Optimal Ensemble-Based Selection of Channels from Advanced Sounders in the Presence of Cloud

STEFANO MIGLIORINI

European Centre for Medium-Range Weather Forecasts, and National Centre for Earth Observation, University of Reading, Reading, United Kingdom

(Manuscript received 29 July 2014, in final form 21 May 2015)

ABSTRACT

This study aims to illustrate a general procedure based on well-known information theory concepts to select the channels from advanced satellite sounders that are most advantageous to assimilate both in clear-sky and overcast conditions using an ensemble-based estimate of forecast uncertainty. To this end, the standard iterative channel selection method, which is used to select the most informative channels from advanced infrared sounders for operational assimilation, was revisited so as to allow its use with measurements that have correlated errors. The method was here applied to determine a 24-humidity-sensitive-channel set that is small in size relative to a total of 8461 channels that are available on the Infrared Atmospheric Sounding Interferometer (IASI) on board the EUMETSAT Polar System MetOp satellites. The selected channels can be used to perform all-sky data assimilation experiments, in addition to those currently used for operational data assimilation of IASI data at ECMWF. Care was taken to include in the observation uncertainty used for channel selection the contributions arising from imperfect knowledge of the concentration of contaminants (except for cloud) in a given spectral channel. Also, (cumulative) weighting functions that provide a vertically resolved picture of the (total) number of degrees of freedom for signal expressed by a given set of measurements were introduced, which allows for the definition of a novel channel selection merit function that can be used to select measurements that are most sensitive to variations of a given parameter over a given atmospheric region (e.g., in the troposphere).

1. Introduction

Over the last decade or recent decades there has been a formidable increase in the amount of data that is being acquired by satellite sounding instruments and disseminated to operational meteorological centers for assimilation, particularly in the infrared spectral region. At ECMWF, the infrared sounding instruments that are currently monitored or assimilated are the High Resolution Infrared Radiation Sounder (HIRS), on board the EUMETSAT Polar System MetOp polar-orbiting satellites, with 20 channels; the Advanced Infrared Sounder (AIRS) on board Aqua and measuring over 2378 channels; the Infrared Atmospheric Sounding Interferometer (IASI) also on board MetOp with 8461 channels; and the Cross-track Infrared Sounder (CrIS) on board the Suomi–National Polar-Orbiting Partnership (Suomi-NPP) satellite, with 1305 channels. In each case only a subset of channels are assimilated.

To be able to exploit such a wealth of data, operational centers had to overcome numerous technological and scientific challenges, including making appropriate choices about which subset of channels from each instrument to consider for assimilation. The problem was put on firm theoretical grounds by Rodgers (1996), who described an iterative method to determine an optimal set of channels by maximizing a figure of merit based on their Shannon information content. In particular, according to this method, a new channel is selected if it provides the largest information increment with respect to the information content already provided by all previously selected channels, in the assumption that the considered measurements have mutually independent errors. Hereafter, iterative channel selection methods that assume a diagonal measurement error covariance will be referred to as (iterative) sequential methods. A subsequent study (Rabier et al. 2002) showed that the
iterative sequential method provided a more effective channel selection (i.e., a larger state estimate error reduction) than other existing methods based on a non-iterative approach (see, e.g., Prunet et al. 1998). As recognized in Rabier et al. (2002), a likely reason for the shortcomings of the noniterative methods is their difficulties in providing a selection of channels that are informative over atmospheric partial columns at different height ranges and that are representative of different spectral regions.

The iterative sequential channel selection method—after appropriate prescreening of channels with too large forward-model uncertainty or with characteristics that make them more sensitive to misspecifications of forecast error uncertainty in observation space (e.g., with multiple gas sensitivities or with Jacobians that have multiple peaks or long tails)—was used at ECMWF (Collard 2007) to determine an optimal set of (currently 373) IASI channels sensitive to atmospheric temperature, water vapor, ozone, and surface conditions in the clear sky for operational monitoring or assimilation. The impact on analyses and forecasts of the selected IASI channels with respect to the ECMWF operational system at the time of the experiments is discussed in Collard and McNally (2009). The same channel selection methodology was also used to select the most informative channels from AIRS (Fourrié and Thépaut 2003) in the clear sky.

The aim of this work is first to use channel selection techniques to check whether the informative potential on key atmospheric variables of a set of instrument channels changes when cloud is present in the instrument field of view, and then to select a set of channels that provide significant information on those variables (in this case atmospheric humidity), both in clear-sky and overcast conditions. A recent study (Martinet et al. 2014) also investigated the use of the iterative channel selection technique to complement the IASI channels, which are already in use for operational assimilation of IASI data in clear-sky conditions (see Collard and McNally 2009), with additional channels that are most effective for the joint retrieval of ice and liquid water content using IASI data without solar contamination. The impact of the additional channels on water vapor estimates was also assessed. The authors found that the additional channels provided at best only marginal improvements with respect to the case when only the standard channels are used in the retrieval. Note, however, that although in both studies the state vector is augmented with cloud-related fields, a key difference between this work and that discussed in Martinet et al. (2014) is that the sensitivity of channel selection results on cloud is assessed in this study in the assumption that the cloud fields (cloud fraction and liquid and ice water contents) are not part of the data assimilation system control vector, so as to be consistent with the assumptions currently made within operational numerical weather prediction (NWP) data assimilation systems.

Also, in this study the iterative channel selection methodology was revisited and modified to be used in a consistent way with observations having correlated errors. This novel formulation of the iterative—and “non-sequential” [i.e., without the sequential updating of the forecast error covariance matrix discussed in Rodgers (1996)]—selection method was then used to select the most effective IASI channels for the estimation of atmospheric water vapor profiles both in clear-sky and overcast conditions. To this end, an ensemble-based estimate of forecast errors, derived from an ECMWF's ensemble of data assimilations (EDA) run on a 91-level and 50-member configuration, was used for a case study during summer 2012. It is important to note that the main aim of this work is not to replace existing sets of IASI channels selected for assimilation in clear sky, but rather to determine a relatively small number of additional channels that can provide the largest impact on meteorological analyses in all-sky conditions. The present inability of including cloud fields in the control vector led to the decision of focusing on humidity-sensitive channels, as less affected by misspecification of cloud forecast fields than those mainly sensitive to temperature. This means that the IASI channels selected in this study are considered to be best suited to assimilate water vapor sensitive observations of radiation emerging from either a clear-sky or a cloud-affected scene with a single observation operator that includes a parameterization of multiple scattering by clouds and no need for cloud detection (see, e.g., Bauer et al. 2010).

The paper is structured as follows. Section 2 provides a detailed description of the channel selection methodology and a step-by-step algorithm. Also in this section, the standard information-based figure of merit used for selection is extended to allow a selection that is optimal for estimation over a subspace of the state space (e.g., over a given height range or a given parameter). In section 3 a description of the case study is provided. Section 4 discusses the effects of the chosen forecast and observation error as well as of the observation operator specifications on the signal-to-noise characteristics of the satellite instrument, while section 5 provides details on the selection of optimal channels for atmospheric humidity estimation in all-sky conditions as resulting from the use of the selection method described in this paper, including a list of the selected channels. Finally, a summary of the work and its main conclusions are given.
section 6, while the appendix provides the details of the computational costs of a sequential version of the channel selection algorithm discussed in this paper.

2. Iterative channel selection with correlated observation errors

The channel selection method as described in Rodgers (1996) is based on finding the channel that, at each iteration, provides the largest increment to the number of degrees of freedom for signal (DFS) already provided by the previously selected channels. This procedure is repeated until the required number of channels have been selected. To reduce the computational costs of the iterative selection process, the original algorithm also assumes that the measurement error covariance for the considered channels is diagonal. In this way it is possible to calculate the maximum a posteriori retrieval error covariance found when making use of a set of \( k \) measurements [see, e.g., Rodgers (2000), his section 5.4] as an update of the retrieval error covariance valid for a set of \( k - 1 \) measurements.

