Bias correction and covariance parameters for optimal estimation by exploiting matched in-situ references

Bias Correction and Covariance Parameters for Optimal Estimation by Exploiting Matched In-situ References

Christopher J Merchant[a,b], Stéphane Saux-Picart[c] and Joanne Waller[a,b]

a. Department of Meteorology, University of Reading, Reading, RG6 6AL, UK; c.j.merchant@reading.ac.uk & j.a.waller@reading.ac.uk
b. National Centre for Earth Observation, University of Reading, RG6 6AL, Reading, UK
c. CNRM, Université de Toulouse, Météo-France, CNRS, 22300 Lannion, France; stephane.sauxpicart@meteo.fr

Corresponding author: c.j.merchant@reading.ac.uk

ABSTRACT

Optimal estimation (OE) is a core method in quantitative Earth observation. The optimality of OE depends on the errors in the prior, measurements and forward model being zero mean and having well-known error covariance. Often these assumptions are not met. We show how to use matches of satellite observations to in situ reference measurements to estimate parameters for use in OE that bring the retrieval framework closer to the theoretical optimality. This is done by retrieving bias correction and error covariance parameters. Bias correction parameters for some components of the retrieved state and for the satellite radiances are anchored by the in situ reference measurements, and are obtained by a modification of Kalman filtering. Error covariance matrices for the prior state and for the observation-simulation difference are iteratively obtained by applying equations for diagnosing internal retrieval consistency. The theory is applied to the case of OE of sea surface temperature from a sensor on a geostationary platform. Relative to an initial OE implementation, all measures of retrieval performance are improved when the optimised OE is tested on independent data: mean difference from validation data is reduced from -0.08 K to -0.01 K, and the standard deviation from 0.47 to 0.45 K; retrieval sensitivity to sea surface temperature increases from 71% to 76%; and a 20% underestimation of retrieval uncertainty is corrected. Perhaps more significant than the quantitative improvements are the coherent new insights into the forward model simulations and prior assumptions that are also obtained. These include estimates of prior bias in the absence of in situ information, an important consideration when in situ information is not globally distributed. Biases and lack of information about error covariances arise in remote sensing very often. While illustrated here for a particular case, the principles and methods we present for constraining that lack of knowledge systematically using ground truth will be widely applicable in remote sensing.

HIGHLIGHTS

- Method to determine bias and covariance parameters for optimal estimation
- Ensures assumptions underlying optimal estimation are more closely met
- Observation and prior state biases are constrained using matched ground truth
- Objective evaluation of prior and observation-simulation error covariances
- Example application to sea surface temperature, but method is widely applicable

Keywords: optimal estimation; remote sensing; retrieval theory; parameter estimation; bias correction; error covariance; sea surface temperature; SEVIRI

1 Introduction

Optimal estimation (OE) is an application of Bayes' theorem to the inverse problem of retrieving useful geophysical parameters from Earth observations (Rodgers, 2000). OE has been applied to the remote sensing of many geophysical parameters, including atmospheric trace gases (Carboni *et al.*, 2019; Buchwitz *et al.*, 2017; Munro *et al.*, 1998), atmospheric aerosol (Thomas *et al.*, 2009), cloud properties (McGarragh *et al.*, 2018; Heidinger, 2003; Poulsen *et al.*, 2012) and sea surface temperature (SST) (Merchant *et al.*, 2008; Merchant *et al.*, 2013). Strengths of OE include (McGarragh *et al.*, 2018): flexible use of information from all available wavebands, mutual consistency of multiple retrieved variables, multivariate characterisation of uncertainty (error covariance) and a framework for investigating information content of measurements.

Optimal estimation is not a trivial approach to implement, requiring availability of a forward model that can usefully simulate the observed satellite radiances and their local derivatives with respect to the variables describing the observed state. Moreover, for the retrieval result to be truly optimal a number of further conditions must be met (Rodgers, 2000), as follows.

Firstly, the forward model must have zero mean error relative to the satellite measurement. Herein, "error" is used strictly to mean the difference between the measured value and a true value of the measurand, and never as a synonym for "uncertainty" (JCGM, 2008). The combined error in the simulated minus the measured value of radiance is often called the "observation error", a term that will be used hereafter bearing in mind that it includes both measurement and forward model error. In practice, the accuracy of the calibration of satellite sensors and the accuracy of radiative transfer simulation are not generally sufficient to guarantee that observation errors have zero mean.

Secondly, the prior estimate of the state must be unbiased (meaning there are no systematic dependencies of the errors of the prior, and the errors have zero mean). This also is not generally the case.

Thirdly, the observation and the prior uncertainties (or, in the multivariate case, their error covariances) need to be well quantified in order to obtain the optimal solution and a realistic uncertainty evaluation for that solution. (The optimal solution is usually defined as the solution that minimises the retrieval uncertainty given the prior and the measurements.) In practice, it can be difficult to obtain representative observation or prior error covariance matrices, particularly as these error covariances are likely to vary with the retrieval context. The error covariance matrices are typically inferred from information such as sensor specifications, the degree of discrepancy between radiative transfer models and differences between data used as prior and other measurements of similar quantities. Expert judgement and a degree of arbitrariness has typically been involved. Specific criticism of the application of OE to SST (Koner *et al.*, 2015) has centred on this problem of determining appropriate error covariance matrices.

It is clearly desirable to put estimation of OE retrieval parameters on an objective footing which enables the conditions underlying the OE solution to be closely met. That is our aim in this paper: to describe and demonstrate a systematic framework to determine consistent estimates for the relative biases of forward model and instrument, prior biases, and parameters quantifying both the observation and prior error covariance matrices. The key is availability of independent data that act as a reference that is taken to be unbiased. The framework is demonstrated here in the context of joint OE of SST and total column water vapour (TCWV) from observations of an infrared radiometer, but is more widely applicable.

The essence of the method to obtain bias corrections is to retrieve bias-correction parameters progressively from many satellite-reference matches, thereby extending OE to be "bias-aware". The method is analogous to bias correction practices in data assimilation (Dee, 2005; Auligne *et al.*, 2007), and can be considered as a form of Kalman filtering for parameter estimation (Kalman, 1960), although not here applied sequentially in time or space.

The essence of the method to obtain error covariance parameters is to interrogate the pre- and post-retrieval residuals between the forward model and observations, using diagnostic formulations derived for application in data assimilation (Desroziers *et al.*, 2005). The "Desroziers" diagnostics have been used in the context of numerical weather prediction to estimate uncertainties for a variety of atmospheric observations including those from the Infrared Atmospheric Sounding Interferometer and the Spinning Enhanced Visible and InfraRed Imager (SEVIRI) (Stewart *et al.*, 2014; Stewart *et al.*, 1997; Waller *et al.*, 2016a; Waller *et al.*, 2016b; Cordoba *et al.*, 2017). The use of improved observation error statistics in operational assimilation has resulted in improved analyses and forecast skill (Weston *et al.*, 2014; Bormann *et al.*, 2016; Campbell *et al.*, 2017).

The remainder of this paper is structured as follows. The mathematics used to derive the bias and covariance parameter estimates is presented in section 2. Implementation will be illustrated with reference to retrieval of SST and TCWV using SEVIRI. The data used and the specifics of the implemented example are explained section 3. Section 4 presents the results for the example implementation, which is followed by a wider discussion and conclusions (section 5).

2 Expressions for Estimating the Retrieval Parameters

2.1 Preliminaries

Expressions for estimating the retrieval parameters are developed here specifically in the context of a maximum *a posteriori* (MAP) retrieval in the nearly linear case. In this case, the optimal estimate is formulated (Rodgers, 2000) as in Eq. 1.

$$\hat{\boldsymbol{z}} = \boldsymbol{z}_a + (\boldsymbol{K}^{\mathrm{T}}\boldsymbol{S}_{\epsilon}^{-1}\boldsymbol{K} + \boldsymbol{S}_a^{-1})^{-1}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{S}_{\epsilon}^{-1}(\boldsymbol{y} - \boldsymbol{F}) \qquad \text{Eq. 1}$$

Here: $\boldsymbol{z}$ is a vector containing variables adequate to describe the state of the observed system, $\boldsymbol{z}_a$ being the prior estimate of the state (from climatological or other background information) and $\hat{\boldsymbol{z}}$ being the OE retrieval of the state; $\boldsymbol{y}$ is a vector of observations which depend on the state, and $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{z}_a)$ is the corresponding simulation of the expected observations given the prior; the difference $\boldsymbol{y} - \boldsymbol{F}$ is transformed from "observation space" to "state space" by multiplication by a "gain" equal to $(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{S}_{\epsilon}^{-1}\boldsymbol{K} + \boldsymbol{S}_a^{-1})^{-1}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{S}_{\epsilon}^{-1}$ thus providing an update of the prior estimate of the state; the term $\boldsymbol{K} = \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{z}}|_{\boldsymbol{z}_a}$ contains the partial derivative of each observation with respect to each state variable, and is provided by the forward model; and the error covariance matrices for the prior, $\boldsymbol{S}_a$, and the simulation-minus-observation, $\boldsymbol{S}_{\epsilon}$, are square positive definite matrices that have to be specified and "given" to the OE.

Eq. 1 can be understood as follows. The starting point (prior) is $\boldsymbol{z}_a$ and the retrieval results, $\hat{\boldsymbol{z}}$, is improved relative to the prior by use of observations. In a case where $\boldsymbol{z}_a$ is a close approximation of the state, we would expect the observations to be close to $\boldsymbol{F}$ and then the retrieval is close to the prior. Non-zero differences $\boldsymbol{y} - \boldsymbol{F}$ that are significant compared to the uncertainties, in contrast, will significantly update the prior. The gain determines how strongly the observation-simulation difference updates the prior. The form of the gain gives the weight to $\boldsymbol{y} - \boldsymbol{F}$ using the inverse error covariance matrices, analogously to weighting the average of measured values of a quantity by their inverse squared uncertainty to give a best estimate for the quantity.

