-sky spectra and partly cloudy spectra, respectively. These biases are relatively small, but indicate that our bias model is not fully applicable to all IRIS spectra, as suggested from our results in chapter 4.

To compare the IRIS SSTs with the historical in-situ based SST datasets, they were gridded and estimates were made of the gridded uncertainties. Intraorbital correlations in the SST errors were accounted for in the calculation of the gridded uncertainties. These correlations were estimated based on the assumption that the errors in the IRIS SSTs are correlated over the same spatiotemporal scales as the errors in the IRIS window channel BTs. This is a fair assumption for the clear-sky SSTs because they are sensitive mainly to the observed window channel BTs. This assumption does not necessarily hold for the partly cloudy SSTs, where the SST errors are correlated with errors in the cloud fraction. As the errors in the IRIS BTs are unknown, the error correlations were calculated from a proxy most likely to have appropriate error characteristics, the IRIS-THIR window BT differences. This has the obvious caveat that there may be different error correlations in the THIR BTs, but this is the best available choice.

To fulfill the final objective of this thesis, we computed the normalised differences between the gridded IRIS SSTs and in-situ based SSTs. For the two datasets of clear-sky gridded SSTs, we find a lack of overall agreement between IRIS and HadISST2

and between IRIS and HadSST3. The two possible explanations for this are that the uncertainties in HadSST3 and HadISST2 are underestimated or the uncertainties in the IRIS SSTs are underestimated (or a combination of the two). Uncertainties of 0.46 K are assumed for each IRIS SST retrieved using the clear-sky retrieval technique, derived from the IRIS-like IASI validation statistics. This uncertainty assumes that the level of cloud contamination is no more than 2 %. If the level of cloud contamination in the IRIS observations is in fact higher, the uncertainty in the IRIS SSTs may be underestimated.

Greater discrepancies are found for the IDCS-derived cases than the CERA-20C-derived cases. This may be because of the difference between the approaches used to filter each set of SSTs; stricter filtering was applied to the CERA-20C-derived SSTs.

The comparison of the partly cloudy gridded SSTs with HadISST2 and HadSST3 shows a similar lack of overall agreement. There are clear spatial correlations in the gridded IRIS minus in-situ SSTs with a zonal trend in the differences. Since a zonal bias was seen in the partly cloudy IRIS-like IASI SSTs, we expect this zonal trend to be attributable to biases in the IRIS SSTs rather than HadISST2/HadSST3. There are therefore likely to be intergridbox error correlations in the IRIS SSTs. The limitation of using the overall distribution of z-values to test the IRIS vs in-situ agreement is that it hides these intergridbox error correlations. From the maps of monthly gridded z-values, it appears that there are certain regions where the IRIS and in-situ based SSTs agree more closely, although further analysis would be needed to quantify the level of agreement in these regions.

We investigated the sensitivity of the IRIS-HadISST2 differences to the a priori cloud height. We found that restricting the IRIS SSTs used in the construction of the gridded dataset to cases with a cloud top height below 2 km had very little physical effect on the IRIS-HadISST2 differences.

As a final test of the IRIS SSTs, we compared the clear-sky and partly cloudy gridded IRIS SSTs for gridboxes where both are available. The aim of this test was

to assess the level of agreement between the two different techniques of retrieving SST from IRIS. From the distribution of z-values, we found that the clear-sky and partly cloudy SSTs agree to within their uncertainties, with the CERA-20C-derived SSTs in fact being too consistent with the partly cloudy SSTs given their respective uncertainties. The partly cloudy SSTs have an overall warm bias of $\approx 0.2\,\mathrm{K}$ relative to the clear-sky SSTs. This may be due to a combination of potential cloud contamination in the clear-sky SSTs and a cloud-related warm bias in the partly cloudy SSTs.

In summary, the clear-sky IRIS SSTs have poor spatial coverage but nonetheless provide a potential indication that the uncertainties in HadISST2 and HadSST3 are underestimated. We have established that the SSTs retrieved from IRIS are not of sufficient quality for a robust assessment of the SST reconstructions, due to cloud related biases in the partly cloudy SSTs.