Radiance data calculated with fast-forward models used for operational assimilation, however, can have spectrally correlated errors (Matricardi 2010), with spectral distances that can be significantly larger than those due to apodization, which only involves adjacent channels. Other relevant sources of interchannel error correlation include the variations of atmospheric species such as water vapor or ozone when selecting temperature-sensitive channels, errors arising from shortcomings in accounting for cloud as well as surface emissivity errors, and representativeness errors (Bormann et al. 2010). A recent channel selection study (Ventress and Dudhia 2014) investigated a way to account for observation-error correlations arising from imperfect knowledge of the concentration of unconstrained (i.e., not retrieved) atmospheric constituents with absorption lines in the spectral regions that are sampled by the set of channels considered for selection. Observation errors are expressed in Ventress and Dudhia (2014) as a combination of random and systematic components, with the random component being assumed as spectrally uncorrelated and as the only observation-error component that is relevant to update the retrieval error covariance calculated using the previously selected channels. With this assumption it is still possible to make use of the sequential method to update the retrieval error covariance, with some computational savings. At the same time, both observation-error components—the diagonal random error and the spectrally correlated systematic error components—are considered to compute the information-content-based figure of merit used for channel selection.

The iterative channel selection method used here avoids the use of the sequential covariance update formula and can then consistently be used in the presence of correlated observation errors with no simplifying assumptions. The increased computational costs incurred when avoiding the sequential update proved to be affordable, also considering that the method is not supposed to be used for near-real-time applications. But before discussing the algorithm further let us define some relevant quantities. To retrieve an estimate of \( \mathbf{x} \in \mathbb{R}^m \), the true state of the system, we can make use of a set of measurements that are assumed to be components of the measurement vector \( \mathbf{y} \in \mathbb{R}^m \). The relationship between the measurement vector and the state vector can be written as

\[
\mathbf{y} = \mathbf{H}(\mathbf{x}) + \mathbf{e}_o,
\]

where \( \mathbf{H}(\mathbf{x}) \in \mathbb{R}^{m \times n} \) is the observation operator and \( \mathbf{e}_o \in \mathbb{R}^m \) is an additive observation-error vector, assumed to be unbiased and Gaussian distributed with nonsingular covariance \( \mathbf{R}_m \). If the observation operator is approximately linear in a given region of the state space around \( \mathbf{x}_0 \), it is meaningful to approximate Eq. (1) as

\[
\mathbf{y} \approx \mathbf{H}(\mathbf{x}_0) + \mathbf{H}(\mathbf{x} - \mathbf{x}_0) + \mathbf{e}_o,
\]

where \( \mathbf{H} = (\partial \mathbf{H}/\partial \mathbf{x}) \) calculated at \( \mathbf{x}_0 \) is the Jacobian matrix. In this case, we can define \( \mathbf{y} \) as [see Rodgers (2000), his section 8.3]

\[
\mathbf{y} = \mathbf{y}^o - \mathbf{H}(\mathbf{x}_0) + \mathbf{H}(\mathbf{x} - \mathbf{x}_0) \approx \mathbf{H}(\mathbf{x} - \mathbf{x}_0) + \mathbf{e}_o.
\]

The algorithm used in this work is based on maximizing a figure of merit given by the number of DFS. The total number of DFS using a given set of measurements, \( d_s \), is given by [e.g., Rodgers (1996), his section 2.5]

\[
d_s = \text{tr}(\mathbf{I}_n - \mathbf{P}^f \mathbf{B}^{-1}) = \text{tr}(\mathbf{A}),
\]

where \( \mathbf{P}^f \in \mathbb{R}^{n \times n} \) is the retrieval error covariance matrix when a measurement vector with \( m \) components is used in the retrieval, \( \mathbf{B} \in \mathbb{R}^{n \times n} \) is the forecast (also denoted as prior or background) error covariance matrix (assumed nonsingular), and \( \mathbf{I}_n \) is the identity matrix of order \( n \) and where \( \mathbf{A} \in \mathbb{R}^{n \times n} \) is the so-called averaging kernel matrix. Let us now introduce an extension of the DFS-based figure of merit that may be used to maximize the number of DFS over a given number of state vector components, that is, over a given range of model levels (e.g., in the troposphere) or a given state vector parameter (e.g., specific humidity or the profile of a given atmospheric chemical compound of interest).
a. DFS weighting functions and the effective DFS

If now \( B \) is expressed as \( B = XX^T \) we can define the signal-to-noise matrix \( S \) of the considered measurements as \( S = R^{-1/2}HX \) [e.g., Rodgers (2000), his section 2.4]. The signal-to-noise matrix can be expressed in terms of its singular value decomposition as \( S = U \Lambda V^T \), where \( U \in \mathbb{R}^{m \times m} \) and \( V \in \mathbb{R}^{m \times n} \) are orthogonal matrices whose columns are the left and right singular vectors of \( S \), respectively, and where \( \Lambda \in \mathbb{R}^{m \times n} \) is the matrix whose nonzero elements have the same row and column indices and are given by the singular values \( \lambda_j \) of \( S \), with \( j = 1, \ldots, r \) where \( r \) is the rank of \( S \), with \( r \leq \min(m, n) \).

It is possible to show (see Migliorini 2013) that the retrieval error covariance \( \mathbf{P}^r \) can be expressed in terms of the singular values and the right singular vectors of \( S \) as

\[
\mathbf{P}^r = \mathbf{XV} \begin{bmatrix} (\Lambda^2_j + 1)_{1 \times (n-r)} & 0_{r \times (n-r)} \\ 0_{(n-r) \times r} & 1_{n-r} \end{bmatrix} \mathbf{V}^T \mathbf{X}^T
\]

\[
= \sum_{j=1}^{r} \frac{n}{1 + \lambda_j^2} \mathbf{Xv}_j \mathbf{v}_j^T \mathbf{X}^T. \tag{5}
\]

Noting that \( \sum_{j=1}^{n} \mathbf{v}_j \mathbf{v}_j^T = I_r \), Eq. (5) can be written as

\[
\mathbf{P}^r = \mathbf{B} - \sum_{j=1}^{r} \frac{n}{1 + \lambda_j^2} \mathbf{Xv}_j \mathbf{v}_j^T \mathbf{X}^T. \tag{6}
\]

From Eqs. (6) and (4) we can write

\[
d_j = \text{tr} \left( \sum_{j=1}^{r} \frac{\lambda_j^2}{1 + \lambda_j^2} \mathbf{Xv}_j \mathbf{v}_j^T \mathbf{X}^T \mathbf{B}^{-1} \right)
\]

\[
= \text{tr} \left( \sum_{j=1}^{r} d_{sj} \mathbf{v}_j \mathbf{v}_j^T \mathbf{X}^T \mathbf{B}^{-1} \right)
\]

\[
= \sum_{j=1}^{r} d_{sj} \text{tr}(\mathbf{v}_j \mathbf{v}_j^T \mathbf{X}^T \mathbf{B}^{-1}), \tag{7}
\]

as \( d_j \) can also be expressed as [see Migliorini (2013), his section 4] \( d_j = \sum_{j=1}^{r} \lambda_j^2 / (1 + \lambda_j^2) = \sum_{j=1}^{r} d_{sj} \). As the trace of a matrix is invariant under similarity transformations [e.g., Golub and Van Loan (2013), their section 7.1.1] Eq. (7) can be written as

\[
d_j = \sum_{j=1}^{r} d_{sj} \text{tr}(\mathbf{v}_j \mathbf{v}_j^T) = \sum_{j=1}^{r} d_{sj} \mathbf{v}_j \mathbf{v}_j^T. \tag{8}
\]