Error covariance matrices describe the uncertainty associated with a set of variables, and the correlations between the errors in different variables. Since we will be interested in interpreting the uncertainty and error correlations implied by the covariance matrices we estimate, it is convenient to note that any error covariance matrix can be simply decomposed into matrices that separate out these properties:

$$\boldsymbol{S} = \boldsymbol{URU} \qquad \text{Eq. 2}$$

where $\boldsymbol{U}$ is a diagonal matrix whose diagonal terms correspond to the uncertainty values of each variable, and $\boldsymbol{R}$ has off-diagonal terms equal to the coefficient of correlation of errors each pair of different variables (with 1s on its diagonal). If $\boldsymbol{R}$ is diagonal (equal to $\boldsymbol{I}$, the identity matrix), the errors are independent between variables.

2.2 Parameters for Correction of Bias

The differences $\boldsymbol{y} - \boldsymbol{F}$ are in general subject to errors that do not have zero mean over a large ensemble of retrievals: i.e., there are observation biases. These may arise in the measured values from the sensor and/or in the forward model, and have the equivalent effect of biasing the retrieved values irrespective of their origin. We therefore wish to estimate parameters for observation bias correction, $\boldsymbol{\beta}$, defined such that adding $\boldsymbol{\beta}$ to the forward model corrects for bias (relative to the observations). This definition is a convenient choice, and is not intended to imply that the forward model is the source of all the biases.

The prior estimate of the state may also be biased (i.e., may have a spatio-temporally persistent non-zero mean error across many instances). A vector $\boldsymbol{\gamma}$ is defined such that $\boldsymbol{z}_a + \boldsymbol{\gamma}$ is unbiased, and $\boldsymbol{\gamma}$ also needs to be estimated. If any elements of $\boldsymbol{z}_a$ are known or are defined to be unbiased, then the corresponding elements of $\boldsymbol{\gamma}$ contain zero.

The method for estimating the $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ is essentially to retrieve them as part of an extended state vector, $\tilde{\boldsymbol{z}}$. This is achieved progressively, refining the estimates of the parameters over many retrievals. Consider the i^{th} retrieval, where we have estimates from the previous retrieval for the bias correction parameters, written as $\boldsymbol{\gamma}_{i-1}$ and $\boldsymbol{\beta}_{i-1}$. The extended optimal estimate in the i^{th} retrieval is formulated in Eq. 3.

$$\tilde{\boldsymbol{z}}_i = \tilde{\boldsymbol{z}}_a + \left(\widetilde{\boldsymbol{K}}^{\mathrm{T}}\boldsymbol{S}_\epsilon^{-1}\widetilde{\boldsymbol{K}} + \tilde{\boldsymbol{S}}^{-1}\right)^{-1}\widetilde{\boldsymbol{K}}^{\mathrm{T}}\boldsymbol{S}_\epsilon^{-1}\left(\boldsymbol{y} - \left(\boldsymbol{F}(\boldsymbol{z}_a + \boldsymbol{\gamma}_{i-1}) + \boldsymbol{\beta}_{i-1}\right)\right) \qquad \text{Eq. 3}$$

$$\tilde{\boldsymbol{z}}_a = \begin{bmatrix} \boldsymbol{z}_a + \boldsymbol{\gamma}_{i-1} \\ \boldsymbol{\gamma}_{i-1} \\ \boldsymbol{\beta}_{i-1} \end{bmatrix}$$

$$\widetilde{\boldsymbol{K}} = \begin{bmatrix} \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{z}}|_{\boldsymbol{z}_a} & \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{z}}|_{\boldsymbol{z}_a} & \boldsymbol{I} \end{bmatrix}$$

$$\tilde{\boldsymbol{S}} = \begin{bmatrix} \boldsymbol{S}_a + \boldsymbol{S}_{\gamma_{i-1}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_{\gamma_{i-1}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_{\beta_{i-1}} \end{bmatrix}$$

In Eq. 3 the state is retrieved jointly with bias correction parameters for the prior estimate of the state and for the observations. The bias correction for the prior modifies the prior estimate of the state, which is why $\boldsymbol{\gamma}_{i-1}$ appears in the extended state vector both in the term $\boldsymbol{z}_a + \boldsymbol{\gamma}_{i-1}$ and in its own right as a retrieved vector. The forward model is also calculated for the bias corrected prior state $\boldsymbol{z}_a + \boldsymbol{\gamma}_{i-1}$. The partial derivatives of the forward model are identical with respect to the corresponding elements of $\boldsymbol{z}_a$ and $\boldsymbol{\gamma}_{i-1}$, as reflected in the formulation of $\widetilde{\boldsymbol{K}}$. The use of $\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{z}}|_{\boldsymbol{z}_a}$ is an approximation convenient for small corrections, and $\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{z}}|_{\boldsymbol{z}_a+\boldsymbol{\gamma}_i}$ must be evaluated otherwise. The final columns of $\widetilde{\boldsymbol{K}}$ are the partial derivatives of the bias corrected forward model with respect to $\boldsymbol{\beta}_{i-1}$. Since the bias correction has been formulated here as purely additive and independent between channels, these partial derivatives all equal 1 and the final columns are an identity matrix. More complex formulations of $\boldsymbol{\beta}$ would involve calculating appropriate partial derivatives here.

The extended prior error covariance matrix, $\tilde{\boldsymbol{S}}$, is block diagonal. The blocks relating to the bias correction parameters having been carried forward from the error covariance matrix of the solution of the previous iteration. Considering the result of the i^{th} retrieval, the error covariance matrix of the solution is:

$$\boldsymbol{S}_{\tilde{z}_i} = \left(\tilde{\boldsymbol{K}}^{\mathrm{T}}\boldsymbol{S}_{\epsilon}^{-1}\tilde{\boldsymbol{K}} + \tilde{\boldsymbol{S}}^{-1}\right)^{-1} = \begin{bmatrix} \boldsymbol{S}_{z_i} & \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{A}^{\mathrm{T}} & \boldsymbol{S}_{\gamma_i} & \boldsymbol{C} \\ \boldsymbol{B}^{\mathrm{T}} & \boldsymbol{C}^{T} & \boldsymbol{S}_{\beta_i} \end{bmatrix} \quad \text{Eq. 4}$$

The matrices $\boldsymbol{S}_{\gamma_i}$ and $\boldsymbol{S}_{\beta_i}$ are taken from the evaluation of Eq. 4 and passed to the $\tilde{\boldsymbol{S}}$ of the subsequent retrieval. The blocks $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ are not passed forward to the subsequent $\tilde{\boldsymbol{S}}$ which imposes independence between the errors in the state and bias correction vectors. This is in contrast to Kalman filtering, where the iterations are sequential in space and/or time, and the assumption is made that $\boldsymbol{S}_{\tilde{z}_i}$ in its entirety is a good estimate for $\tilde{\boldsymbol{S}}$ in iteration $i+1$.

The bias correction for the prior affects the calculated value of $\boldsymbol{F}$ and we are also attempting to derive a bias correction for $\boldsymbol{y} - \boldsymbol{F}$ simultaneously. Thus, there may be ambiguity between these bias corrections which could affect convergence. Here, we will anchor one or more of the elements of $\boldsymbol{\gamma}$ and or $\boldsymbol{\beta}$ to zero using in situ reference data, which we find to be sufficient for convergence. A full analysis of convergence conditions is beyond the scope of this paper, although criteria for monitoring progress to convergence are provided below.

Note that if any element of $\boldsymbol{\gamma}$ or $\boldsymbol{\beta}$ is externally constrained to a fixed value, the corresponding rows and columns are deleted from the vectors and matrices of Eq. 3.

2.3 Observation Error Covariances

The observation error covariance matrix, $\boldsymbol{S}_\epsilon$, is estimated given specified bias corrections, since the observation bias correction is effectively part of the forward model: $\boldsymbol{F}' = \boldsymbol{F} + \boldsymbol{\beta}$. To estimate $\boldsymbol{S}_\epsilon$, we make use of an equation derived for diagnosing the consistency of a data assimilation system (Desroziers *et al.*, 2005). Re-written in the retrieval nomenclature we have:

$$E[(\boldsymbol{y} - \boldsymbol{F}'(\hat{\boldsymbol{z}}))(\boldsymbol{y} - \boldsymbol{F}'(\boldsymbol{z}_a))^{\mathrm{T}}] = \boldsymbol{S}_\epsilon \qquad \text{Eq. 5}$$

where $E[.]$ signifies expectation. The expression says that expectation of the outer product of two terms equals the observation error covariance for a well-formulated optimal estimate. The two terms in the outer product are the difference between the observations and the simulation for the retrieved state, and the difference between the observations and the simulation for the prior state. Here, we reverse the application of the diagnostic equation, and use an approximation to the left-hand side as a new estimate for $\boldsymbol{S}_\epsilon$.

To apply Eq. 5 in this way**,** three adaptations are made. First, we must estimate the expectation as the average across many instances. Second, since Eq. 5 assumes the bias-free case, and biases may not on any given evaluation have been fully removed, the differences are shifted to give zero mean. Third, we must force the result to be strictly symmetric. Using $\langle . \rangle$ to indicate the arithmetic average over an ensemble of instances, we obtain:

$$\widehat{\boldsymbol{S}}_\epsilon = \frac{1}{2}\langle \boldsymbol{d}_r^o {\boldsymbol{d}_a^o}^{\mathrm{T}} + \boldsymbol{d}_a^o {\boldsymbol{d}_r^o}^{\mathrm{T}} \rangle \qquad \text{Eq. 6}$$

$$\boldsymbol{d}_r^o = \boldsymbol{y} - \boldsymbol{F}'(\hat{\boldsymbol{z}}) - \langle \boldsymbol{y} - \boldsymbol{F}'(\hat{\boldsymbol{z}}) \rangle$$

$$\boldsymbol{d}_a^o = \boldsymbol{y} - \boldsymbol{F}'(\boldsymbol{z}_a) - \langle \boldsymbol{y} - \boldsymbol{F}'(\boldsymbol{z}_a) \rangle$$

2.4 Prior Error Covariances

Another data assimilation diagnostic (Desroziers *et al.*, 2005) corresponds to:

$$E[(\boldsymbol{F}'(\hat{\boldsymbol{z}}) - \boldsymbol{F}'(\boldsymbol{z}_a))(\boldsymbol{y} - \boldsymbol{F}'(\boldsymbol{z}_a))^{\mathrm{T}}] = \boldsymbol{K}\boldsymbol{S}_a\boldsymbol{K}^{\mathrm{T}} \qquad \text{Eq. 7}$$