## 7.5 Future work

In this section we discuss some ideas for further work that may more fully address the original aim of the research. The key to this is retrieving better quality SSTs from IRIS.

One way of achieving this is to identify a much larger number of clear-sky IRIS spectra, because far more accurate (and less biased) SSTs can be retrieved using the clear-sky retrieval technique. An approach that can be taken further is to maximise the benefit from the other sensors onboard Nimbus-4. It may be possible to use data from the THIR to obtain a larger set of clear-sky IRIS observations. This would probably require the use of both the window and water vapour channels, however, and since both channels are available for only 9 % of IRIS observations, this is unlikely to provide many clear-sky observations. If the IDCS data could be fully geocorrected, this would result in a several-fold increase in the number of clear-sky spectra identified. However, this is very challenging because the direction and magnitude of the geolocation error

varies within a single IDCS image. However, it may be possible develop a physical model for the geolocation error, to partially correct the 70 % of the IDCS data where collocated THIR data is unavailable.

Since some SST errors are cloud related, a more sophisticated cloud detection method may be useful. It may be worth testing the cloud detection method of *McNally and Watts* (2003) on the bias-corrected IRIS spectra. *Poli and Brunel* (2016) identified relatively large numbers of clear-sky IRIS spectra using this method. However, its success depends heavily on the quality of the observations; the level of radiometric noise and remaining spectral bias may prevent this from being a reliable method of cloud detection. This method can be validated to some extent against the IDCS-based cloud detection.

We can also try to obtain better quality SSTs from partly cloud-affected observations. This requires further development of the cloudy retrieval technique on the set of partly cloudy IRIS-like IASI spectra, with the aim of reducing the biases in the IRIS-like IASI SSTs. This can be achieved by testing the sensitivity of the retrieved SSTs to changes in the retrieval setup. Some ideas of potential modifications and sensitivity experiments are as follows:

- Retrieve cloud top pressure simultaneously with cloud fraction and SST. This may require the use of more sounding channels from the $CO_2$ band.
- Vary a priori uncertainty in the CF. For example, we could adopt the approach in *Pavelin et al.* (2008), where no prior cloud information is provided in the 1D-Var retrieval of cloud parameters, meaning that the background-error covariances for the cloud parameters are assumed to be very large.
- Choice of NWP for simulations: newly available ERA5 may provide more realistic cloud information, which would help the retrieval to converge to the correct solution.
- Averaging kernels of SST and CF: accuracy of SSTs may improve for cases with

higher A(SST) and/or A(CF).

- Vary CF threshold used to select partly cloudy cases.
- The decision to equally weight the SST and CF uncertainties when selecting the partly cloudy channels was a pragmatic choice that was not fully explored; it would be interesting to test the effect of varying the weighting on the channel selection and then on the retrieval.

If we can better characterise the bias in the IRIS observations, this would contribute to reducing the level of bias in the SSTs. Identifying a much larger set of clear-sky IRIS spectra would help in characterising the dependence of the bias on parameters such as scene temperature and orbital position.

There is an ongoing Copernicus Climate Change Service (C3S) project which aims to recover and assess early satellite data records, including IRIS, for use in future numerical weather prediction. The work on IRIS aims to more fully quantify instrument effects, such as self-apodisation, resulting in an improved forward model. This work may also help to develop an improved bias model. We expect these improvements to result in the retrieval of better quality SSTs from IRIS.

We can potentially improve estimation of bulk SSTs from the IRIS skin SSTs by using a more sophisticated model of the skin effect, rather than assuming a constant bulk-skin offset. Without accurate NWP wind speed data this will be difficult to model, however.

It would be valuable to compare the gridded IRIS SSTs with the other two in-situ based reconstructions mentioned in chapter 2. As with HadSST3 and HadISST2, ERSSTv4 uses an ensemble approach to quantify uncertainties. We are not aware of uncertainty estimates being available for COBE-SST/COBE-SST2, which is not available as an ensemble. The HadSST dataset has recently been updated to version 4 (*Kennedy et al.*, 2019b), so it would also be worth including this in a future expanded analysis.