Noting that \( \mathbf{v}_j \mathbf{v}_j^T = 1 \) as \( \mathbf{v}_j \) is an orthonormal basis vector, the expression in Eq. (7) linking the DFS in observation space with the DFS in state space represents a trivial identity that follows from the fact that the two quantities are equivalent [e.g., Rodgers (2000), his section 2.5]. However, it is possible to make use of the diagonal elements of the matrix \( d_j \mathbf{v}_j \mathbf{v}_j^T \) given by

\[
s_j = \text{diag}\left( d_j \mathbf{v}_j \mathbf{v}_j^T \right) = d_j (v_{j1}^2, v_{j2}^2, \ldots, v_{jn}^2) = d_j \mathbf{v}_j \circ \mathbf{v}_j, \tag{9}
\]

where \( \circ \) denotes the elementwise (or Schur) product, in order to determine how the number \( d_j \) of DFS are vertically distributed for each of the variables included in the state vector. From Eq. (9) it follows that the vertical distribution of the total number of DFS for a given model variable is given by

\[
s = \sum_{j=1}^{r} s_j = \left( \sum_{j=1}^{r} d_{sj} v_{j1}^2, \sum_{j=1}^{r} d_{sj} v_{j2}^2, \ldots, \sum_{j=1}^{r} d_{sj} v_{jn}^2 \right). \tag{10}
\]

In other words, the quantity \( d_{sj} v_{j1}^2 = (s_j)_k \) represents the fraction of the number of DFS \( d_{sj} \) along the \( \mathbf{v}_j \) direction in the state space that is achieved on the \( k \)th component of the state vector [i.e., on a given model variable (e.g., temperature) at a given model level] while \( \sum_{j=1}^{r} d_{sj} v_{jk}^2 = s_k \) is the total number of DFS achieved on the same model variable at the chosen model level. The components of \( s_j \) and \( s \) define the \( j \)th DFS weighting function and the cumulative DFS weighting function, respectively, along all relevant model variables and model levels considered in the state vector. Note that, according to the above definitions, the sum of the (always nonnegative) elements of \( s \) coincides with the trace of \( \mathbf{A} \).

But while the \( k \)th element \( a_{kk} \) of the diagonal of the averaging kernel matrix provides the (possibly negative) sensitivity of the \( k \)th component of the retrieval vector to infinitesimal variations of \( k \)th component of the state vector [see, e.g., Rodgers (2000), his section 3.1.4], the \( k \)th element of the cumulative DFS weighting function (always nonnegative) provides instead the total number of DFS retrieved on that component of the state. To compare the two quantities for a specific example, the temperature averaging kernels and the DFS weighting functions at a clear-sky location for the case study discussed in section 3 and the 373 IASI channels currently monitored at ECMWF are shown in Fig. 1. From Fig. 1 it follows that the atmospheric layers where the DFS weighting functions are largest coincide with those where the averaging kernels have the largest magnitude. In particular, the cumulative DFS weighting function and the diagonal of the averaging kernel matrix take up very similar values except close to the surface.

It is also useful to define the quantity \( d_{sj}^{\text{eff}} \) given by the sum of a subset of the components of \( s \) over a chosen region of the state space, which can be written as

\[
d_{sj}^{\text{eff}} = \sum_k s_{kj} = \sum_{j=1}^{r} d_{sj} \sum_k v_{kj}^2, \tag{11}
\]
which represents the effective degrees of freedom for signal achieved by a set of measurements over that region of the state space, where the index $k$ identifies the (not necessarily adjacent) components of a given parameter of interest (e.g., atmospheric humidity) over some atmospheric (partial) column(s) of interest. For example, for the location and time as in Fig. 1 the 373 IASI channels currently monitored at ECMWF provide 4.479 DFS for humidity and 0.0038 for ozone over the whole atmosphere. As discussed in the next section, the (effective) degrees of freedom for signal can be used, for example, as a figure of merit for channel selection in order to select the set of $n_{sel}$ channels that provide most information about the whole (selected region of the) state space. Note that the minimum and maximum averaging kernel values (found close to or at the surface) are equal to $-0.267$ and $0.989$, respectively, and the surface value of the cumulative DFS weighting function is equal to 0.372. The total number of DFS for temperature is 4.143.

![Fig. 1.](image-url)

Finally, it is useful to compare the newly introduced DFS weighting function with the familiar Jacobian defined as a given row of $\mathbf{H}$ in Eqs. (1) and (2). To determine the region of the state space a given instrument can sense it is necessary to explore the Jacobians for all instrument channels of interest to check where they are significantly different from zero and possibly where they peak. This region of the state space, however, in general does not coincide with that on which a given sounder can provide most of its information, which is also a function of the observation and forecast error covariance matrices used for assimilation. Also, the Jacobians can be negative and this may cause confusion when the Jacobians are used as a measure of vertical resolution for a given variable. The cumulative DFS weighting function of a set of $\lambda_i$ is used instead. In particular, the same set of channels are selected when using the Shannon information content $I$ defined as [see, e.g., Rodgers (2000), his section 2.5.2] $I = 0.5 \sum_i \ln(1 + \lambda_i^2)$. 
s includes contributions from the Jacobians of all considered channels and then it provides a concise depiction of the information that a given instrument can acquire on different vertical atmospheric layers for different model variables. Also, from the fact that the sum of the components of s is equal to \( d_i \), it follows that the sum of the elements of s over a given atmospheric partial column and model variable quantifies the number of DFS that the instrument can provide on the chosen model variable over that region. As explained above, however, the characteristics of the DFS weighting functions vary according to the forecast error uncertainty used for retrieval or assimilation so that they are not only linked to the instrument specifications.

b. Description of the selection algorithm

An iterative, nonsequential, channel selection algorithm can be devised as follows. At the first iteration step \( l = 1 \) the instrument channel \( i \) is considered and the signal-to-noise matrix \( S_{l=1,j} \in \mathbb{R}^{1 \times n} \) is calculated as

\[
S_{l=1,j} = \frac{h_j^T}{\sigma^2_i} X,
\]

(12)

where \( h_j^T \) and \( \sigma^2_i \) are the row of \( H \) and the observation-error standard deviation corresponding to the measurement channel \( i \), respectively. Then the Gramian \( G_{l=1,j} \in \mathbb{R} \) of \( S_{l=1,j} \), given by

\[
g_{l=1,j} = S_{l=1,j} S_{l=1,j}^T = \lambda^2_{l=1,j=1},
\]

(13)

is computed, where the index \( j \) orders the singular values of \( S_{l=1,j} \). Then the number \( d_{s_{l=1,j}} \) of DFS for channel \( i \) is calculated as

\[
d_{s_{l=1,j}} = \frac{\lambda^2_{l=1,j=1}}{1 + \lambda^2_{l=1,j=1}}. \tag{14}
\]

At iteration step \( l = 1 \) the Gramian \( G_{l=1,j} \) and the number \( d_{s_{l=1,j}} \) of DFS are calculated \( m \) times, with \( i = 1, \ldots, m \). It is important to note that these computations can be performed in parallel, in the same way as when sequential channel selection methods are used. The channel selected at \( l = 1 \) is the channel \( i_1 \) with \( d_{s_{l=1,j}} = \max(d_{s_{l=1,j}}) \), with \( i = 1, \ldots, m \).