We adapt this to provide an estimate of $\boldsymbol{S}_a$ as follows. First, note that $\boldsymbol{K}$ is variable between instances, but we have an estimate of $\boldsymbol{K}$ from the forward model in each case. While, in data assimilation, $\boldsymbol{K}\boldsymbol{S}_a\boldsymbol{K}^{\mathrm{T}}$ is often assessed "in observation space", here we isolate $\boldsymbol{S}_a$ by pre-multiplication of both sides by $(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K})^{-1}\boldsymbol{K}^{\mathrm{T}}$ and post-multiplication of both sides by $\boldsymbol{K}(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K})^{-1}$. Again, we adapt the diagnostic equation by averaging over an ensemble of instances and imposing a re-zeroed, symmetric form, obtaining:

$$\widehat{\boldsymbol{S}}_a = \frac{1}{2}\langle (\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K})^{-1}\boldsymbol{K}^{\mathrm{T}} \left(\boldsymbol{d}_a^r {\boldsymbol{d}_a^o}^{\mathrm{T}} + \boldsymbol{d}_a^o {\boldsymbol{d}_a^r}^{\mathrm{T}}\right)\boldsymbol{K}(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K})^{-1} \rangle \qquad \text{Eq. 8}$$

$$\boldsymbol{d}_a^r = \boldsymbol{F}'(\hat{\boldsymbol{z}}) - \boldsymbol{F}'(\boldsymbol{z}_a) - \langle \boldsymbol{F}'(\hat{\boldsymbol{z}}) - \boldsymbol{F}'(\boldsymbol{z}_a) \rangle$$

2.5 A Convergence Metric

A final diagnostic relationship using both $\widehat{\boldsymbol{S}}_\epsilon$ and $\widehat{\boldsymbol{S}}_a$ can be re-cast as a metric of self-consistency. In a consistent system (Desroziers *et al.*, 2005):

$$E[(\boldsymbol{y}-\boldsymbol{F}'(\boldsymbol{z}_a))(\boldsymbol{y}-\boldsymbol{F}'(\boldsymbol{z}_a))^{\mathrm{T}}]=\boldsymbol{S}_\epsilon+\boldsymbol{K}\boldsymbol{S}_a\boldsymbol{K}^{\mathrm{T}} \qquad \text{Eq. 9}$$

and therefore, if the newly estimated error covariance matrices are well quantified, we should find that:

$$\langle\widehat{\boldsymbol{S}}_\epsilon+\boldsymbol{K}\widehat{\boldsymbol{S}}_a\boldsymbol{K}^{\mathrm{T}}\rangle^{-1}\langle\boldsymbol{d}_a^o\boldsymbol{d}_a^{o\mathrm{T}}\rangle-\boldsymbol{I}\approx\boldsymbol{0} \qquad \text{Eq. 10}$$

The element-wise sum of squares of the expression on the left-hand side is a measure of the inconsistency of the error covariance assumptions: as the value decreases, inconsistency decreases and the assumptions are more consistent with the data. Note that the metric involves both covariance matrices and, via $\boldsymbol{d}_a^o$, the prior and observation bias corrections, and therefore tests the consistency of all the estimates. This metric enables us to verify that internal consistency is improved when we revise an estimate of any error covariance parameters, and that there is convergence in the system of parameters being obtained.

3 Example Implementation

3.1 Formulation of Optimal Estimator

We apply these expressions for estimating retrieval parameters to the case of OE of SST (x) and TCWV (w) from SEVIRI. The retrieval formulation has been developed primarily to retrieve SST for operational meteorology and oceanography (Merchant *et al.*, 2008), and is used for SST climate data records (Merchant *et al.*, 2014). The optimal estimator has the same form as Eq. 1, except that a reduced state vector, $\boldsymbol{z} = \begin{bmatrix} x \\ w \end{bmatrix}$, is retrieved. $\boldsymbol{z}_a$ is derived from a full prior state vector, $\boldsymbol{x}_a$, consisting of the complete profiles of temperature and humidity from operational numerical weather prediction (NWP) fields from forecasts of the European Centre for Medium-range Weather Forecasting (Vitart, 2014). The full prior is used for the forward model simulation: thus $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{x}_a)$ and $\boldsymbol{K} = \frac{\partial \boldsymbol{F}(\boldsymbol{x}_a)}{\partial \boldsymbol{z}}|_{\boldsymbol{z}_a}$. To relate changes in w to changes in the humidity variables in $\boldsymbol{x}$ the assumption is made that the absolute humidity changes by the same fraction throughout the atmospheric column.

The reduced state vector is used for retrieval because there is limited amount of information about TCWV available in the infrared window channels used for SST determination, although there is sensitivity to the column water vapour (Merchant *et al.*, 2006b). The reduced state vector formulation neglects less dominant terms (the vertical distribution of water vapour, the atmospheric temperature profile, aerosols, etc). This approximation may be a further source of bias in the optimal estimator, if any prior information for these terms is biased.

The observation vector is $\boldsymbol{y} = \begin{bmatrix} y_{8.7} \\ y_{10.8} \\ y_{12.0} \end{bmatrix}$, where y_λ refers to the brightness temperature (BT) of the SEVIRI channel centred on a wavelength of λ μm. Thus, we use the three thermal window channels of SEVIRI that are useable for SST retrieval both night and day. BTs are used rather than radiances because this renders the retrieval nearly linear and amenable to solution in one step. The forward model is RTTOV v11.2 (Saunders *et al.*, 2018).

3.2 Data

We use a dataset of observations from SEVIRI matched to drifting buoy measurements. The SEVIRI sensor in question is operational on the platform Meteosat-09, which was launched in December 2005. The buoy measurements are within the field of view of the SEVIRI pixel and within 30 minutes of the pixel acquisition time. The SEVIRI cloud screening, quality flagging, initial radiance bias correction and matching are done within the systems of the Ocean and Sea-Ice Satellite Applications Facility (OSI-SAF).

Two years of data are exploited: data from the year 2011 are used as a training set from which retrieval parameters are derived, and the quoted results are for the application of those parameters to data from the year 2012. There is no particular significance of these years, other than match-up data (MD) being accessible with an augmented set of contextual information.

There are 167,808 satellite-buoy matches in the 2011 (training) MD, and 153,394 in the 2012 (application) MD. The distribution of matches in 2011 is illustrated in Figure 1. In 2012 they are similarly distributed. The information in the dataset includes: the satellite (brightness temperature, BT) and drifting buoy (SST) measurements; a quality level (QL), derived in the operational system from a number of considerations such as proximity to flagged clouds; a numerical weather prediction (NWP) forecast of the atmospheric temperature and humidity profiles, needed as input for radiative transfer simulation of SEVIRI BTs; an operational estimate of the simulation bias relative to the satellite observations, estimated on timescales of 3 days on spatial scales of order 5 degrees from averages of simulation minus observation differences in night-time data; spatio-temporal geolocation information, such as satellite zenith angle; and the value of SST from the operational SST analysis, OSTIA, for the location and day. All the above fields are available within the OSI-SAF operational processing system and can be exploited in near-real time.

Since the recommended OSI-SAF SSTs comprise those from pixels with QL 4 and 5, only those pixels are included in the MD. Quality control flags for identifying bad quality drifting-buoy temperatures have been applied. 2.8% of matches have been rejected where the drifting buoy temperature differs from the SST of OSTIA by more than 1.6 K, which is around eight times the expected uncertainty in drifting buoy SST (Lean and Saunders, 2013). A similar proportion of matches is excluded where an index of desert dust (Merchant *et al.*, 2006a) indicates elevated tropospheric aerosol.

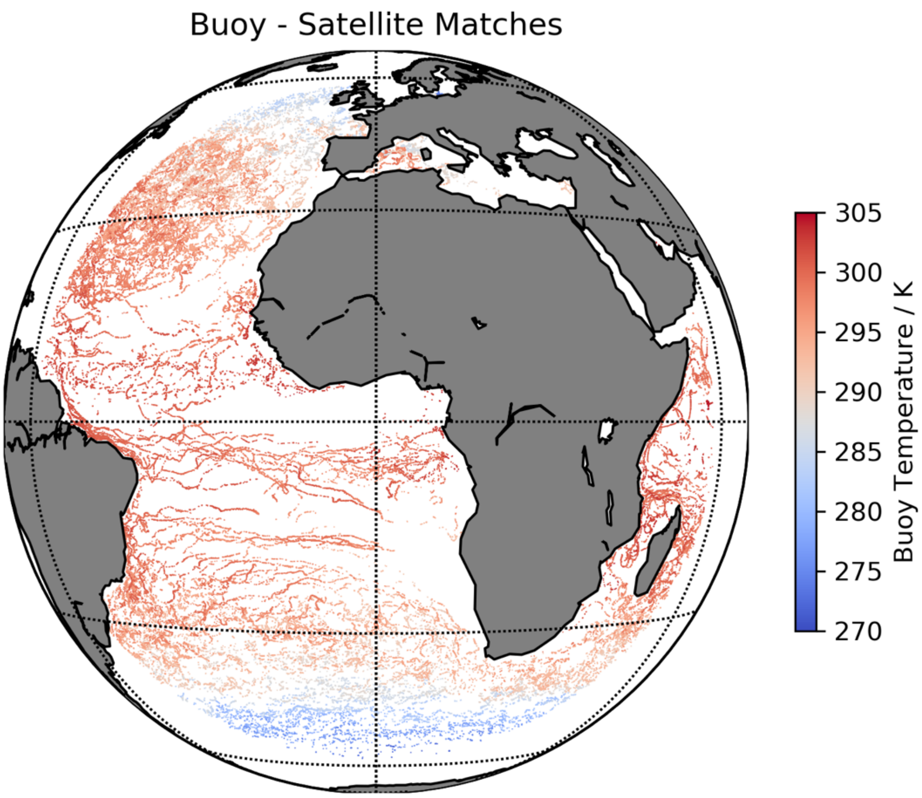


Figure 1. Distribution of satellite-buoy matches used in this study. The locations shown are for 2011, and the distribution in 2012 is similar. Matched locations are coloured with the measured buoy sea surface temperature.