This thesis has focused solely on IRIS, but it may be possible to retrieve SST from other early satellite sensors such as THIR and MRIR. These sensors have their own relative advantages and limitations in terms of SST retrieval, but may offer the opportunity to test the reliability of the reconstructions in other periods of the pre-AVHRR era.

## 7.6 Overall conclusions

This thesis set out to investigate whether data from IRIS can be used to test the reliability of the historical reconstructions of SST. Instrument bias with more complex characteristics than previously understood has been exposed by this work, and proved to be a greater issue than anticipated. Bias correction of a satellite sensor from 1970 is challenging, due to the lack of raw calibration data and limited instrument documentation, among other factors. Nonetheless, we have contributed to characterising this bias in a spectrally resolved sense, which will directly contribute to future use of IRIS to provide observational constraints to future NWP reanalyses. As seen from the comparison of IRIS with THIR, the bias is complex (e.g. varies with orbital position and scene temperature) and further work is needed to understand and reduce it.

Our validation study using IRIS-like IASI data has shown that accurate and relatively unbiased SSTs can be retrieved from IRIS spectra known a priori to be clear-sky. However, the spatial resolution of IRIS means that very few cloud-free spectra can be identified. We have developed a method for retrieving SSTs from partly cloud-affected IRIS observations, enabling SSTs to be retrieved with near-global coverage. Biases in the retrieved SSTs remained too large for a robust test of the in-situ based reconstructions. Further testing of the cloudy retrieval scheme is needed to investigate whether these biases can be reduced, but this is unlikely to reduce the biases to the level required ($\sim 0.1\,\mathrm{K}$).

The IRIS dataset is valuable for improving future climate reanalysis datasets in

addition to SST retrieval. There is a renewed interest in IRIS for this purpose, and so it is hoped that our work on quality assurance and bias correction can feed into this. Moreover, the original scientific need to exploit all sources of data to test and constrain historical reconstructions of SST still stands, and this work may be further developed along some of the directions discussed in this chapter.

# REFERENCES


Allison, L. J., and J. S. Kennedy (1967), An evaluation of sea surface temperature as measured by the Nimbus I High Resolution Infrared Radiometer, *Tech. rep.*, National Aeronautics and Space Administration, Goddard Space Flight Center.

An, Z., and Z. Shi (2015), Scene Learning for Cloud Detection on Remote-Sensing Images, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(8), 4206–4222, 10.1109/JSTARS.2015.2438015.

Anderson, P. C., F. J. Rich, and S. Borisov (2018), Mapping the South Atlantic Anomaly continuously over 27 years, *Journal of Atmospheric and Solar-Terrestrial Physics*, *177*, 237–246, 10.1016/j.jastp.2018.03.015.

August, T., D. Klaes, P. Schlüssel, T. Hultberg, M. Crapeau, A. Arriaga, A. O'Carroll, D. Coppens, R. Munro, and X. Calbet (2012), IASI on Metop-A: Operational Level 2 retrievals after five years in orbit, *Journal of Quantitative Spectroscopy and Radiative Transfer*, *113*(11), 1340–1371, 10.1016/j.jqsrt.2012.02.028.

Bantges, R. J., H. E. Brindley, X. H. Chen, X. L. Huang, J. E. Harries, and J. E. Murray (2016), On the detection of robust multidecadal changes in Earth's outgoing longwave radiation spectrum, *Journal of Climate*, *29*(13), 4939–4947, 10.1175/JCLI -D-15-0713.1.

Berry, D. I., G. K. Corlett, O. Embury, and C. J. Merchant (2018), Stability assessment of the (A)ATSR sea surface temperature climate dataset from the European Space Agency Climate Change Initiative, *Remote Sensing*, *10*(1), 10.3390/rs10010126.

Capelle, V., A. Chédin, M. Pondrom, C. Crevoisier, R. Armante, L. Crepeau, and N. A. Scott (2018), Infrared dust aerosol optical depth retrieved daily from IASI and comparison with AERONET over the period 20072016, *Remote Sensing of Environment*, *206*, 15–32, 10.1016/j.rse.2017.12.008.