The iteration step \( l = 2 \) consists in calculating \( S_{l=2,j} \in \mathbb{R}^{2 \times n} \) as

\[
S_{l=2,j} = R_{l=2,j}^{-1/2} \begin{pmatrix} h_j^T \end{pmatrix} X,
\]

(15)

where \( R_{l=2,j} \in \mathbb{R}^{2 \times 2} \) is a submatrix of the observation-error covariance \( R \) defined as

\[
R_{l=2,i} = \begin{bmatrix} \alpha^2_i r_{i_1} & r_{i_1,i} \alpha^2_i \sigma^2_i \\ r_{i_1,i} \sigma^2_i & \sigma^2_i \end{bmatrix} = L \Gamma L^T
\]

(16)

and \( R_{l=2,j}^{-1/2} = L \Gamma^{-1/2} L^T \). The Gramian matrix \( G_{l=2,j} \in \mathbb{R}^{2 \times 2} \), with \( i \neq i_1 \), is then given by

\[
G_{l=2,j} = S_{l=2,j} S_{l=2,j}^T = U_{l=2,j} \Lambda^2_{l=2,j} U_{l=2,j}^T. \tag{17}
\]

Then the number of DFS for channel \( i \neq i_1 \) is calculated as

\[
d_{s_{l=2,j}} = \frac{\lambda^2_{l=2,i,j=1}}{1 + \lambda^2_{l=2,i,j=1}}.
\]

(18)

with \( r \leq \min(2, n) \). At iteration step \( l = 2 \) the Gramian matrix \( G_{l=2,j} \) and the number of DFS \( d_{s_{l=2,j}} \) are calculated \( m - 1 \) times, with \( i = 1, \ldots, m \neq i_1 \). The channel selected at \( l = 2 \) is the channel \( i_2 \) with \( d_{s_{l=2,j}} = \max(d_{s_{l=2,j}}) \), with \( i = 1, \ldots, m \neq i_1 \).

The algorithm is iterated until \( l = n_{sel} \) channels are selected. Note that if the state vector \( x \) is composed of two independent components \( x_1 \) and \( x_2 \)—say temperature and humidity—with background error covariance matrix \( B \) block diagonal over the subspaces defined by \( x_1 \) and \( x_2 \), then the signal-to-noise matrix \( S \) can be calculated as \( S = (S_{x_1}, S_{x_2}) \). In the appendix it is shown that a sequential version of the channel selection algorithm discussed above makes use of significantly fewer computational resources than the standard sequential-update algorithm [e.g., Rodgers (2000), his section 5.8.1.3].

Finally, note that when there is the need to select channels that are particularly suited to reduce retrieval errors over a given region of the state space, it is possible to replace \( d_i \) with \( d_i^{\text{eff}} \) as a channel selection figure of merit. This can be achieved by replacing \( G_{ij} \in \mathbb{R}^{l \times l} \) in Eqs. (13) and (17) with \( G_{ij}^{\text{eff}} \in \mathbb{R}^{n \times n} \), the Gramian of \( S_{ij} \), given by

\[
G_{ij}^{\text{eff}} = S_{ij}^T S_{ij} = V_{ij} \Lambda_{ij}^2 V_{ij}^T,
\]

(19)

and then by the \( j \)th column \( v_{ij} \) of \( V_{ij} \) and the \( j \)th diagonal element \( \lambda_{ij}^2 \), with \( j = 1, \ldots, r \) to calculate \( d_i^{\text{eff}} \) from Eq. (11) over a selected region of the state space. It is then possible to replace \( d_i \) with \( d_i^{\text{eff}} \) in Eqs. (14) and (18) and use it as a figure of merit for channel selection over a target subspace of the state space. Note that with these modifications the computational expense of the selection algorithm becomes significantly larger, in the typical case when \( n > n_{sel} \geq l \), as at each iteration it is necessary to compute the eigenvector decomposition of a \( n \times n \) rather than of a \( l \times l \) matrix. This can still be computationally affordable (e.g., when only temperature...
is considered in the state vector and there is the need to find the best channels for temperature estimation, e.g., in the troposphere).

3. Description of the case study

To study channel selection strategies that are effective and robust both in clear sky and in the presence of cloud a case study was selected at 2100 UTC 30 June 2012. A 50-member ensemble of short-range forecasts—including also cloud liquid and ice water content and cloud fraction—generated from an EDA at ECMWF was used in this study to define a location-dependent estimate of the forecast error covariance over 91 model levels. Care was taken to inflate the variance of the forecast ensemble in order to be approximately equal to the EDA mean square error (routinely estimated from operational ECMWF analyses) so as to lead to more reliable forecasts (Bonavita et al. 2012). For each ensemble member at each location it is possible to calculate the cumulative cloud cover $N_{\text{tot}}$, defined as

$$N_{\text{tot}} = 1 - (1 - N_l) \prod_{l=2}^{R} \frac{1 - \max(N_{l-1},N_l)}{1 - N_{l-1}}$$

(20)

when all cloud layers between the top ($l = 1$) and the bottom layer ($l = n$) are considered. Here we have assumed cloud layers in the column to have maximum-random overlap (Räisänen 1998; Matricardi 2005).

Figure 2 shows the ensemble-mean values of $N_{\text{tot}}$ for the whole atmospheric depth as well as the 135 locations of atmospheric columns over ocean that are cloud free ($N_{\text{tot}} = 0$, marked with a red cross) or the 169 overcast ($N_{\text{tot}} = 1$, marked with a blue cross). Given that cloud fraction is a nonnegative quantity, the locations that are cloud free or overcast in the forecast ensemble mean are also such for all ensemble members. We can then denote the cross-marked columns as either almost surely (a.s.) (i.e., with probability 1, as defined by the considered forecast ensemble) cloud free or a.s. overcast. Hereafter, when referring to clear-sky or overcast conditions, it is assumed to consider them in an a.s. sense. Also, in Fig. 3 the vertical distribution of ensemble mean cloud quantities at the a.s. overcast locations is shown.

4. Evaluation of signal-to-noise characteristics of IASI channels

As recognized by previous channel selection studies cited in this paper, from the discussion presented in section 2 it follows that information-content-based channel selection results depend critically on the signal-to-noise characteristics of a given instrument as expressed in a particular data assimilation or retrieval system. In
particular, a meaningful expression for the number of DFS of a set of measurements requires full-rank expressions for the vertical forecast error covariance \( B \) and observation-error covariance \( R \), as well as an expression for the Jacobian \( H \) of the observation operator. A discussion of how these quantities have been determined in this work is provided below.

a. Forecast error specifications

A forecast ensemble of large size \( K \) (i.e., with \( K > n \)), for example, can be used to provide a full-rank approximation of \( B \). The dimension \( n \) of the state space, however, is usually significantly larger than the ensemble size so that the directions spanned by the forecast error vectors derived from the ensemble explore only a subset of the state space at observation location. This means that the vertical forecast error covariance estimated from the ensemble is rank deficient and it is well known that this may lead to spurious long-range correlations (e.g., Hamill et al. 2001). To avoid this problem, two complementary strategies are here adopted: (i) vertical localization of the state space to suppress correlations beyond a given threshold distance from each model level by multiplying element by element the forecast error covariance with a correlation matrix with compact support (e.g., Houtekamer and Mitchell 1998, 2001); (ii) to combine the ensemble-based forecast error variances from EDA with climatological (i.e., full rank) vertical forecast error correlation matrices derived from EDA forecasts over two seasons and geographically varying.