The radiative transfer model, RTTOV v11.2, was run for each match on the NWP profiles for the SEVIRI observation geometry, assuming cloud-free no-aerosol conditions. The SST used in the simulation for the training year was the drifting buoy SST minus a static adjustment for the ocean thermal skin effect of 0.17 K (Donlon *et al.*, 2002). The ocean skin effect is variable (e.g., Saunders, 1967; Wong and Minnett, 2018), and for the present purpose, this adjustment is intended to correct for the mean skin effect to within an uncertainty of order 0.1 K. A climatological SST was used for the simulation for the test year, acting as both a prior and linearization point for the test retrievals. The climatology used was the average for the day of year over the complete years 1982 to 2010 from a satellite-based analysis of SST at 20 cm (Merchant *et al.*, 2019).

3.3 Implementation

3.3.1 Overview

Section 2 provides equations for three steps of parameter estimation to improve OE results for SST (bias correction, observation error covariance estimation, and prior error covariance estimation). The parameters estimated in each step are contained in two vectors of bias correction parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and two covariance matrices, $\boldsymbol{S}_{\epsilon}$ and $\boldsymbol{S}_{a}$. We implement the equations sequentially, but the estimates of the retrieval parameters are not independent, in that the current evaluation of each parameter set influences the evaluation of the others. The optimisation of the retrieval parameters is therefore done by iterating the estimation sequence, as shown in Figure 2 and explained in the following subsections.

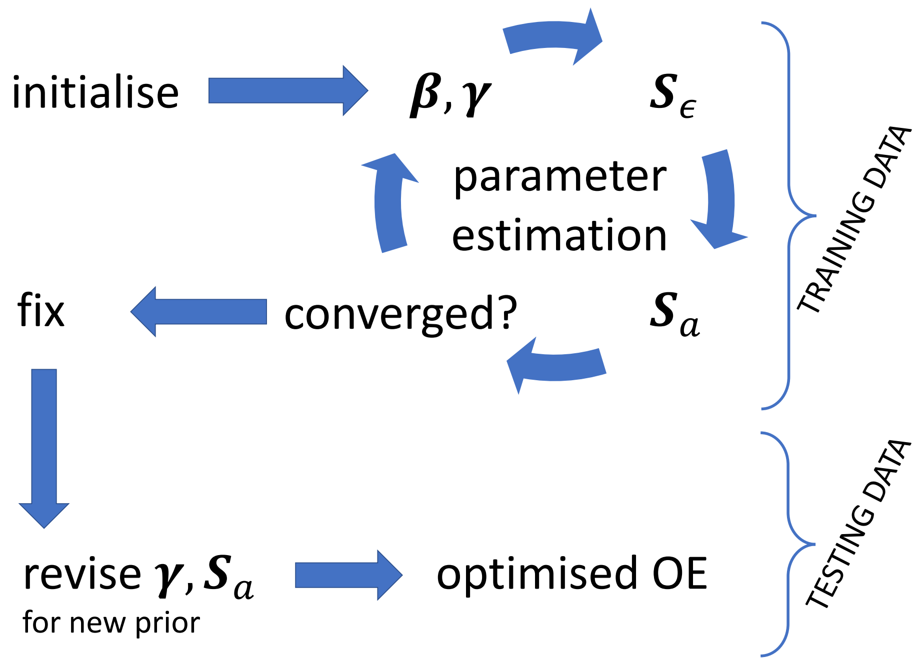


Figure 2. *The sequence of estimation of three sets of parameters for optimal estimation. For symbols, see the main text.*

3.3.2 Initialisation

The first step is to initialise all retrieval parameters. The OSI-SAF ran an OE retrieval experimentally (Merchant *et al.*, 2009b) and the initial estimates of the retrieval parameters are the values of the parameters trialled in that experimental chain. These were specified based on case studies, expert judgement, understanding of SEVIRI sensor characteristics, etc. The observation biases (after the operational radiance bias adjustments are applied) are initialised as zero in each channel. Zero mean bias is also assumed for TCWV. The initial model for the observation error covariance is given by Eq. 11.

$$\boldsymbol{S}_{\epsilon} = \begin{bmatrix} {u^{o}_{8.7}}^{2} & 0 & 0 \\ 0 & {u^{o}_{10.8}}^{2} & 0 \\ 0 & 0 & {u^{o}_{12.0}}^{2} \end{bmatrix} + \begin{bmatrix} {u^{s}_{8.7}}^{2} s^{2} & 0 & 0 \\ 0 & {u^{s}_{10.8}}^{2} s^{2} & 0 \\ 0 & 0 & {u^{s}_{12.0}}^{2} s^{2} \end{bmatrix} \qquad \text{Eq. 11}$$

Here, $s = \sec(\theta)$, where θ is the satellite zenith angle, and s is therefore the length of the path of the ray from the surface to the satellite through the atmosphere relative to a nadir ray (hereafter referred to as the 'path'); u^{o}_{λ} is the measurement uncertainty for the channel centred on λ μm; and u^{s}_{λ} is the corresponding simulation uncertainty. The numerical values are given in Table 1. Eq. 11 embodies some understanding about the observation error-covariance structure and has some limitations. The diagonal form of the measurement error covariance expresses the understanding measurement errors are dominated by radiometric noise which is independent between the BTs of different channels. The values of the noise levels were estimated in (Merchant *et al.*, 2013). The simulation uncertainties are modelled as being proportional to the path, expressing the understanding that the parameterisation of the RTTOV model is more accurate for a nadir path than at high zenith angles. Since the parameterisation of the RTTOV model has the same form for all three channels, it is reasonable to expect that the simulation errors have some degree of cross-channel correlation, but in the absence of quantitative information, the initial assumption is to set the correlations to zero.

Table 1. Initial assumptions about observation-simulation uncertainties.

	Measurement Uncertainty / K			**Nadir simulation Uncertainty / K**		
Channel	8.7	10.8	12.0	8.7	10.8	12.0
Estimate	0.11	0.11	0.15	0.15	0.15	0.15

The initial model for the prior error covariance is also diagonal:

$$\boldsymbol{S}_a = \begin{bmatrix} u_x^2 & 0 \\ 0 & u_w^2 \end{bmatrix}; u_w = aw_a + bw_a^2 \qquad \text{Eq. 12}$$

with the values $a = \frac{3}{10}$ and $b = -\frac{1}{30}$, for the prior total column water vapour, w_a in g cm^{-2}. The initial assumption about the uncertainty of drifting buoy SST is 0.2 K. This is a little greater than inferred by (Lean and Saunders, 2013), which gives some leeway for skin and point-to-pixel variability. It is reasonable to expect that the SST and TCWV errors are uncorrelated, although off-diagonal parameters will be estimated.

3.3.3 Bias-correction Parameters

Having set initial values of all retrieval parameters, the cycle of parameter estimation begins with bias estimation, using the training data subset. Four bias parameters are to be estimated: a brightness temperature correction for each SEVIRI channel (i.e., $\boldsymbol{\beta} = [\beta_{8.7} \quad \beta_{10.8} \quad \beta_{12.0}]^{\mathrm{T}}$) and a bias correction for the prior TCWV only (i.e., $\boldsymbol{\gamma} = [0, \gamma_w]^{\mathrm{T}}$). No bias correction for SST is estimated because the skin-adjusted drifting buoy SSTs collectively provide an SST reference and are the anchor for the other bias corrections.

The number of bias-correction parameter values to be estimated is larger than four ($\beta_{8.7}$, $\beta_{10.8}$, $\beta_{12.0}$ and γ_w) since the values depend on retrieval context. Parameterising the bias parameter dependencies requires scientific insight and judgement. Here, we assume, first, that the observation biases depend on the quality level attributed to the observation in the operational system. Since we are addressing QL = 4 and 5 data, there are two bias-correction parameter values estimated for each channel. Next, we assume that the TCWV bias may be a function of TCWV itself. This is achieved by estimating a parameter value from matches stratified within each quintile (containing ~33562 matches) of the TCWV range, i.e., 5 parameter values for the bias correction of prior TCWV are obtained. The TCWV bias parameters are also derived per quality level. We do not consider that the prior TCWV bias truly depends on the quality level, but it turns out that the apparent TCWV bias does differ between quality level 4 and 5; the interpretation of this outcome will be discussed in the results section.

To estimate the bias-correction parameter values, Eq. 3 is applied repeatedly on matches drawn at random from the training data. Each extended retrieval updates the values of $\boldsymbol{\beta}$ for either QL = 4 or 5 (according to the QL of the match drawn) and of γ_w for the TCWV stratum in which the match features. The updated values are passed to the next extended retrieval for a randomly selected match. The retrieved state is not re-used in any later iteration (which distinguishes this approach for using reference data from Kalman filtering). The bias-correction parameters for all strata of the data stabilise after ~20,000 iterations, by which point most matches remain unused in a given cycle. Note that randomly selecting matches allows matches to be reused, and convergence may be obtainable even where the number of training matches are fewer than the required number of iterations. The parameter values obtained are then fixed during the next step in the cycle of parameter estimation.

3.3.4 Observation Error Covariances

To obtain a revised estimate for the observation error covariance matrix, OE retrieval is undertaken using the bias corrections just obtained, using Eq. 13.

$$\hat{\boldsymbol{z}} = \boldsymbol{z}_a + \boldsymbol{\gamma} + (\boldsymbol{K}^{\mathrm{T}}\boldsymbol{S}_\epsilon^{-1}\boldsymbol{K} + \boldsymbol{S}_a^{-1})^{-1}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{S}_\epsilon^{-1}\left(\boldsymbol{y} - \left(\boldsymbol{F} + \gamma_w \frac{\partial \boldsymbol{F}}{\partial w}|_{w_a} + \boldsymbol{\beta}\right)\right) \qquad \text{Eq. 13}$$

Note that in this step the optimal estimator is not extended as it was when using Eq. 3: bias corrections are applied but are not re-estimated. The formulation of Eq. 13 assumes that the prior TCWV correction is sufficiently small that a first-order term adequately represents the effect of the adjustment of the prior on the forward model BTs. This is convenient in that it avoids recalculation of the simulations, but if the changes are beyond the linear range $\boldsymbol{F} + \gamma_w \frac{\partial \boldsymbol{F}}{\partial w}|_{\boldsymbol{w_a}}$ should be replaced with $\boldsymbol{F}(\boldsymbol{z}_a + \boldsymbol{\gamma})$ in Eq. 13.

