Methodology and Information Content of the NOAA NESDIS Operational Channel Selection for the Cross-Track Infrared Sounder (CrIS)

Antonia Gambacorta and Christopher D. Barnet

***Abstract*—The Cross-track Infrared Sounder (CrIS) was launched on October 28, 2011 aboard the Suomi National Polar-orbiting Partnership platform and is scheduled to become operational in 2012. The purpose of this paper is to describe the National Oceanic and Atmospheric Administration/National Environmental Satellite and Information Service (NOAA/NESDIS) channel selection methodology applied to the CrIS instrument and to present the main spectral characteristics of the final channel subset that will be operationally distributed to the scientific community for near-real-time data assimilation and retrieval applications. We perform an information content analysis and show that this selection, composed of 399 channels, is capable of fully representing the total atmospheric variability contained in the original 1305-channel spectrum, up to instrumental noise. These results ensure that the replacement of the full 1305-channel list in favor of the proposed 399-channel selection will have no detrimental effects on data assimilation and retrieval performance.**

***Index Terms*—Information content, spectral channel selection, variance explained.**

I. Introduction

LAUNCHED onboard the Suomi National Polar-orbiting Partnership platform on October 28, 2011, the Cross-track Infrared Sounder (CrIS) is a Fourier transform spectrometer with a total of 1305 infrared sounding channels covering the long-wave (655–1095 cm^{-1}), midwave (1210–1750 cm^{-1}), and short-wave (2155–2550 cm^{-1}) infrared spectral regions [1]. Given the large yield of the expected data flow, a *radiance thinning* procedure is required that will expedite the near-real-time data distribution and assimilation processes. Although hyperspectral measurements are affected by high redundancy, careful attention must be paid during this radiance thinning procedure, aimed at minimizing the loss of atmospheric information, such that the final retrieval quality is not deteriorated.

A distinction must be made among two categories of radiance thinning. One category is *data compression* by means of principal component analysis (PCA), and the other is *spectral thinning* by means of a spectral channel selection. PCA compression of hyperspectral radiance data [2] is often used for data distribution but has also found a large number of applications in the field of inversion algorithms, such as the following: 1) fast regression computations, which are often used as the first guess of physical retrieval algorithms [2], [3], and 2) instrumental noise reduction of reconstructed radiances [4].

Channel selection thinning consists in selecting a subset of channels carrying unique pieces of information such that the total variability contained in the full spectrum is still represented, up to instrumental noise. These channel subsets are used to improve the computation efficiency of data assimilation and retrieval operations.

Existing channel selection methodologies, as described in [5]–[7], follow a statistical iterative methodology [8] where successive tests are performed on the incremental addition of each channels. At each addition, the channel is retained if the information content (based on the entropy of the system) is observed to increase. For near-real-time operational applications, a *constant* channel selection is normally used, which is derived as an average from multiple optimal selections computed over the globe. Differently from these statistical techniques [8], the NOAA/NESDIS methodology employs a physically based procedure where channels are selected solely upon their spectral properties: High priority is given to spectral purity, avoidance of redundancy, and vertical sensitivity properties, along with low instrumental noise and global optimality.

The NOAA/NESDIS methodology is a heritage technique previously adopted for the operational channel selection of the National Aeronautics and Space Administration (NASA) Atmospheric Infrared Sounder (AIRS) [3] and later employed for the Metop-B Infrared Atmospheric Sounding Interferometer (IASI) NOAA/NESDIS channel selection [9]. Since the beginning of the AIRS and IASI postlaunch operations (2003 and 2008, respectively), both the National Weather Prediction and research users have found great advantage in employing these AIRS and IASI near-real-time channel selections to expedite data assimilation processing while preserving good quality in the retrieval results [10]–[12]. Current users of the AIRS channel selection are the NOAA National Centers for Environmental Prediction/Joint Center for Satellite Data Assimilation (NCEP/JCSDA), the Global Modeling and Assimilation Office (GMAO), the National Research Laboratory/Air Force Weather Agency (NRL/AFWA), the European Centre for

Manuscript received January 31, 2012; revised June 25, 2012; accepted August 11, 2012. Date of publication December 11, 2012; date of current version May 16, 2013. The views, opinions, and findings contained in this paper are those of the authors and should not be construed as an official National Oceanic and Atmospheric Administration or U.S. Government position, policy, or decision.