The left side of Fig. 4 shows the vertical temperature forecast error correlations from EDA at two overcast and clear-sky locations. The figure shows that for the overcast location in the boundary layer below about model level 75 (i.e., below about 800 hPa) the temperature error correlations at different model levels are relatively large. This is consistent with a well-mixed boundary layer that is decoupled from the above free troposphere. In the clear-sky case, these large correlations are only relevant to levels closest to the surface. Both locations also show the presence of spurious long-range correlations. An eigenvector decomposition of the correlation matrices for both locations shows that the rank of the matrix is insufficient (i.e., less than \( n = 91 \)) and equal to its theoretical maximum value (\( = K - 1 = 49 \)). As anticipated, we try to address this shortcoming by localizing the raw correlation functions with a correlation function [as described in Gaspari and Cohn (1999)] that is set to zero beyond a given distance, here chosen to be reached when the ratio \( r \) of the distance between two different model levels and the atmospheric scale height is equal to 2.0. The localized correlation matrices are shown in the right side of Fig. 4. The localized correlation matrices look reasonable as they retain the physically consistent large correlations in the short range while the spurious ones at distance larger than two atmospheric scale heights are suppressed. The rank of both localized correlation matrices is now full and equal to 92, the dimension of the state space at observation location when only the atmospheric temperature profile and the surface skin temperature are included in the state vector.

As discussed above, it is also possible to consider a regional climatology of vertical forecast error correlations—in addition to the EDA-derived variances—to calculate the signal to noise matrix for a given linearized observation operator. Vertical correlations are available for temperature, humidity, and ozone over regions of 625-km grid size and averaged over a month to a season (Anderson and Fisher 2001). Recent investigations (Holm and Kral 2012) show that the seasonal dependence of the correlations is small with respect to...
their geographical variability. In Fig. 5 (top panel) the climatological vertical temperature forecast error correlation over 91 model levels, interpolated at the selected overcast location is shown. A comparison between the top panels of Figs. 4 and 5 shows that the localization procedure applied to the raw EDA vertical error correlation matrix can make the correlation length scales of the raw matrix comparable to those characterizing the climatological covariance, with still some differences in the upper stratosphere above about model level 20 and, as to be expected, in the boundary layer. An evaluation of the eigenvalues of the correlation matrices for the overcast location (see Fig. 5, bottom panel) also shows that the absolute differences between the 49 largest eigenvalues of the climatological and EDA correlation matrices are dramatically reduced when the localized EDA correlation matrix is considered instead of the raw matrix. The localized EDA correlation matrix, however, is less conditioned than the climatological one due to the lower magnitude of the eigenvalues corresponding to eigenvectors of the localized-EDA correlation matrix spanning the subspace of the state space that is not represented by the raw forecast ensemble.

Overall, the comparison of the characteristics of the localized version of the forecast error covariance based on EDA and of that from a regional climatology shows that it is reasonable to make use of a localized ensemble-based forecast error covariance for a set of model fields to provide an estimate of the information content of a number of measurements that is as consistent as possible with the actual information content provided by the same number of measurements when assimilated in an operational data assimilation system. In view of these results, the channel selection method applied to IASI data in this work always made use of a localized

![Fig. 4. Temperature vertical forecast error correlation from a 50-member EDA. (top) The same overcast location (57.76°N, 26.25°W) as in Figs. 6 and 7, (bottom) the same clear-sky location over the Mediterranean Sea as in Fig. 1 (36.45°N, 17.5°E), (left) the raw EDA correlation, and (right) the localized correlations matrices, with $r = 2.0$ (see text). Note that level 1 is the top model level at 0.01 hPa, level 20 is at 13.50 hPa, and level 40 is at 109.9 hPa. Model levels 60, 80, and the lowest model level (91) are at 390.9, 896.7, and 998.4 hPa for the chosen overcast location and at 394.2, 912.0, and 1016.2 hPa for the chosen clear location, when ensemble mean pressure is considered.](image-url)
EDA-based forecast error covariance to determine an expression for $B$ to be used in the calculation of $G$, as discussed in section 2. In particular, the state vector as defined in this study includes temperature, humidity, and ozone components. The temperature components are defined over 91 atmospheric model levels and one surface level (surface skin temperature), while the humidity and ozone components are defined over 91 atmospheric model levels. Note that errors on the atmospheric and surface skin temperature, humidity, and ozone components of the state are assumed to be mutually uncorrelated.

b. Observation operator characterization

In this study the radiation emerging from the atmosphere was simulated using version 11 of the Radiative Television and Infrared Observation Satellite (TIROS) Operational Vertical Sounder (RTTOV; Hocking et al. 2014) in the “scattering parameterization” configuration (Matricardi 2005) to account for cloud radiative effects. To investigate the sensitivity of IASI observations to temperature and water vapor at a given channel depending on cloud conditions it is possible to calculate the Jacobian matrix, that is, the linearized observation operator $H$ [see Eq. (2)] about the ensemble mean forecast. Figure 6 shows water vapor and cloud ensemble-mean profiles at the chosen overcast location over the Atlantic Ocean. In Fig. 7 are shown the elements corresponding to atmospheric temperature and humidity of a row of the Jacobian matrix linearized about the ensemble mean forecast corresponding to IASI channel 921 (centered at 875 cm$^{-1}$ in the infrared atmospheric absorption window). In particular, it is interesting to check how the atmospheric temperature and humidity Jacobians change when cloud is removed from the instrument field of view. Figure 7 shows that the height of the peak of the temperature Jacobian in the presence of cloud is very close to that where cloud ice water density reaches its maximum value, while in the clear sky the measurements of brightness temperature in the considered atmospheric window.
channel are mainly sensitive to temperature variations in the lower troposphere, as expected. Also, the measurement’s vertical resolution is considerably higher when cloud is present in the instrument field of view and the peak temperature sensitivity is more than 20 times larger than that experienced in the clear sky. The changes in the water vapor Jacobian due to the presence of cloud are largely similar, although the altitude of the peak of the Jacobian in the presence of cloud is higher and its width remains relatively large. From Fig. 7 it follows that the presence of cloud can result in observations in the infrared atmospheric window to acquire an increased vertical resolution (on temperature estimations) and sensitivity to temperature and humidity in the vicinity and above the cloud top, as well as in a loss of sensitivity below about the cloud top. It is worth noting, however, that to translate these expected resolution and sensitivity benefits above about the cloud top into actual observational signal-to-noise ratio improvements, it is necessary that systematic radiative transfer errors and predictability issues in the NWP model’s representation of the cloud fields do not cancel out the potential advantages discussed above.

c. Observation-error specifications

The IASI observation-error covariance $R$ used in this study—in brightness temperature units—is that provided with the 1DVar scheme developed by the Met Office for the Satellite Application Facility for numerical weather prediction (NWP-SAF) and includes (uncorrelated) forward model error. The apodization process applied during the radiometric calibration of the observations from IASI introduces correlations between adjacent channels. As the channel selection method used in this work allows for the presence of correlations, there was no need to exclude adjacent channels once a given channel was selected, as required when uncorrelated observation errors are assumed. More generally, the selection method presented here allows the inclusion of systematic error components that present relatively long-range spectral correlations, which may not be compatible with channel exclusion procedures. Of this kind are errors arising from imperfect knowledge of “contaminant” species affecting the estimate of a “target” component of the state vector, such as errors due to incorrect specification of water vapor or ozone concentrations within a temperature retrieval [see, e.g., Rodgers (2000), his section 4.1.2; Dudhia et al. (2002); Ventress and Dudhia (2014)], provided that the contaminant and the target species have independent forecast errors. In this case, the total observation-error covariance $R^{\text{tot}}$ to be used in the place of $R$ to select channels that are best suited to estimate a target species using the method described in section 2 can be calculated as

$$R^{\text{tot}} = R + \sum_i H_i B_i H_i^T,$$

where $H_i$ and $B_i$ are the Jacobian and background error matrices, respectively, for the contaminant $c_i$. Note that the use of the exact expression for the “systematic” components $H_i B_i H_i^T$ of the observation-error covariance
makes the procedure, used in Dudhia et al. (2002) and Ventress and Dudhia (2014) to represent the systematic error covariance by means of an ensemble of perturbed measurement vectors, unnecessary here. Figure 8 shows the spectral dependence of the standard deviation of the components of the IASI observation-error covariance used in this work for channel selection purposes (left panel) and their spectral correlations for channels below 2200 cm$^{-1}$ (right panel).