Eq. 13 is applied to all matches in the training data, and the retrieval results are used to evaluate the observation error covariances using Eq. 6. As seen in Eq. 11, we expect observation error covariances to depend on the path, s, because the forward model uncertainty is likely to increase with satellite zenith angle (other factors being equal). $\widehat{\boldsymbol{S}}_\epsilon$ is therefore found subsets of the data stratified by s. The strata boundaries are defined by the quintiles of the s distribution in the training data, so effectively five estimates of $\widehat{\boldsymbol{S}}_\epsilon$ are formed for different ranges of satellite zenith angle.

3.3.5 Prior Error Covariances

To obtain a revised estimate for the prior error covariance matrix, OE retrieval is undertaken using the bias corrections and the observation error covariances just obtained, using Eq. 14.

$$\hat{\boldsymbol{z}} = \boldsymbol{z}_a + \boldsymbol{\gamma} + \left(\boldsymbol{K}^{\mathrm{T}}\widehat{\boldsymbol{S}}_\epsilon^{-1}\boldsymbol{K} + \boldsymbol{S}_a^{-1}\right)^{-1}\boldsymbol{K}^{\mathrm{T}}\widehat{\boldsymbol{S}}_\epsilon^{-1}\left(\boldsymbol{y} - \left(\boldsymbol{F} + \gamma_w \frac{\partial \boldsymbol{F}}{\partial w}|_{w_{\boldsymbol{a}}} + \boldsymbol{\beta}\right)\right) \qquad \text{Eq. 14}$$

This differs from Eq. 13 only in using the new estimate for observation error covariance. $\widehat{\boldsymbol{S}}_\epsilon$ is determined for a given match by piecewise linear interpolation of the five stratified estimates with respect to s.

Eq. 8 is evaluated using the retrieval result of Eq. 14. Again, this is done for strata of the training data, since we expect u_w to vary with w_a (as also seen in the initial formulation, Eq. 12. The strata are defined by the quintiles of the w_a distribution of the training data.

3.3.6 Consistency and Convergence

Eq. 10 is evaluated to verify that the result of the three parameter estimation steps has led to a set of retrieval parameters that are more consistent with the data, i.e., that this metric has decreased towards zero.

Convergence is assessed pragmatically on the basis of how much change there is in the retrieved SST since the previous cycle of parameter estimation. If the differences in retrieved SST (the latest results minus either the initial results or the results of the previous cycle) are small, then further cycles serve no practical purpose in improving SST retrieval.

Specifically, we take the estimation process as having converged if the standard deviation of these differences is less than 0.01 K.

If this criterion is not met, then $\boldsymbol{\beta}$, γ_w, $\widehat{\boldsymbol{S}}_\epsilon$ and $\widehat{\boldsymbol{S}}_a$ are carried forward to a further cycle of parameter estimation, commencing with the refine of the bias correction parameter values in the light of the improved estimates of the error covariance matrices.

3.3.7 Revision of Prior Parameters

The evaluations of $\boldsymbol{\beta}$, γ_w, $\widehat{\boldsymbol{S}}_\epsilon$ and $\widehat{\boldsymbol{S}}_a$ are for use in OE of SST and TCWV in circumstances where the prior SST is not provided by reference temperatures from drifting buoys.

In this study, the prior SST in the test set, x_c, is from a climatology based on the period 1982 - 2010. This prior SST is not assumed to be unbiased relative to SST in 2012, so in general $\gamma_x \neq 0$, and the prior SST uncertainty, u_{x_c}, when using a climatology is greater than when using drifting buoy SSTs. γ_x and u_{x_c} need to be estimated. An important point to note is that in this step the revision of the parameters is done independently of any drifting buoy information: the anchoring of the prior SST bias correction comes from having BT and TCWV bias correction parameters available. The assumption is that $\boldsymbol{\beta}$, γ_w need not change, which means in this case that parameter values derived from training data from 2011 are valid for the NWP data and SEVIRI BT calibration in 2012.

To estimate the prior SST bias, Eq. 3 is adapted to an expression, Eq. 15, that uses the OE parameters previously obtained and enables iterative calculation of γ_x over many random cases.

$$\tilde{\boldsymbol{z}}_i = \tilde{\boldsymbol{z}}_a + \left(\widetilde{\boldsymbol{K}}^{\mathrm{T}}\widehat{\boldsymbol{S}}_{\epsilon}^{-1}\widetilde{\boldsymbol{K}} + \tilde{\boldsymbol{S}}^{-1}\right)^{-1}\widetilde{\boldsymbol{K}}^{\mathrm{T}}\widehat{\boldsymbol{S}}_{\epsilon}^{-1}\left(\boldsymbol{y} - \boldsymbol{\beta} - \boldsymbol{F}(\tilde{\boldsymbol{z}}_a)\right) \qquad \text{Eq. 15}$$

$$\tilde{\boldsymbol{z}}_a = \begin{bmatrix} x_c + \gamma_{x_{i-1}} \\ w_a + \gamma_w \\ \gamma_{x_{i-1}} \end{bmatrix}$$

$$\widetilde{\boldsymbol{K}} = \left[\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{z}}\Big|_{z_a} \quad \frac{\partial \boldsymbol{F}}{\partial x}\Big|_{x_c}\right]$$

$$\tilde{\boldsymbol{S}} = \begin{bmatrix} \widehat{\boldsymbol{S}}_a + \boldsymbol{S}_{\gamma_{i-1}} & 0 \\ 0 & {\sigma_{x_{i-1}}}^2 \end{bmatrix}$$

Here, $\sigma_{x_{i-1}}$ is the uncertainty in the estimate of $\gamma_{x_{i-1}}$ from the previous iteration. In this implementation, γ_x has been estimated in the annual average for each of 8 zones of latitude, each 15° of latitude wide, spanning 60°S to 60°N. (Since the presence of in situ matches is not necessary for estimating the prior SST bias, an operational implementation using frequent imagery data could provide an estimate on a time-variable basis with greater spatial resolution, including longitudinal discrimination.) Once repeated application of Eq. 15 has converged on stable values of γ_x, Eq. 8 is evaluated over the whole dataset to obtain an updated prior error covariance matrix, from which the u_{x_c} estimate is substituted into $\widehat{\boldsymbol{S}}_a$.

4 Results

The cycle of estimation of $\boldsymbol{\beta}$, γ_w, $\boldsymbol{S}_\epsilon$ and $\boldsymbol{S}_a$ was applied four times. The consistency metric (Eq. 10) was calculated at the end of each cycle, and decreased as expected: the initial value was 2.3, followed by 0.75, 0.26, 0.10 and 0.05 respectively after each iteration. The standard deviation of the change in retrieved SST between cycles also decreased monotonically, being 0.35 K, 0.025 K, 0.014 K and, between the results of the penultimate and final iteration, 0.009 K. This represents adequate convergence in retrieved SST.

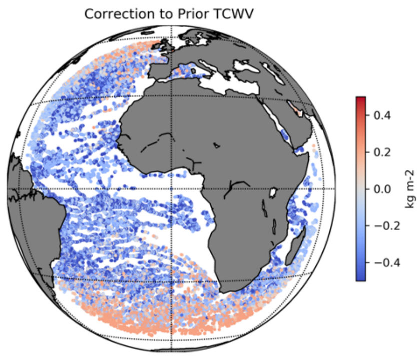
The estimates of $\boldsymbol{\beta}$ were stable (to 0.01 K) after the second cycle. Their final values are shown in Table 2. These values are added to the forward model simulation to bring BT observations and simulations into agreement on average. This means the apparent calibration of the observations relative to the simulation is marginally cooler (by ~0.04 K in all three channels) for QL 4 than for QL 5. Neither the instrument nor the simulation “know” about quality level, so this discrepancy arises from another factor. The sign of the discrepancy is consistent with more residual cloud contamination affecting the nominally clear QL 4 BTs than the QL 5 BTs, which is plausible. The smallness of the discrepancy supports the designation QL 4 and QL 5 pixels as ‘good’ and ‘excellent’ for SST retrieval.

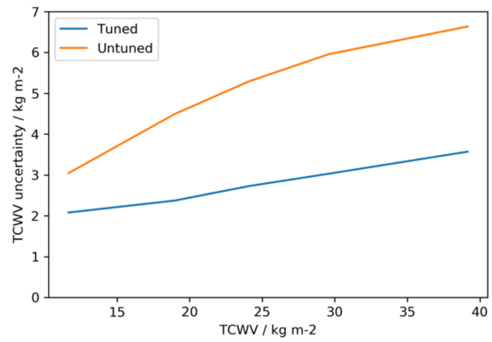
Table 2. Estimated observation-simulation bias / K

	8.7 μm	10.8 μm	12.0 μm
Quality level = 4	0.01	0.00	0.02
Quality level = 5	0.05	0.04	0.06

The bias corrections to be added to the prior TCWV, γ_w, and the prior TCWV error variance (an element of $\boldsymbol{S}_a$) were estimated as a function of TCWV and QL, with the results shown in Figure 3. The estimated prior TCWV biases are modest, typically 1% of the prior TCWV. For QL 5, the correction is always to reduce the prior TCWV, at all latitudes. For QL 4, the correction is less negative in humid latitudes and is positive for the driest locations. The expectation is that the prior TCWV should need a modest negative correction, for the following reason. The NWP humidity fields represent model-cell averages, including the fraction of the cell that is cloudy and where the air is saturated. SST retrievals are made only where skies are clear and therefore where humidity is less than the local average including clouds. This is consistent with the QL 5 result. The error in prior TCWV cannot in reality depend on the quality level of a satellite observation, so the interpretation of the QL 4 result is attribution to prior TCWV bias of an unrepresented factor, that differentially affects QL 4 compared to QL 5. Pixels with QL 4 are (by design) more likely to be subject to residual influences of uncleared clouds. Where the spectral signature of residual cloud across the three thermal channels is similar to that of additional water vapour, such pixels are both more difficult to detect and screen (because truly water-vapour influenced pixels must be retained for the retrieval) and more likely to appear to the retrieval to have high TCWV. Residual cloud contamination of this sort is most likely to arise close to identifiable clouds, and proximity to identified clouds is a criterion for flagging a pixel as QL 4 rather than QL 5.