Carrea, L., O. Embury, and C. J. Merchant (2015), Datasets related to in-land water for limnology and remote sensing applications: distance-to-land, distance-to-water, water-body identifier and lake-centre co-ordinates, *Geoscience Data Journal*, *2*(2), 83–97, 10.1002/gdj3.32.

Chahine, M. T. (1974), Remote Sounding of Cloudy Atmospheres. I. The Single Cloud Layer, *Journal of the Atmospheric Sciences*, *31*(1), 233–243, 10.1175/1520-0469(1974)031⟨0233:RSOCAI⟩2.0.CO;2.

Chahine, M. T. (1977), Remote Sounding of Cloudy Atmospheres. II. Multiple Cloud Formations, *Journal of the Atmospheric Sciences*, *34*(5), 744–757, 10.1175/1520-0469(1977)034⟨0744:RSOCAI⟩2.0.CO;2.

Clough, S., M. Shephard, E. Mlawer, J. Delamere, M. Iacono, K. Cady-Pereira, S. Boukabara, and P. Brown (2005), Atmospheric radiative transfer modeling: a summary of the AER codes, *Journal of Quantitative Spectroscopy and Radiative Transfer*, *91*(2), 233 – 244, https://doi.org/10.1016/j.jqsrt.2004.05.058.

Conrath, B. J., R. A. Hanel, V. G. Kunde, and C. Prabhakara (1970), The Infrared Interferometer Experiment on Nimbus 3, *Journal of Geophysical Research*, *75*(30), 5831–5857, 10.1029/JC075i030p05831.

Dommenget, D. (2007), Evaluating EOF modes against a stochastic null hypothesis, *Climate Dynamics*, *28*(5), 517–531, 10.1007/s00382-006-0195-8.

Donlon, C. J., P. J. Minnett, C. Gentemann, T. J. Nightingale, I. J. Barton, B. Ward, and M. J. Murray (2002), Toward improved validation of satellite sea surface skin temperature measurements for climate research, *Journal of Climate*, *15*(4), 353–369, 10.1175/1520-0442(2002)015⟨0353:TIVOSS⟩2.0.CO;2.

Dybbroe, A., K. G. Karlsson, and A. Thoss (2005), NWCSAF AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modeling. Part I: Algorithm description, *Journal of Applied Meteorology*, *44*(1), 39–54, 10.1175/JAM-2188.1.

Embury, O., C. J. Merchant, and G. K. Corlett (2012), A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Initial validation, accounting for skin and diurnal variability effects, *Remote Sensing of Environment*, *116*, 62–78, https://doi.org/10.1016/j.rse.2011.02.028.

Emery, W. (2014), *Air Sea Interactions: Sea Surface Temperature*, pp. 136–143, 10 .1016/B978-0-12-382225-3.00065-7.

Eresmaa, R. (2014), Imager-assisted cloud detection for assimilation of Infrared Atmospheric Sounding Interferometer radiances, *Quarterly Journal of the Royal Meteorological Society*, *140*(684), 2342–2352, 10.1002/qj.2304.

EUMETSAT (2011), Infrared Atmospheric Sounding Interferometer (IASI) data taken on-board the Eumetsat Polar System (EPS) Metop-A Satellite. NERC Earth Observation Data Centre, http://catalogue.ceda.ac.uk/uuid/1977a2db646c506d0fc31bca6c6c080e.

EUMETSAT (2015), Ocean and Sea Ice Satellite Application Facility (9999): Global Sea Ice Concentration reprocessing dataset 1978-2015 (v1.2). Norwegian and Danish Meteorological Institutes, http://catalogue.ceda.ac.uk/uuid/8bbde1a8a0ce4a86904a3d7b2b917955.

Evans, W. D. (1942), The Standard Error of Percentiles, *37*(219), 367–376.