A. Gambacorta is with IM System Group, NOAA/NESDIS, Camp Springs, MD 20746 USA (e-mail: antonia.gambacorta@noaa.gov).

C. D. Barnet is with NOAA/NESDIS, Camp Springs, MD 20746 USA (e-mail: chris.barnet@noaa.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2012.2220369

U.S. Government work not protected by U.S. copyright.

Medium-Range Weather Forecasts (ECMWF), the U.K. Met Office, Meteo France, the Deutscher Wetterdienst, the Canadian Meteorology Center (CMC), the Japanese Meteorological Agency, the Australian Bureau of Meteorology, the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), and, through EUMETSAT, all the 23 European member states. Current users of the IASI NOAA/NESDIS channel selection are NCEP/JCSDA, GMAO, NRL/AFWA, and CMC.

The CrIS instrument has been designed to guarantee continuity to the 1:30 A.M./P.M. equatorial crossing time orbit, in replacement of the AIRS instrument. As for AIRS and IASI, the operational CrIS channel selection is scheduled to remain unchanged during the whole life mission of the instrument, in order to ensure a uniform long-term data record of reanalysis and climate variables. For this purpose, robustness and global applicability are fundamental prerequisites to an equivalent success and durability of the CrIS operational channel selection.

The objective of this paper is to describe the NOAA/NESDIS channel selection methodology applied to the CrIS instrument, to present the main spectral characteristics, and to assess the information content and global optimality of the final channel subset that will be operationally distributed to the worldwide scientific community for near-real-time data assimilation and retrieval applications. The CrIS NOAA/NESDIS channel selection described in this paper will have the same operational user outreach as the AIRS channel selection, with the addition of the full World Meteorological Organization/Global Telecommunication System user network [13].

This channel selection is composed of a total of 399 channels consisting of 24 surface temperature and emissivity, 87 temperature, 62 water vapor, 53 ozone, 27 carbon monoxide, 54 methane, 53 carbon dioxide, 24 N_2O, 28 HNO_3, and 24 SO_2 sounding channels. Preliminary test applications have shown this channel selection to be significantly advantageous in terms of computation efficiency. For example, the computational time of routine operations, such as the computation of radiance covariance matrices, has been observed to improve by a factor of 10 when going from assimilating full spectra down to the 399-channel selection.

This paper is divided into two parts. Section II hereinafter provides a detailed description of the channel selection methodology. The focus of this section is an assessment of the degree of spectral purity associated with each selected channel. Sections III and IV assess the information content present in the selection by performing both a principal component and a regression analysis aimed at quantifying the fraction of the full spectrum atmospheric variance explained by this selection.

II. Channel Selection Methodology

The finite spectral resolution of the instrument does not allow for *spectral purity*. Specifically, in the infrared domain, the signal associated with a given channel of nominal frequency ν is, in reality, the result of multiple molecular rotovibrational transitions whose spectral range of occurrence overlaps with the spectral resolution $\Delta\nu$ of the channel. For a given atmospheric species of interest then, the accuracy of the retrieval and its associated error characterization rest on the knowledge of the degree of spectral purity present in the subset of channels selected to run the retrieval.