The uncertainty estimate for the prior TCWV is an order of magnitude greater than the bias, and increases linearly with TCWV. The new, initial prior TCWV uncertainty parameterisation corresponds on average to a fractional uncertainty of around 12%. This is around half of that assumed in previous work (Merchant *et al.*, 2013), and the new estimate is more credible. An estimate for the uncertainty in the skin-adjusted drifting buoy SST as an estimate of the SEVIRI pixel-area skin SST is also obtained from estimating the prior error covariance matrix. This match uncertainty is on average 0.25 K, which is plausible in the context of a buoy SST measurement uncertainty of 0.2 K augmented by unaccounted-for variability in the skin effect and in the difference between SST at the point measurement and over the SEVIRI pixel footprint. Consistent with this interpretation, the estimated match uncertainty increases towards the limb view (not shown), where the pixels are larger and point-to-pixel variability increases.



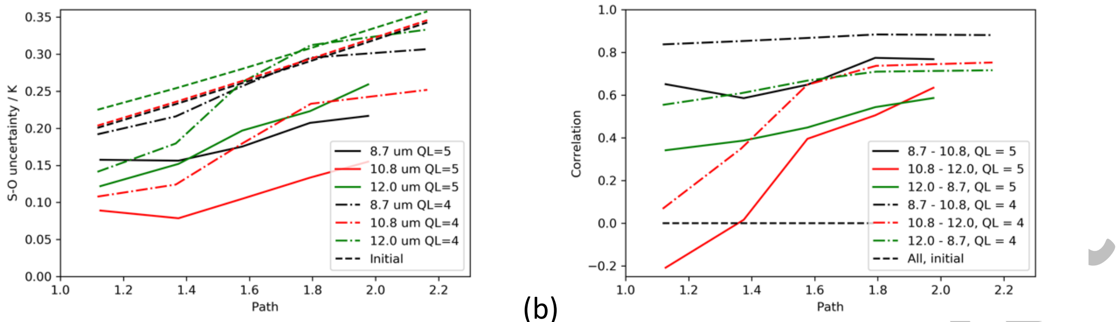


(a) (b)

Figure 3. Biases and uncertainty in prior total column water vapour. (a) Estimated prior TCWV bias per match. (b) Red line: uncertainty in prior TCWV as previously assumed (Merchant et al., 2009b) as a function of TCWV. Blue line: new estimate of the uncertainty in prior TCWV.

The parameters of the observation error covariance matrix, $\boldsymbol{S}_{\epsilon}$, are shown decomposed into uncertainty and error correlations in Figure 4. Instrument noise contributes to the uncertainty but is not dependent on the path, whereas the evaluated uncertainty generally increases with the path. This is consistent with the expectation that the uncertainty in the RTTOV simulations increases when simulating radiative transfer through a greater optical depth of atmosphere. (The reason for the slight upturn in uncertainty for the lowest paths in the 8.7 and 12.0 μm channels at QL 5 may be confounding between the path and the locations of highest TCWV, which tend to occur disproportionately near the satellite nadir.) The 8.7 μm channel has relatively high uncertainty near-nadir, which implies the RTTOV model is not as effective at modelling this channel in humid atmospheres as it is at modelling the others. The evaluated uncertainty is greater for QL 4 than for QL 5. This likely reflects the tendency for the lower quality level observations to be more influenced by residual cloud contamination or atmospheric aerosol. Neither of these factors is simulated by the forward model, and to the degree they are present, they add some variability to the difference $\boldsymbol{y} - \boldsymbol{F}$, which then appears as uncertainty. In general, the previous assumptions about noise and simulation uncertainties were pessimistic, and the new estimates indicate lower uncertainty.

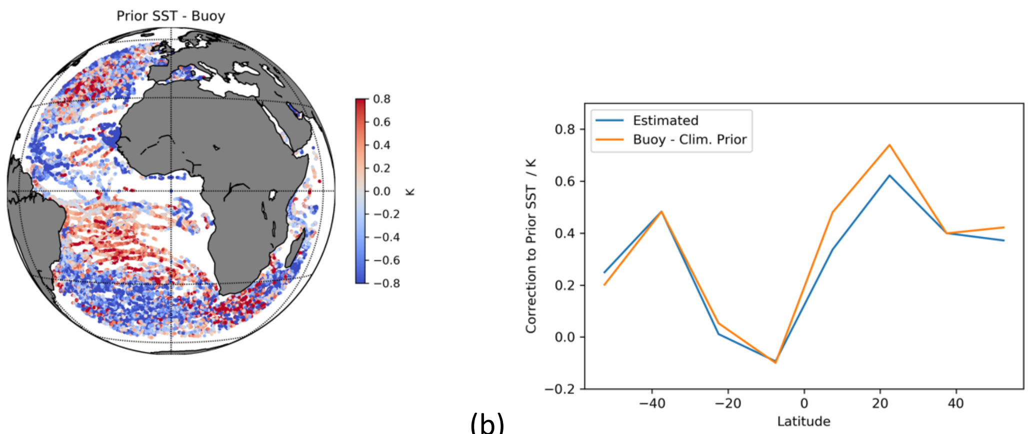
Cross-channel correlations of simulation-observation errors between RTTOV and SEVIRI channels at large satellite zenith angle have previously been inferred by an independent method using residuals from assimilation of SEVIRI data (Waller *et al.*, 2016a), and we obtain numerically similar results here. Towards the edge of the usable disk, errors in all pairs of window channels are correlated with coefficients about 0.7. Failing to account for such correlation (by assuming a diagonal observation error covariance matrix, as done in previous implementations of OE for SST) leads to sub-optimal solutions and underestimation of retrieval uncertainty, so this confirmation is valuable.



(a) (b)

Figure 4. Properties of observation errors. (a) Uncertainty as a function of path (the secant of the satellite zenith angle). (b) Inter-channel correlation of errors.

The bias and uncertainty of the climatological SST used as prior for the initial retrievals were estimated in a single pass of bias estimation and application of Eq. 8. Although not used in this process or in the initial retrievals, drifting buoy SSTs are available in the dataset to quantify the prior SST bias and uncertainty to a good approximation. Figure 5 shows the differences of the prior SST and drifting buoys (accounting for skin effect). Some regional effects are visible, such as cold bias of the prior in the east tropical Atlantic, associated with desert dust outbreaks. However, the dominant variation is latitudinal, and the prior SST correction was estimated in latitudinal bands 15° wide. The validation of the estimate using the differences to drifting buoys in the same latitudinal bands confirms that the prior SST correction is usefully estimated, with much of the latitudinal variation captured to within 0.1 K. This demonstrates that, having bias corrected the SEVIRI radiances, prior SST biases can be estimated independently of the presence of in situ measurements (i.e., valid results can be obtained also in the areas where drifting buoy data are absent).



(a) (b)

Figure 5. *Characteristics of climatological prior SST. (a) Differences of prior SST from matched drifting buoy SST (accounting for skin effect). (b) Blue line: estimated correction of prior SST stratified in bands of 15 ° of latitude (evaluated without in situ references). Red line: the mean prior SST minus drifter SST difference in the same latitudinal bands.*

Table 3 shows the performance of the initial optimal estimator compared to other retrievals and the prior SST using a number of metrics. These include the mean, difference and robust standard deviation (RSD) of the retrieved SST compared to buoy SST (adjusting for the skin effect). The RSD is the median absolute deviation, scaled to be equivalent to standard deviation for a Gaussian distribution. Next is shown the retrieval sensitivity (Merchant *et al.*, 2009a), which indicates the fraction of local SST variability (across fronts or from diurnal cycling of temperature) captured by the retrieval. Sensitivity should ideally be 100%. Lastly, the uncertainty estimates that are obtained for the OE retrievals are assessed via the standard deviation of satellite-buoy differences normalised by the estimated uncertainty in the difference (combining the uncertainty estimate for the retrieval and that of the buoy matched SST). To calculate this metric stably, a trimmed standard deviation is used, excluding a small fraction (0.2%) of outliers beyond five standard deviations. When the retrieved SST uncertainty estimates are ideal, this ratio is 1. As well as the uninitial and the newly initial OE results, the results for the operational algorithm for SEVIRI SST (Le Borgne *et al.*, 2011) are given for comparison. (No sensitivity or uncertainty evaluation is available for this algorithm.)

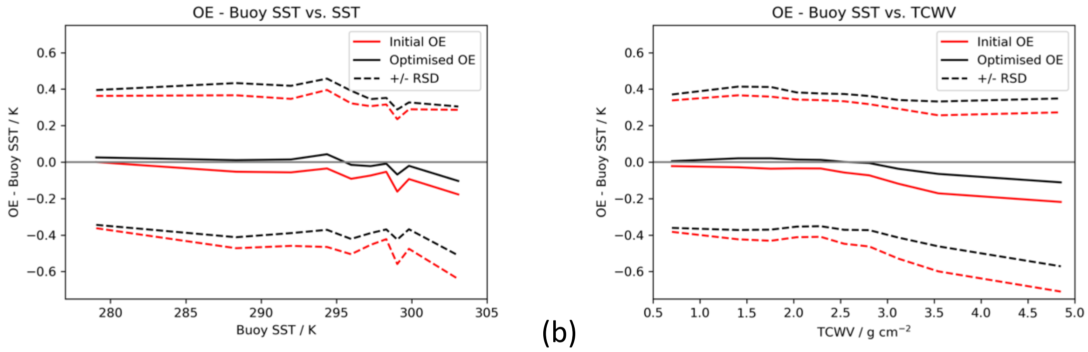
The operational algorithm gives low (<0.1 K) bias and good metrics of scatter against drifting buoys. The initial OE has marginally smaller scatter, and a negative bias of -0.08 K that is larger than the operational algorithm's results, but is still low. In comparison, the optimised OE has negligible bias and scatter that is further improved. The optimised OE improves the validation statistics and simultaneously improves SST sensitivity; this combination is the mark of a valid improvement in retrieval (Petrenko *et al.*, 2014). The improvement in bias reflects the use of the bias corrections. The reduction in standard deviation also comes in part from the bias corrections and from the adjusted balance between prior and observation error covariances. The increased sensitivity arises from the reduced magnitude of the observation uncertainties. Estimates of SST uncertainty are significantly more realistic than before, being 5% pessimistic rather than 20% optimistic. This reflects the smaller, more realistic, error covariance matrices that have been estimated.