Ewing, G. (Ed.) (1965), *Oceanography from Space; proceedings of the Conference on the Feasibility of Conducting Oceanographic Explorations from Aircraft, Manned Orbital and Lunar Laboratories. Held August 24-28 at Woods Hole, Massachusetts, USA*.

Eyre, J. R. (1989a), Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. I: Theory and simulation for TOVS, *Quarterly Journal of the Royal Meteorological Society*, *115*, 1001–1026, 10.1002/qj.49711548902.

Eyre, J. R. (1989b), Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. II: Application to TOVS data, *Quarterly Journal of the Royal Meteorological Society*, *115*, 1027–1037, 10.1002/qj.49711548903.

Freeman, E., S. D. Woodruff, S. J. Worley, S. J. Lubker, E. C. Kent, W. E. Angel, D. I. Berry, P. Brohan, R. Eastman, L. Gates, W. Gloeden, Z. Ji, J. Lawrimore, N. A. Rayner, G. Rosenhagen, and S. R. Smith (2017), ICOADS Release 3.0: a major update to the historical marine climate record, *International Journal of Climatology*, *37*(5), 2211–2232, 10.1002/joc.4775.

Gandin, L. S. (1963), Objective analysis of meteorological fields, *ISPT*, p. 242.

GCOS (2009), Progress Report on the Implementation of the Global Observing System for Climate in Support of the UNFCCC 2004-2008. GCOS-129, WMO/TD No. 1489, GOOS-173, GTOS-70, August 2009, *Tech. rep.*

Greaves, J. R., R. Wexler, and C. J. Bowley (1966), The feasibility of sea surface temperature determination using satellite infrared data, *Tech. rep.*, Aracon Geophysics Co.; Concord, MA, United States.

Hanel, R. A., B. Schlachman, D. Rogers, and D. Vanous (1971), Nimbus 4 Michelson Interferometer, *Applied Optics*, *10*(6), 1376, 10.1364/ao.10.001376.

Hanel, R. A., B. J. Conrath, V. G. Kunde, C. Prabhakara, I. Revah, V. V. Salomonson, and G. Wolford (1972), The Nimbus 4 infrared spectroscopy experiment: 1.

Calibrated thermal emission spectra, *Journal of Geophysical Research*, *77*(15), 2629–2641, 10.1029/jc077i015p02629.

Hannachi, A. (2004), A primer for EOF analysis of climate data, *Tech. rep.*, University of Reading.

Hansel, R. A. (1994), IRIS/Nimbus-4 Level 1 Radiance Data V001, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://disc.gsfc.nasa.gov/datacollection/IRISN4RAD_001.html.

Harries, J. E., H. E. Brindley, P. J. Sagoo, and R. J. Bantges (2001), Increases in greenhouse forcing inferred from the outgoing longwave radiation spectra of the Earth in 1970 and 1997, *410*, 355–357, 10.1038/35066553.

Hay, W. W. (1996), Tectonics and climate, *Geologische Rundschau*, *85*(3), 409–437, 10.1007/BF02369000.

Hess, M., P. Koepke, and I. Schult (1998), Optical Properties of Aerosols and Clouds: The Software Package OPAC, *Bulletin of the American Meteorological Society*, *79*(5), 831–844, 10.1175/1520-0477(1998)079⟨0831:OPOAAC⟩2.0.CO;2.

Hirahara, S., M. Ishii, and Y. Fukuda (2014), Centennial-scale sea surface temperature analysis and its uncertainty, *Journal of Climate*, *27*(1), 57–75, 10.1175/JCLI-D-12-00837.1.

Huang, B., P. W. Thorne, V. F. Banzon, T. Boyer, G. Chepurin, J. H. Lawrimore, M. J. Menne, T. M. Smith, R. S. Vose, and H. M. Zhang (2017), Extended reconstructed Sea surface temperature, Version 5 (ERSSTv5): Upgrades, validations, and intercomparisons, *Journal of Climate*, *30*(20), 8179–8205, 10.1175/JCLI-D-16-0836.1.