Finally, note that Eq. (21) could in principle also be used to account for uncertainty contributions due the variability of cloud fields, which are here not included in the control vector (see section 1). To do so, however, a number of theoretical and practical issues need to be addressed, such as those related to the specification of error correlations, both within the cloud state vector as well as those between the cloud state vector and the temperature and humidity components. Also, nontrivial difficulties would arise due to the large asymmetries between total observation errors in a.s. clear-sky and in cloudy conditions, which may also be significantly non-Gaussian. While cloud-related errors are empirically evaluated and included in the observation-error covariance matrix used for all-sky assimilation of microwave radiances (Geer and Bauer 2011), a thorough investigation of the effects of cloud uncertainties on channel selection results is left for future work.

5. Channel selection implementation and results

The channel selection method discussed in section 2 was applied to each of the 135 clear-sky and 169 overcast columns in our case study, with the aim of selecting a number of humidity-sensitive IASI channels to be used for all-sky data assimilation experiments in addition to the temperature-, humidity-, and ozone-sensitive IASI channels already assimilated operationally in the clear sky. Note that the channel selection figure of merit used here is the number of DFS expressed by a set of measurement channels, but a figure of merit given by the number of effective DFS [see Eq. (11)] could have been used instead if the aim was to select a number of humidity-sensitive IASI channels over a given atmospheric region.

Similarly to the previous studies cited in section 1, the first step was to select channels primarily sensitive to atmospheric temperature profile variations located in the 15-$\mu$m carbon dioxide band, in a way to minimize contaminations from atmospheric species such as water vapor, ozone, and carbon monoxide that are radiatively active in the infrared, as well as to avoid nonlocal-thermodynamic-equilibrium (non-LTE) effects and solar contributions. An additional benefit of this channel prescreening procedure is that it reduces the nonlinearity of the observation operator, which could make the temperature Jacobians dependent on the state of the...
contaminant species and potentially lead the data assimilation analysis to be critically dependent on the minimization first guess: a temperature Jacobian, for example, may result in having its peak at an incorrectly lower height in the troposphere when the short-range model’s forecast underestimates the mixing ratio of the contaminant species (e.g., water vapor). To this end, 100 temperature-sensitive IASI channels were selected among those with wavenumber less than 900 cm$^{-1}$ (i.e., out of a total of 1020 IASI channels) at each considered location. For temperature channel selection, the total observation-error covariance matrix included systematic contributions [see Eq. (21)] to account for contaminations due to uncertainty on humidity and ozone while for humidity channel selection the only additional systematic contribution was that due to ozone uncertainty.

Once the 100th channel was added to the list of those maximizing the number of DFS for temperature, at each location the temperature state vector was augmented with the 91 components of the specific humidity vertical profile. A further set of 50 channels at each location were chosen this time among those with wavenumber between 1100 and 2200 cm$^{-1}$ (i.e., out of a total of 4399 IASI channels) to exclude the channels already selected for temperature as well as to avoid solar contamination and non-LTE effects. The number of DFS achieved by the selected channels is shown in Fig. 9. Note that the two locations (one in clear-sky and one in overcast conditions) where the overall maximum number of DFS is achieved are different from the locations where the maximum number of DFS for temperature is captured. Note also that the maximum number of DFS with 150 selected channels in clear-sky (overcast) conditions is 74.93% (64.52%) of the 15.10 (21.48) DFS achieved when all 8461 IASI channels are considered, which is still only 8.2% (11.7%) of the value that would be necessary to achieve the ideal goal of a direct and error-less joint estimate of the whole 183-component state vector.

a. Channel selection dependence on the presence of cloud

It is now interesting to discuss the different channel selection results obtained in clear-sky and in overcast conditions. Considering the cloud vertical distribution at overcast locations shown in Fig. 3, which indicates that overcast conditions are reached below about 800 hPa, it is reasonable to expect that the most informative water vapor channels selected in overcast conditions have Jacobians that are mainly different from zero above about 800 hPa. In the clear sky, however, it is expected that the selected channels also provide information about humidity in the lower troposphere. In Fig. 10 are shown the water vapor Jacobians for the 10 most informative humidity-sensitive channels at four selected clear-sky and overcast locations. Figure 10 indeed confirms that in the clear sky the selected channels can provide an estimate of water vapor mixing ratio over a wider vertical range, although the largest contributions to the total humidity DFS both in clear-sky and overcast conditions come from channels that are sensitive to water vapor in the middle and upper troposphere.

The IASI water vapor channel that, in combination with the previously selected 100 temperature channels, is mostly selected (over 28 out of 135 clear-sky locations) to provide the largest number of DFS in clear sky is channel 3446 (centered at 1506.25 cm$^{-1}$), while in
overcast conditions is channel 3244 (centered at 1455.75 cm\(^{-1}\)) to be mostly selected (over 25 out of 169 overcast locations). Note that IASI channel 3244—whose water vapor Jacobian when a 1% humidity mixing ratio perturbation is considered peaks at about 300 hPa—is also selected at 16 out of 135 clear-sky locations as the most informative humidity-sensitive channel in the clear sky and it is as well the most important water vapor channel selected during the “main run” in Collard (2007). This shows that within the NWP system used for this work, IASI is most effective in estimating water vapor in the upper-to-middle troposphere even in clear-sky conditions.

b. A strategy to select additional channels for all-sky data assimilation

To make sure the selected channels provide the largest amount of information over a broad range of...
atmospheric conditions it is important to determine how many times the 50 channels selected at a given clear-sky (overcast) location—regardless of their selection ranking—are also selected at the other 134 clear-sky (168 overcast) locations. Figure 11 shows the number of times (in percentage) that a given channel is selected at the considered clear-sky (overcast) locations relative to the total number of clear-sky (overcast) locations. The 24 humidity-sensitive channels that are selected over at least 40% of the clear-sky (overcast) locations and that are also selected over at least 40% of the overcast (clear sky) locations are denoted in Fig. 11 by red dots. These

FIG. 11. The channel selection frequency at the (left) 135 and (right) 169 clear-sky (overcast) locations. Selection frequency values are denoted with red dots for the 24 channels that are selected both in clear-sky and overcast conditions among the set of channels that are selected over at least 40% of all clear-sky or overcast locations.

FIG. 12. Spectral locations (indicated by colored circles) of a set of 24 humidity-sensitive IASI channels as selected using the iterative, nonsequential channel selection procedure described in this work. The color associated to the marker showing the spectral location of channel $i$ reflects the ranking value $q_i$ for that channel (see text).
24 channels—out of the 6750 (8450) nonunique channels that are selected at all the clear-sky (overcast) locations—populate the final channel selection shortlist. It is important to note that other criteria may be used to select the final shortlist of \( n'_{\text{sel}} \) humidity-sensitive channels. Following Collard (2007), an alternative strategy could be to compute the number of DFS achieved by each additional channel at different clear-sky (overcast) locations and then to use as figure of merit the total number of DFS computed as the sum of the number of DFS obtained by each additional channel at each considered clear-sky (overcast) location. The drawback of this procedure is that the total number of DFS may be dominated by the values obtained at specific locations where the number of DFS achieved is relatively large (e.g., in the tropics when the DFS for humidity are considered), given the large variability of the DFS results over different locations (see Fig. 9). To avoid this shortcoming, the histogram approach discussed above was used instead, with the drawback of requiring us to define an arbitrary selection frequency threshold.