Table 3. Comparison of retrieval results via several metrics.

Retrieval	*Mean diff.* / *K*	*SD. diff.* / *K*	*RSD. diff.* / *K*	$\partial\hat{x}/\partial x_{true}$	$SD\left(\frac{\hat{x}-x_b}{\sqrt{S_{\hat{x}}+u_{x_b}^2}}\right)$
Clim. (prior)	-0.16	0.78	0.73	0%	-
Operational	-0.03	0.48	0.42	-	-
Initial OE	-0.08	0.47	0.40	71%	0.80
Optimised OE	-0.01	0.45	0.38	76%	1.05

The results in Table 3 show that, given the operational bias correction, OE initialised from climatology is comparable to the operational retrieval. The parameter retrieval process leads to improved retrievals with less bias, smaller standard deviation when validated on independent data, and improved retrieval sensitivity. Other than the improvement in the retrieval uncertainty estimate, the retrieval improvements are fairly modest. This reflects that in the initial OE formulation, both the observation and prior error covariances were over-estimated by a similar factor. The SST solutions obtained, which represent an optimal compromise between the prior and added information, are broadly similar between the initial and new OE formulations. Nonetheless, the reduction in robust SD corresponds to removal of an independent uncertainty of 0.12 K. The evaluation of the uncertainty in these SST solutions, is, in contrast, significantly changed and improved.

Within the initial OE, the 71641 matches with QL = 4 are biased on average by -0.11 K, and by -0.06 K for the 81753 matches with QL = 5. Since the channel bias corrections are stratified by QL, the relative bias between quality levels is negligible for the optimised OE. The independent uncertainty reduction is similar for QL = 4 and 5.



(a) (b)

Figure 6. Statistics of OE minus buoy SST as a function of (a) buoy SST and (b) prior TCWV. Red lines: initial OE retrieval. Black lines: for optimised OE retrieval. Solid: mean difference. Dashed: mean plus/minus robust standard deviation of difference. Statistics are calculated for deciles of the variable along the abscissa.

Statistics of OE minus buoy SSTs are shown for deciles of SST and TCWV in Figure 6. Against both these factors, a similar pattern of improvement is seen for the optimised OE compared to the initial OE, which reflects that SST and TCWV are well correlated. For the cooler and drier ~50% of matches, the main improvement is reduction of bias of OE relative to buoy SST, while the scatter around the bias is little changed. Larger negative biases are present for warmer and wetter matches in the initial OE results, and these are approximately halved using the optimised OE. The scatter around the mean difference for the warmest and wettest deciles is reduced by 7% to 14%. Overall, the independent uncertainty reduction arises from a combination of reducing functional dependencies in the retrieval bias and reducing retrieval scatter.

5 Discussion and conclusions

This study demonstrates how independent reference data can be used to refine our knowledge of the bias and error covariance parameters needed to obtain best results from optimal estimation. The reference data, available for at least some elements of the state vector being retrieved, enable estimates to be made of biases in non-referenced aspects of the state and of biases in observations relative to the forward model being used for retrieval. This is achieved using a method similar to Kalman filtering for parameter estimation, modified to account for the nature of satellite-reference matched data. Having reduced biases in the system in this way, adapted error-diagnostic relationships are iteratively applied to converge upon parameters describing error covariances. The framework for these parameter estimates systematically integrates available data with knowledge brought to the problem in the form of specifications of the factors likely to be associated with variations in the values of parameters—for example, the expectation that simulations are less precise for off-nadir paths through the atmosphere. In addition to obtaining improvements in retrieval, the estimates of bias and error covariance parameters in themselves provide useful gains in knowledge of the context of the OE retrieval.

In our example application to SST retrieval, the bias parameters obtained for prior water vapour and observed radiances (in different channels and quality levels) have plausible physical interpretations which have been stated. While physical plausibility builds confidence in the outcome, we expect the solutions obtained also to be influenced by other factors. These may include the approximation of using a reduced state space for the retrieval (such that air temperature and water vapour vertical distribution are not retrieved) and the impacts of unmodelled influences on brightness temperature (such as tropospheric aerosol). The choices made about functional dependencies of retrieval parameters also affect the partitioning of bias between different terms. Ultimately, the method is empirical, and it may not always be possible to interpret in terms of likely sources the bias-correction values that are found.

Note also that the bias corrections obtained interact with the radiance calibration, the choice of forward model for simulation, and with cloud detection (as indicated by the different results with respect to QL in our SST retrieval). When these factors change, at least some OE parameters need to be re-estimated to continue to minimise the biases in retrieved quantities.

The results suggest a number of possible directions for further research.

First, we note that the prior correction to the ECMWF NWP humidity fields needed when simulating radiances for only clear-sky areas is not sensor-dependent. Application of the method of estimating this correction using other satellite sensors and in situ references should obtain similar estimates, which would build confidence in their validity.

Second, the ability to estimate parameters for the prior error covariance matrix provides a route to re-visiting the reduced state space used for the SST retrieval. Retrieving only SST and TCWV is an extreme reduction of the state space, since window-channel brightness temperatures are sensitive to a few leading modes of the vertical variability of humidity and temperature (Merchant *et al.*, 2006b), not only to the total amount of water vapour. As noted earlier, any biased prior information relating to the neglected modes will contribute to bias in simulation-observation differences. Since these modes are not retrieved in the reduced-state-space formulation, any such biases cannot be attributed and directly corrected; instead, they are likely folded into the bias parameters obtained. Particularly where three or more channels are used in the retrieval, better solutions may be found with a less restricted state vector. Adding terms to the state vector also requires expanding the prior error covariance matrix, and the approach of this paper may provide a means to obtain a suitable parameterisation for this.

Third, we note that the reference data need not be in situ references, as used here, but could be the retrievals of a different satellite sensor. In the case where a constellation of sensors is in use, each with differing channels, noise and sampling characteristics, it may be relevant to improve the uncertainty of some members of the constellation by bringing them into better consistency with a reference sensor. In the case of SST, dual-view infrared radiometers have been discussed as satellite reference sensors (Donlon *et al.*, 2007) because they are less prone to regional and aerosol-related SST biases (Embury and Merchant, 2012; Embury *et al.*, 2012). In the contemporary context, for SST, the maturing of the drifting buoy and fiducial reference measurement networks (Poli *et al.*, 2019) makes use of in situ references compelling, but in the context of lower global coverage of in situ SST measurements through to the early 2000s, inter-satellite references remain highly relevant for climate data record development. Inter-satellite matches may also be rapidly accumulated at the start of a new mission, enabling rapid inference of OE parameters. Use of reference data from different sources (e.g., satellite and in situ, drifters and ships) with differing uncertainty characteristics is possible within the framework, since the differing uncertainty of the reference measurements can be accounted for (and, indeed, re-estimated).

Finally, we note that the concepts developed here for bias and error covariance parameter estimation using reference data are quite general. Applicability to other variables will depend on factors such as the availability of reference measurements (whether satellite, in situ, or both). Another factor to consider is whether in a particular case, the bias and error covariance properties may be estimated as well or better within a data assimilation system. It is worth noting two differences between parameters estimated within a data assimilation context and relative to independent reference measurements. Biases estimated within a data assimilation system are informed by any in situ measurements that are also assimilated, but additionally reflect any model biases that project through the observation operator. The "observation errors" obtained when applying Desroziers diagnostics in a data assimilation system do not have quite the same meaning as those estimated here, since the uncertainties associated with representativity between the satellite and model grid are additionally convolved with the instrumental and forward model uncertainties.

"Optimal estimation" is a powerful methodology for retrieval. In this paper, we have presented a new approach to using reference data systematically to improve the bias and uncertainty properties of OE retrievals by developing well founded estimates of retrieval parameters, bringing OE closer in practice to the optimality assumed by its underlying theory.

6 Authors' responsibilities

Merchant proposed the study, undertook the quantitative analysis, figure preparation, interpretation and lead writing of the manuscript. Saux-Picart undertook data preparation and forward modelling, advised on analysis implementation and reviewed the manuscript. Waller advised on mathematical formulation and implementation, and reviewed the manuscript.

7 Funding

This work was supported by: the National Centre for Earth Observation (UK) core science programme; the Visiting Scientist programme of the Ocean and Sea Ice Satellite Application Facility of EUMETSAT [OSI_VS18_03]; UK EPSRC Grant No. EP/P002331/1 “Data Assimilation for the Resilient City (DARE); and the European Space Agency Climate Change Initiative for Sea Surface Temperature [4000109848/13/I-NB].

8 References

Auligne, T., McNally, A. P. & Dee, D. P. (2007). Adaptive bias correction for satellite data in a numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society,* **133,** 631-642.

Bormann, N., Bonavita, M., Dragani, R., Eresmaa, R., Matricardi, M. & McNally, A. (2016). Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quarterly Journal of the Royal Meteorological Society,* **142,** 1767-1780.

Buchwitz, M., Reuter, M., Schneising, O., Hewson, W., Detmers, R. G., Boesch, H., Hasekamp, O. P., Aben, I., Bovensmann, H., Burrows, J. P., Butz, A., Chevallier, F., Dils, B., Frankenberg, C., Heymann, J., Lichtenberg, G., De Maziere, M., Notholt, J., Parker, R., Warneke, T., Zehner, C., Griffith, D. W. T., Deutscher, N. M., Kuze, A., Suto, H. & Wunch, D. (2017). Global satellite observations of column-averaged carbon dioxide and methane: The GHG-CCI XCO2 and XCH4 CRDP3 data set. *Remote Sensing of Environment,* **203,** 276-295.

Campbell, W. F., Satterfield, E. A., Ruston, B. & Baker, N. L. (2017). Accounting for Correlated Observation Error in a Dual-Formulation 4D Variational Data Assimilation System. *Monthly Weather Review,* **145,** 1019-1032.