Huang, H. L., P. Yang, H. Wei, B. A. Baum, Y. Hu, P. Antonelli, and S. A. Ackerman (2004), Inference of ice cloud properties from high spectral resolution infrared ob-

servations, *IEEE Transactions on Geoscience and Remote Sensing*, *42*(4), 842–853, 10.1109/TGRS.2003.822752.

Iacono, M. J., and S. A. Clough (1996), Application of infrared interferometer spectrometer clear sky spectral radiance to investigations of climate variability, *Journal of Geophysical Research Atmospheres*, *101*(23), 29,439–29,460.

IPCC (2013), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 1535 pp., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 10.1017/CBO9781107415324.

Irion, F. W., B. H. Kahn, M. M. Schreier, E. J. Fetzer, E. Fishbein, D. Fu, P. Kalmus, R. W. Chris, S. Wong, and Q. Yue (2018), Single-footprint retrievals of temperature, water vapor and cloud properties from AIRS, *Atmospheric Measurement Techniques*, *11*(2), 971–995, 10.5194/amt-11-971-2018.

Johnson, J., E. Esfandiari, E. Zamkoff, I. Gerasimov, A. Al-Jazrawi, and G. Alcott (2017), Data Recovery Effort of Nimbus era Observations by the NASA GES DISC. Goddard Earth Sciences Data and Information Services Center (GES DISC), NASA GSFC, *5th International Conference on Reanalysis 2017*.

Johnston, T., S. R. Young, D. Hughes, R. M. Patton, and D. White (2017), Optimizing Convolutional Neural Networks for Cloud Detection, *Proceedings of the Machine Learning on HPC Environments*, pp. 4:1–4:9, 10.1145/3146347.3146352.

Kennedy, J., N. Rayner, H. Titchner, S. Millington, R. Smith, and M. Saunby (2019a), The Met Ofce Hadley Centre sea ice and sea-surface temperature data set, version 2, part 2: Sea surface temperature analysis, in preparation.

Kennedy, J. J. (2014), A review of uncertainty in in situ measurements and data sets of sea surface temperature, *Reviews of Geophysics*, *52*(1), 1–32, 10.1002/2013RG000434.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby (2011a), Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties, *Journal of Geophysical Research: Atmospheres*, *116*(D14), 10.1029/2010JD015218.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby (2011b), Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization, *Journal of Geophysical Research: Atmospheres*, *116*(D14), 10.1029/2010JD015220.

Kennedy, J. J., N. A. Rayner, C. P. Atkinson, and R. E. Killick (2019b), An ensemble data set of seasurface temperature change from 1850: the Met Office Hadley Centre HadSST.4.0.0.0, *Journal of Geophysical Research: Atmospheres*, *124*.

Kent, E. C., J. J. Kennedy, D. I. Berry, and R. O. Smith (2010), Effects of instrumentation changes on sea surface temperature measured in situ, *Wiley Interdisciplinary Reviews: Climate Change*, *1*(5), 718–728, 10.1002/wcc.55.

Kent, E. C., J. J. Kennedy, T. M. Smith, S. Hirahara, B. Huang, A. Kaplan, D. E. Parker, C. P. Atkinson, D. I. Berry, G. Carella, Y. Fukuda, M. Ishii, P. D. Jones, F. Lindgren, C. J. Merchant, S. Morak-Bozzo, N. A. Rayner, V. Venema, S. Yasui, and H. M. Zhang (2017), A call for new approaches to quantifying biases in observations of sea surface temperature, *Bulletin of the American Meteorological Society*, *98*(8), 1601–1616, 10.1175/BAMS-D-15-00251.1.

Kiehl, J. T., and K. E. Trenberth (1997), Earth's Annual Global Mean Energy Budget, *Bulletin of the American Meteorological Society*, *78*(2), 197–208, 10.1175/1520-0477(1997)078⟨0197:EAGMEB⟩2.0.CO;2.

Laloyaux, P., M. Balmaseda, D. Dee, K. Mogensen, and P. Janssen (2016), A coupled data assimilation system for climate reanalysis, *Quarterly Journal of the Royal Meteorological Society*, *142*(694), 65–78, 10.1002/qj.2629.