As discussed above, the importance of a selected channel depends on both its selection frequency \( f_i \) over the considered locations and the iteration step \( p_i \) (in %) in which the channel \( i \) was selected—the earlier the step the more important the channel—defined as \( p_i = 100 \times (n_{\text{sel}} - l_i)/(n_{\text{sel}} - 1) \), where \( n_{\text{sel}} \) is the number of selected channels (in this case 50) and \( l_i \) is the iteration in which channel \( i \) was selected. For example, \( l_i = 1 \) means that channel \( i \) provides the largest increase in DFS on temperature and humidity given the previously selected—in this case 100—temperature-sensitive channels, leading to a value of \( p_i \) given by \( p_i = 100\% \). Given that the present channel selection needs to be suitable both in clear-sky and overcast conditions, a weighted average of both the selection frequency \( f_i \) and the iteration step percentage \( p_i \) according to the number considered clear-sky and overcast locations was calculated. In this way the weights reflect the fact that, in general (including in this case), the number of clear-sky locations differs from the number of considered overcast locations. A final ranking \( q_i \) of the shortlisted channels (see Fig. 12 and Table 1) was then obtained by calculating the arithmetic mean between the weighted \( f_i \) and the weighted \( p_i \). For sake of illustration, let us calculate step by step the final ranking for one of the shortlisted channels (i.e., channel 2675). This channel was selected at 68.89% (66.86%) of all clear-sky (overcast) locations and its iteration step percentage averaged over all considered clear-sky (overcast) locations was 67.46% (75.08%). This means that channel 2675 has a weighted average selection frequency given by \( \overline{f_i} = (135 \times 68.89 + 169 \times 66.86)/(135 + 169) \) and a weighted average iteration step percentage given by \( \overline{p_i} = (135 \times 67.46 + 169 \times 75.08)/(135 + 169) \) corresponding to a final ranking \( q_i \) for channel 2675 given by \( q_i = (\overline{f_i} + \overline{p_i})/2 = 69.73\% \). The weighting choices here adopted to determine a final channel selection ranking order are somehow arbitrary.

It is important to note, however, that only six channels out the 24 shown in Table 1 are not already included in the set chosen by Collard (2007) in clear-sky conditions and the highest ranking among these “new” channels is placed in tenth position. This provides an important confirmation of the soundness of the selection strategy presented here, aiming to choose channels that are informative both in clear-sky and cloudy conditions.

Finally, in Fig. 13 (left panel) are shown the humidity Jacobians for the channels listed in Table 1 calculated using RTTOV v11—with coefficients based on the Line-By-Line Radiative Transfer Model (LBLRTM) over 101 vertical levels—at a nonisolated clear-sky location (i.e., surrounded by other clear-sky columns) over the Mediterranean Sea at 36.45°N, 17.5°W. It is also interesting to calculate the DFS weighting functions for the selected 24 channels, which are shown in Fig. 13 (right panel) in the case when no forecast error variance inflation (briefly mentioned in section 3) is applied. A

<table>
<thead>
<tr>
<th>IASI channel No.</th>
<th>Wavenumber (cm(^{-1}))</th>
<th>Channel ranking (%)</th>
<th>Ranking order</th>
</tr>
</thead>
<tbody>
<tr>
<td>2675</td>
<td>1313.5</td>
<td>69.73</td>
<td>12</td>
</tr>
<tr>
<td>2868</td>
<td>1361.75</td>
<td>71.50</td>
<td>10</td>
</tr>
<tr>
<td>2939</td>
<td>1379.5</td>
<td>55.15</td>
<td>21</td>
</tr>
<tr>
<td>2959</td>
<td>1359.5</td>
<td>62.90</td>
<td>16</td>
</tr>
<tr>
<td>2991</td>
<td>1392.5</td>
<td>62.55</td>
<td>17</td>
</tr>
<tr>
<td>3002</td>
<td>1395.25</td>
<td>73.75</td>
<td>9</td>
</tr>
<tr>
<td>3009</td>
<td>1397</td>
<td>76.63</td>
<td>6</td>
</tr>
<tr>
<td>3064</td>
<td>1410.75</td>
<td>56.39</td>
<td>20</td>
</tr>
<tr>
<td>3093</td>
<td>1418</td>
<td>74.70</td>
<td>8</td>
</tr>
<tr>
<td>3105</td>
<td>1421</td>
<td>65.72</td>
<td>13</td>
</tr>
<tr>
<td>3244</td>
<td>1455.75</td>
<td>83.36</td>
<td>2</td>
</tr>
<tr>
<td>3248</td>
<td>1456.75</td>
<td>39.16</td>
<td>24</td>
</tr>
<tr>
<td>3252</td>
<td>1457.75</td>
<td>78.31</td>
<td>5</td>
</tr>
<tr>
<td>3312</td>
<td>1472.75</td>
<td>83.25</td>
<td>3</td>
</tr>
<tr>
<td>3321</td>
<td>1475</td>
<td>63.87</td>
<td>15</td>
</tr>
<tr>
<td>3411</td>
<td>1497.5</td>
<td>70.04</td>
<td>11</td>
</tr>
<tr>
<td>3446</td>
<td>1506.25</td>
<td>84.39</td>
<td>1</td>
</tr>
<tr>
<td>3453</td>
<td>1508</td>
<td>59.73</td>
<td>19</td>
</tr>
<tr>
<td>3509</td>
<td>1522</td>
<td>61.98</td>
<td>18</td>
</tr>
<tr>
<td>3527</td>
<td>1526.5</td>
<td>41.16</td>
<td>23</td>
</tr>
<tr>
<td>3575</td>
<td>1538.5</td>
<td>53.11</td>
<td>22</td>
</tr>
<tr>
<td>3580</td>
<td>1539.75</td>
<td>80.71</td>
<td>4</td>
</tr>
<tr>
<td>3653</td>
<td>1558</td>
<td>75.76</td>
<td>7</td>
</tr>
<tr>
<td>3658</td>
<td>1559.25</td>
<td>64.21</td>
<td>14</td>
</tr>
</tbody>
</table>
comparison between the left and the right panels of Fig. 13 shows that the width of the region where the Jacobians of the selected channels are significantly different from zero (between about 150 and 800 hPa) coincides with that of the cumulative DFS weighting function. The cumulative DFS weighting function when forecast error inflation (see section 3) is applied (not shown) has also a similar dependence with height, although it becomes negligible above about 270 hPa due to both the rapid decline of the water vapor forecast error standard deviation with height and the effects of its calibration, which reduce the magnitude of the standard water vapor mixing ratio by more than 80% in the region above 290 hPa.

6. Summary and conclusions

In this study the iterative channel selection method, which is in standard use at operational meteorological centers to select an optimal subset of all available channels from advanced infrared sounding instruments for assimilation, was revisited in order to select channels with correlated errors (due to both apodization and interference from contaminant species) using an ensemble-based estimation of forecast uncertainty both in clear-sky and overcast conditions. Note, however, that the additional and potentially large observation uncertainty due to erroneous specification of cloud profiles was not taken into account in the channel selection procedure. This limitation was to avoid difficulties such as those arising from significant asymmetries between observation errors in cloud-free and overcast conditions, as in the former case the interference error term as defined in this paper would be zero while in the latter it could be very large. The channel selection implications of additional uncertainty due to cloud is left for future work.