Carboni, E., Mather, T. A., Schmidt, A., Grainger, R. G., Pfeffer, M. A., Ialongo, I. & Theys, N. (2019). Satellite-derived sulfur dioxide (SO2) emissions from the 2014-2015 Holuhraun eruption (Iceland). *Atmospheric Chemistry and Physics,* **19,** 4851-4862.

Cordoba, M., Dance, S. L., Kelly, G. A., Nichols, N. K. & Waller, J. A. (2017). Diagnosing atmospheric motion vector observation errors for an operational high-resolution data assimilation system. *Quarterly Journal of the Royal Meteorological Society,* **143,** 333-341.

Dee, D. P. (2005). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society,* **131,** 3323-3343.

Desroziers, G., Berre, L., Chapnik, B. & Poli, P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society,* **131,** 3385-3396.

Donlon, C., Robinson, I., Casey, K. S., Vazquez-Cuervo, J., Armstrong, E., Arino, O., Gentemann, C., May, D., LeBorgne, P., Piolle, J., Barton, I., Beggs, H., Poulter, D. J. S., Merchant, C. J., Bingham, A., Heinz, S., Harris, A., Wick, G., Emery, B., Minnett, P., Evans, R., Llewellyn-Jones, D., Mutlow, C., Reynolds, R. W., Kawamura, H. & Rayner, N. (2007). The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. *Bulletin of the American Meteorological Society,* **88,** 1197-1213.

Donlon, C. J., Minnett, P. J., Gentemann, C., Nightingale, T. J., Barton, I. J., Ward, B. & Murray, M. J. (2002). Toward improved validation of satellite sea surface skin temperature measurements for climate research. *Journal of Climate,* **15,** 353-369.

Embury, O. & Merchant, C. J. (2012). A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: A new retrieval scheme. *Remote Sensing of Environment,* **116,** 47-61.

Embury, O., Merchant, C. J. & Corlett, G. K. (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.

Heidinger, A. K. (2003). Rapid daytime estimation of cloud properties over a large area from radiance distributions. *Journal of Atmospheric and Oceanic Technology,* **20,** 1237-1250.

JCGM. (2008). *Guide to the expression of Uncertainty in Measurement*.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Trans AMSE, Ser. D, J. Basic Eng.,* **82**.

Koner, P. K., Harris, A. & Maturi, E. (2015). A Physical Deterministic Inverse Method for Operational Satellite Remote Sensing: An Application for Sea Surface Temperature Retrievals. *Ieee Transactions on Geoscience and Remote Sensing,* **53,** 5872-5888.

Le Borgne, P., Roquet, H. & Merchant, C. J. (2011). Estimation of Sea Surface Temperature from the Spinning Enhanced Visible and Infrared Imager, improved using numerical weather prediction. *Remote Sensing of Environment,* **115,** 55-65.

Lean, K. & Saunders, R. W. (2013). Validation of the ATSR Reprocessing for Climate (ARC) Dataset Using Data from Drifting Buoys and a Three-Way Error Analysis. *Journal of Climate,* **26,** 4758-4772.

McGarragh, G. R., Poulsen, C. A., Thomas, G. E., Povey, A. C., Sus, O., Stapelberg, S., Schlundt, C., Proud, S., Christensen, M. W., Stengel, M., Hollmann, R. & Grainger, R.

G. (2018). The Community Cloud retrieval for CLimate (CC4CL) - Part 2: The optimal estimation approach. *Atmospheric Measurement Techniques,* **11,** 3397-3431.

Merchant, C. J., Embury , O., Bulgin , C. E., Block, T., Corlett , G. K., Fiedler , E., Good , S. A., Mittaz , J., Rayner , N. A., Berry , D., Eastwood , S., Taylor , M., Tsushima , Y., Waterfall , A., Wilson , R. & Donlon, C. (2019). Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Scientific Data,* **in press**.

Merchant, C. J., Embury, O., Le Borgne, P. & Bellec, B. (2006a). Saharan dust in nighttime thermal imagery: Detection and reduction of related biases in retrieved sea surface temperature. *Remote Sensing of Environment,* **104,** 15-30.

Merchant, C. J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C. E., Corlett, G. K., Good, S., McLaren, A., Rayner, N., Morak-Bozzo, S. & Donlon, C. (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,** 179-191.

Merchant, C. J., Harris, A. R., Roquet, H. & Le Borgne, P. (2009a). Retrieval characteristics of non-linear sea surface temperature from the Advanced Very High Resolution Radiometer. *Geophysical Research Letters,* **36**.

Merchant, C. J., Horrocks, L. A., Eyre, J. R. & O'Carroll, A. G. (2006b). Retrievals of sea surface temperature from infrared imagery: origin and form of systematic errors. *Quarterly Journal of the Royal Meteorological Society,* **132,** 1205-1223.

Merchant, C. J., Le Borgne, P., Marsouin, A. & Roquet, H. (2008). Optimal estimation of sea surface temperature from split-window observations. *Remote Sensing of Environment,* **112,** 2469-2484.

Merchant, C. J., Le Borgne, P., Roquet, H. & Legendre, G. (2013). Extended optimal estimation techniques for sea surface temperature from the Spinning Enhanced Visible and Infra-Red Imager (SEVIRI). *Remote Sensing of Environment,* **131,** 287-297.

Merchant, C. J., Le Borgne, P., Roquet, H. & Marsouin, A. (2009b). Sea surface temperature from a geostationary satellite by optimal estimation. *Remote Sensing of Environment,* **113,** 445-457.

Munro, R., Siddans, R., Reburn, W. J. & Kerridge, B. J. (1998). Direct measurement of tropospheric ozone distributions from space. *Nature,* **392,** 168-171.

Petrenko, B., Ignatov, A., Kihai, Y., Stroup, J. & Dash, P. (2014). Evaluation and selection of SST regression algorithms for JPSS VIIRS. *Journal of Geophysical Research-Atmospheres,* **119,** 4580-4599.

Poli, P., Lucas, M., Ocarroll, A., Le Menn, M., David, A., Corlett, G. K., Blouch, P., Meldrum, D., Merchan, C. J., Belbeoch, M. & Herklotz, K. (2019). The Copernicus Surface Velocity Platform drifter with Barometer and Reference Sensor for Temperature (SVP-BRST): genesis, design, and initial results. *Ocean Science,* **15,** 199-214.

Poulsen, C. A., Siddans, R., Thomas, G. E., Sayer, A. M., Grainger, R. G., Campmany, E., Dean, S. M., Arnold, C. & Watts, P. D. (2012). Cloud retrievals from satellite data using optimal estimation: evaluation and application to ATSR. *Atmospheric Measurement Techniques,* **5,** 1889-1910.

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

Saunders, P. M. (1967). The Temperature at the Ocean-Air Interface. *Journal of the Atmospheric Sciences,* **24,** 269-273.

Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., Vidot, J., Roquet, P., Matricardi, M., Geer, A., Bormann, N. & Lupu, C. (2018). An update on the RTTOV

fast radiative transfer model (currently at version 12). *Geoscientific Model Development,* **11,** 2717-2732.
Stewart, J. V., Canaria, E. C., Brown, C. E., Yang, A. F., Yang, G. S. & Merchant, C. J. (1997). Toxicity of the synthetic slow release fertilizer, NPK (7-40-0), on various freshwater organisms. *Canadian Technical Report of Fisheries and Aquatic Sciences***,** 110-110.
Stewart, L. M., Dance, S. L., Nichols, N. K., Eyre, J. R. & Cameron, J. (2014). Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Quarterly Journal of the Royal Meteorological Society,* **140,** 1236-1244.
Thomas, G. E., Poulsen, C. A., Sayer, A. M., Marsh, S. H., Dean, S. M., Carboni, E., Siddans, R., Grainger, R. G. & Lawrence, B. N. (2009). The GRAPE aerosol retrieval algorithm. *Atmospheric Measurement Techniques,* **2,** 679-701.
Vitart, F. (2014). Evolution of ECMWF sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society,* **140,** 1889-1899.
Waller, J. A., Ballard, S. P., Dance, S. L., Kelly, G., Nichols, N. K. & Simonin, D. (2016a). Diagnosing Horizontal and Inter-Channel Observation Error Correlations for SEVIRI Observations Using Observation-Minus-Background and Observation-Minus-Analysis Statistics. *Remote Sensing,* **8**.
Waller, J. A., Simonin, D., Dance, S. L., Nichols, N. K. & Ballard, S. P. (2016b). Diagnosing Observation Error Correlations for Doppler Radar Radial Winds in the Met Office UKV Model Using Observation-Minus-Background and Observation-Minus-Analysis Statistics. *Monthly Weather Review,* **144,** 3533-3551.
Weston, P. P., Bell, W. & Eyre, J. R. (2014). Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society,* **140,** 2420-2429.
Wong, E. W. & Minnett, P. J. (2018). The Response of the Ocean Thermal Skin Layer to Variations in Incident Infrared Radiation. *Journal of Geophysical Research-Oceans,* **123,** 2475-2493.

LIST OF FIGURE CAPTIONS

Figure 1. Distribution of satellite-buoy matches used in this study. The locations shown are for 2011, and the distribution in 2012 is similar. Matched locations are coloured with the measured buoy sea surface temperature.

Figure 2. The sequence of estimation of three sets of parameters for optimal estimation. For symbols, see the main text.

Figure 3. Biases and uncertainty in prior total column water vapour. (a) Estimated prior TCWV bias per match. (b) Red line: uncertainty in prior TCWV as previously assumed

(Merchant et al., 2009b) as a function of TCWV. Blue line: new estimate of the uncertainty in prior TCWV.

Figure 4. Properties of observation errors. (a) Uncertainty as a function of path (the secant of the satellite zenith angle). (b) Inter-channel correlation of errors.

Figure 5. Characteristics of climatological prior SST. (a) Differences of prior SST from matched drifting buoy SST (accounting for skin effect). (b) Blue line: estimated correction of prior SST stratified in bands of 15° of latitude (evaluated without in situ references). Red line: the mean prior SST minus drifter SST difference in the same latitudinal bands.

Figure 6. Statistics of OE minus buoy SST as a function of (a) buoy SST and (b) prior TCWV. Red lines: initial OE retrieval. Black lines: for optimised OE retrieval. Solid: mean difference. Dashed: mean plus/minus robust standard deviation of difference. Statistics are calculated for deciles of the variable along the abscissa.