Le Treut, H., R. Somerville, U. Cubasch, Y. Ding, C. Mauritzen, A. Mokssit, T. Peterson, and M. Prather (2007), *Historical overview of climate change. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L.(Eds.), Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of The Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Liu, X., W. L. Smith, D. K. Zhou, and A. Larar (2006), Principal component-based radiative transfer model for hyperspectral sensors: Theoretical concept, *Applied Optics*, *45*(1), 201–209, 10.1364/AO.45.000201.

Luo, T., W. Zhang, Y. Yu, M. Feng, B. Duan, and D. Xing (2019), Cloud detection using infrared atmospheric sounding interferometer observations by logistic regression, *International Journal of Remote Sensing*, *40*(17), 6530–6541, 10.1080/2150704X.2018.1553318.

Matricardi, M., and A. P. McNally (2014), The direct assimilation of principal components of IASI spectra in the ECMWF 4D-Var, *Quarterly Journal of the Royal Meteorological Society*, *140*(679), 573–582, 10.1002/qj.2156.

McNally, A. P., and P. D. Watts (2003), A cloud detection algorithm for high-spectral-resolution infrared sounders, *Quarterly Journal of the Royal Meteorological Society*, *129*(595 PART B), 3411–3423, 10.1256/qj.02.208.

Merchant, C. J., and O. Embury (2014), Chapter 4.3 - Simulation and Inversion of Satellite Thermal Measurements, in *Optical Radiometry for Ocean Climate Measurements*, *Experimental Methods in the Physical Sciences*, vol. 47, edited by G. Zibordi, C. J. Donlon, and A. C. Parr, pp. 489 – 526, Academic Press, https://doi.org/10.1016/B978-0-12-417011-7.00015-5.

Merchant, C. J., A. R. Harris, E. Maturi, and S. MacCallum (2005), Probabilistic

physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval, *Quarterly Journal of the Royal Meteorological Society*, *131*(611), 2735–2755, 10.1256/qj.05.15.

Merchant, C. J., O. Embury, J. Roberts-Jones, E. Fiedler, C. E. Bulgin, G. K. Corlett, S. Good, A. McLaren, N. Rayner, S. Morak-Bozzo, and C. Donlon (2014), Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI), *Geoscience Data Journal*, *1*(2), 179–191, 10.1002/gdj3.20.

O'Carroll, A. G., T. August, P. Le Borgne, and A. Marsouin (2012), The accuracy of SST retrievals from Metop-A IASI and AVHRR using the EUMETSAT OSI-SAF matchup dataset, *Remote Sensing of Environment*, *126*, 184–194, 10.1016/j.rse.2012.08.006.

O'Carroll, A. G., E. M. Armstrong, H. M. Beggs, M. Bouali, K. S. Casey, G. K. Corlett, P. Dash, C. J. Donlon, C. L. Gentemann, J. L. Høyer, A. Ignatov, K. Kabobah, M. Kachi, Y. Kurihara, I. Karagali, E. Maturi, C. J. Merchant, S. Marullo, P. J. Minnett, M. Pennybacker, B. Ramakrishnan, R. Ramsankaran, R. Santoleri, S. Sunder, S. Saux Picart, J. Vázquez-Cuervo, and W. Wimmer (2019), Observational Needs of Sea Surface Temperature, *Frontiers in Marine Science*, *6*, 10.3389/fmars.2019.00420.

Pavelin, E. G., S. J. English, and J. R. Eyre (2008), The assimilation of cloud-affected infrared satellite radiances for numerical weather prediction, *Quarterly Journal of the Royal Meteorological Society*, *134*(632), 737–749, 10.1002/qj.243.

Poli, P., and P. Brunel (2016), Meteorological satellite data rescue: Assessing radiances from Nimbus-IV IRIS (1970-1971) and Nimbus-VI HIRS (1975-1976). ERA Rep. 23, *Tech. rep.*, European Centre for Medium Range Weather Forecasts.