Also, the standard channel selection figure of merit, defined by the number of DFS expressed by the channels already selected plus that of an additional candidate channel, was modified so as to be able to be optionally used for selecting an optimal set of channels for estimation of a portion of the state space (e.g., tropospheric temperatures). To this end, the new concept of (cumulative) DFS weighting function was introduced, which can also be used to provide a synthetic, nondimensional, and normalized picture of the region of the state space from which is possible to extract the (cumulative) contributions to the DFS expressed by a given set of channels.

![Fig. 13](image-url)
The observation-error-correlation-aware channel selection method discussed in this paper was then used—in its standard figure of merit formulation—to select a set of 100 temperature-sensitive (below 900 cm\(^{-1}\)) and 50 humidity-sensitive (between 900 and 2200 cm\(^{-1}\)) IASI channels at a number of clear-sky and overcast locations for a case study in July 2012. Care was taken to select a final shortlist of 24 humidity-sensitive channels from the set of humidity-sensitive channels that were selected both in clear-sky and overcast conditions over at least 40% of all considered locations. Finally, a ranking of the shortlisted channels was provided, based on their selection frequency and average selection iteration step. Future work will investigate the potential of an all-sky assimilation of (a subset of) the selected humidity-sensitive IASI channels on improving the ECMWF forecast skill scores over suitable case studies.

Acknowledgments. The author is partly funded by the NERC National Centre for Earth Observation. The author would like to thank F. Baordo, M. Bonavita, N. Bormann, S. English, R. Eresmaa, A. Geer, M. Hamrud, E. Holm, C. Lupu, M. Matricardi, and T. McNally for their help, suggestions, and comments. The comments and questions of three anonymous reviewers, which helped to improve the paper are gratefully acknowledged.

APPENDIX

Evaluation of the Computational Costs of Channel Selection

When the observations have mutually independent errors, the analysis error covariance matrix \(P^a_{l-1} \in \mathbb{R}^{n \times n}\) at iteration \(l\) can be calculated sequentially as [see, e.g., Rodgers (2000), his section 5.8.1.3]:

\[
P^a_{l} = P^a_{l-1} \left[ 1 - \frac{h_l(P^a_{l-1}h_l)^T}{1 + (P^a_{l-1}h_l)^T h_l} \right], \tag{A1}
\]

where \(h_l \in \mathbb{R}^n\) is the noise-normalized Jacobian of the observation operator for channel \(l\), with \(P^a_0 = B\). The channel selection figure of merit, given by the number of degrees of freedom for signal \(d_f\), can then be calculated as shown in Eq. (4), where the inverse of the forecast error covariance matrix \(B^{-1}\) is assumed known. Let us now calculate the number of floating point operations (FLOPs) to compute \(d_f\) as in Eq. (4) using \(P^a_{l}\) as in Eq. (A1). By noting that the inner product of two vectors of dimension \(n\) consists of \(n\) multiplications and \(n - 1\) summations we get a cost for \(P^a_{l}h_l\) of \(2n^2 - n\) FLOPs and of \(n^2\) FLOPs for \(h_l(P^a_{l}h_l)^T\). An additional \(n\) FLOPs are needed to compute \(1 + (P^a_{l}h_l)^T h_l = 1 + \text{tr}(h_l(P^a_{l}h_l)^T)\) and another \(n^2 + n\) for the terms inside the brackets in Eq. (A1). The covariance update requires instead \(2n^3 - n^2\) FLOPs. This means that we need a total of \(2n^3 + 3n^2 + n\) to calculate \(P^a_{l}\) as in Eq. (A1). Finally, the computation of \(\text{tr}(I_n - P^a_{l}B^{-1})\) requires \(2n^3 - n^2 + n\) FLOPs. In summary, to calculate \(d_f\) for a given channel \(i\) at iteration \(l\) using the sequential algorithm as in Collard (2007), a total of \(4n^3 + 2n^2 + 2n\) FLOPs are required.

Let us now consider a sequential version of algorithm discussed in section 2b and write the signal-to-noise matrix \(S_{l-1} \in \mathbb{R}^{(l-1) \times n}\) at iteration \(l\) as \(S_{l-1} = H_{l-1}X_I\), where \(H_{l-1}\) is now the noise-normalized Jacobian matrix for the already chosen \(l - 1\) channels. The Gramian of \(S^T_{l-1}\) can then be written as

\[
S_{l-1}S^T_{l-1} = \left( S_{l-1}S^T_{l-1} \begin{bmatrix} S_{l-1}X^T_Ih_l \\ h_l^TX^T_{l-1}S_{l-1}h_l \end{bmatrix} \right). \tag{A2}
\]

The cost of the algorithm at iteration \(l - 1\) is then given by the cost of calculating \(S_{l-1}\), equal to \(2(l - 1)n^2 - (l - 1)n\) FLOPs, plus the cost of computing the Gramian of \(S^T_{l-1}\), equal to \((l - 1)^3 + (l - 1)^2/2 - (l - 1)/2\) as only either the lower or the upper triangular part of the Gramian needs to be computed [i.e., only \((l - 1)^2/2 + (l - 1)/2\) elements of \(S_{l-1}S^T_{l-1}\), each with a cost of \(2(l - 1) - 1\) FLOPs]. The additional cost at iteration \(l\) is due to the calculation of \(S_{l-1}X^T_Ih_l\), equal to \(2(n - 1)n + (2n - 1)(l - 1)\) FLOPs, and that of computing \(h_l^TX^T_{l-1}S_{l-1}h_l\), equal to \(2n - 1\) FLOPs. Finally, the cost of finding the eigenvalues of \(S_{l-1}S^T_{l-1}\) is dominated by the reduction of the Gramian to tridiagonal form, given by \(4l^3/3\) FLOPs [e.g., Golub and Van Loan (2013), their section 8.3.1], and the computation of the figure of merit requires additional \(2r + r - 1\) FLOPs, with \(r = \min(l, n)\). In summary, the sequential version of the channel selection algorithm presented in this paper has an iteration-dependent cost, which is equal to \((2n - 1)(n + 1)/2 + 4l^3/3 + (l - 1)^3 + (l - 1)^2/2 - (l - 1)/2 + 3r - 1\) at iteration \(l\). Note, however, that the cost of calculating the new value of the figure of merit for a channel \(j \neq i\) at iteration \(l\) does not include again the cost of computing \(S_{l-1}\) and \(S_{l-1}S^T_{l-1}\).

From the above discussion it follows that the cost of calculating the value of the figure of merit for a given set of channels using the standard sequential algorithm in Eq. (A1) is the same at any given iteration and for each additional channel, while the cost of the algorithm discussed here depends on the number of already selected
channels—as the number of degrees of freedom for signal is calculated in measurement space rather than in state space—and can be made it cheaper (in relative terms) by exploiting economies of scale when selecting from a large pool of channels. For this reason, in order to compare the two algorithm it is convenient to calculate the cost of selecting an increasingly larger number of channels from a given set of available channels, for a given dimension of the state space. From Figure A1 it follows that the channel selection algorithm presented in this paper, in its sequential implementation, can be used to select 100 temperature-sensitive channels and 50 additional humidity-sensitive channels from those acquired by an advanced infrared sounder for less than 12% of the cost than that needed when the standard sequential channel selection algorithm is used.

REFERENCES

Hamill, T., J. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an en-


Matricardi, M., 2005: The inclusion of aerosols and clouds in RTIASI, the ECMWF fast radiative transfer model for the infrared atmospheric sounding interferometer. ECMWF Tech. Memo. 474, 53 pp.