Poli, P., H. Hersbach, D. P. Dee, P. Berrisford, A. J. Simmons, F. Vitart, P. Laloyaux, D. G. Tan, C. Peubey, J. N. Thépaut, Y. Trémolet, E. V. Hólm, M. Bonavita,

L. Isaksen, and M. Fisher (2016), ERA-20C: An atmospheric reanalysis of the twentieth century, *Journal of Climate*, *29*(11), 4083–4097, 10.1175/JCLI-D-15-0556.1.

Prabhakara, C., G. Dalu, and V. G. Kunde (1974), Estimation of sea surface temperature from remote sensing in the 11- to 13-$\mu$m window region, *Journal of Geophysical Research*, *79*(33), 5039–5044, 10.1029/jc079i033p05039.

Rao, P. K., W. L. Smith, and R. Koffler (1972), Global Sea-Surface Temperature Distribution Determined From an Environmental Satellite, *Monthly Weather Review*, *100*(1), 10–14, 10.1175/1520-0493(1972)100⟨0010:gstddf⟩2.3.co;2.

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Scientific, River Edge.

Sabatini, R. R. (1970), The Nimbus IV User's Guide, Nimbus Proj. Office, Goddard Space Flight Center, Greenbelt, Md.

Saunders, R., J. Hocking, D. Rundle, P. Rayer, M. Matricardi, A. Geer, C. Lupu, P. Brunel, and J. Vidot (2013), RTTOV-11 Science and Validation Report, *Tech. rep.*, Met Office, UK.

Serio, C., A. M. Lubrano, F. Romano, and H. Shimoda (2000), Cloud detection over sea surface by use of autocorrelation functions of upwelling infrared spectra in the 800-900 $\mathrm{cm}^{-1}$ window region, *Applied Optics*, *39*(21), 3565, 10.1364/ao.39.003565.

Shi, M., F. Xie, Y. Zi, and J. Yin (2016), Cloud detection of remote sensing images by deep learning, in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 701–704, 10.1109/IGARSS.2016.7729176.

Shine, K. P., I. V. Ptashnik, and G. Rädel (2012), The Water Vapour Continuum: Brief History and Recent Developments, *Surveys in Geophysics*, *33*(3-4), 535–555, 10.1007/s10712-011-9170-y.

Smith, W. L., P. K. Rao, R. Koffler, and W. R. Curtis (1970), The determination of sea-surface temperature from satellite high resolution infrared window radiation measurements, *Monthly Weather Review*, *98*(8), 604–611, 10.1175/1520-0493(1970)098⟨0604:tdosst⟩2.3.co;2.

Susskind, J., C. D. Barnet, and J. M. Blaisdell (2003), Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds, *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2 PART 1), 390–409, 10.1109/TGRS.2002.808236.

Susskind, J., J. M. Blaisdell, and L. Iredell (2011), Improved surface and tropospheric temperatures determined using only shortwave channels: the AIRS Science Team Version-6 retrieval algorithm, 10.1117/12.893558.

Van Den Dool, H. M., S. Saha, and A. Johansson (2000), Empirical orthogonal teleconnections, *Journal of Climate*, *13*(8), 1421–1435, 10.1175/1520-0442(2000)013⟨1421:EOT⟩2.0.CO;2.

Wang, G. (2004), A Conceptual Modeling Study on Biosphere Atmosphere Interactions and Its Implications for Physically Based Climate Modeling, *Journal of Climate*, *17*, 2572–2583, 10.1175/1520-0442(2004)017⟨2572:ACMSOB⟩2.0.CO;2.

Wu, D., Y. Hu, M. P. McCormick, and F. Yan (2011), Global cloud-layer distribution statistics from 1 year CALIPSO lidar observations, *International Journal of Remote Sensing*, *32*(5), 1269–1288, 10.1080/01431160903530821.

Závody, A. M., C. T. Mutlow, and D. T. Llewellyn-Jones (2000), Cloud clearing over the ocean in the processing of data from the along-track scanning radiometer (ATSR), *Journal of Atmospheric and Oceanic Technology*, *17*(5), 595–615, 10.1175/1520-0426(2000)017⟨0595:CCOTOI⟩2.0.CO;2.