We employ the Stand-alone AIRS Radiative Transfer Algorithm (SARTA) radiative transfer model [14] to compute a spectral sensitivity analysis of the full CrIS spectrum. We use standard U.S. atmospheric profiles to perform brute force perturbations of temperature, water vapor, and trace gases (ozone, methane, carbon dioxide, carbon monoxide, HNO_3, N_2O, and SO_2). We first analyze the radiance response to each perturbation, $\Delta R(\nu)$, represented by the difference between the perturbed and the unperturbed radiance spectra. In this methodology, derivatives of the radiance spectrum, defined as spectral sensitivity functions $S(\nu)$, are computed by finite differencing for each perturbation δX_j, according to

$$S(\nu)\delta X_j = \frac{\partial R(X_o,\nu)}{\partial X_j}\partial X_j \approx \Delta R_j(\nu) \tag{1}$$

$$\Delta R_j(\nu) = (R(X_o + \delta X_j) - R(X_o)) \tag{2}$$

where R is the radiance signal, X_o represents the unperturbed atmospheric state, and δX_j is the perturbation associated with species j. These radiance differences indicate the sensitivity of each channel to each specific atmospheric species of interest and are used in our study as indexes of spectral purity.

Fig. 1 provides an illustrative example of this sensitivity analysis, applied to the case of a midlatitude U.S. standard atmospheric profile (see online color figure). Here, the different curves represent differences in brightness temperature due to a perturbation in temperature (1 K perturbation, black curve), water vapor (10% perturbation, red curve), ozone (10% perturbation, blue curve), surface temperature (1 K perturbation, green curve), methane (2% perturbation, magenta), carbon monoxide (1% perturbation, cyan), carbon dioxide (1% perturbation, light purple), HNO_3 (1% perturbation, orange), N_2O (1% perturbation, yellow), and SO_2 (1% perturbation, dark purple).

A preliminary channel screening is computed on the basis of the sensitivity features highlighted in Fig. 1. We section the spectrum into different channel subsets, with each subset including channels highly sensitive to a specific atmospheric species of interest and with the lowest interference from all remaining species. In doing so, for example, temperature sounding channels will be preferentially selected in the 670- and 2300-cm^{-1} regions, where the sensitivity to temperature perturbation is high (black curve) and the interference of other species is minimal or is well known (see ahead discussion on noise covariance).

After this preliminary screening, we closely study the properties of the vertical sensitivity functions, $S(\nu)$, associated with the selected channels. Since the shape of the sensitivity functions is influenced by temperature and pressure as well as by the vertical variation of the atmospheric species mixing ratio, the selection of channels whose sensitivity functions peak at different altitudes enables the vertical sounding of the atmosphere. The finite spectral resolution of the instrument is responsible for broadening the channel sensitivity functions into coarse vertical ranges. One-peaked channels as well as channels with sharper sensitivity functions are then preferentially

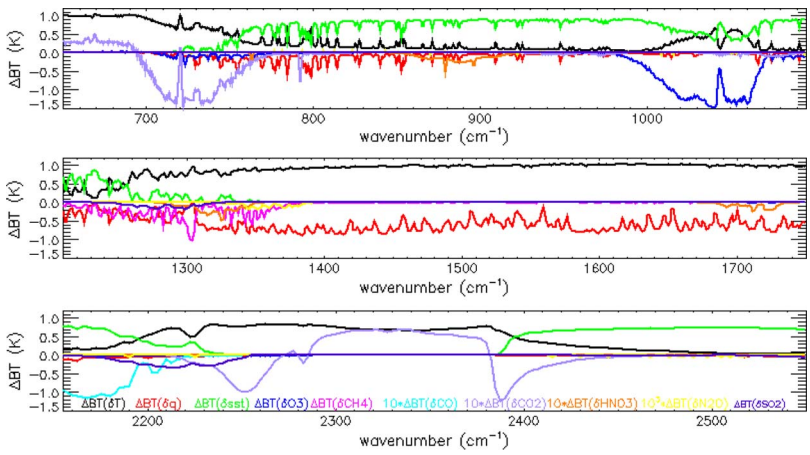

Fig. 1. Sensitivity analysis of CrIS channels. Black curve: Temperature sensitivity analysis. Red curve: Water vapor sensitivity analysis. Green curve: Surface temperature sensitivity analysis. Blue curve: Ozone sensitivity analysis. Magenta curve: Methane sensitivity analysis. Cyan curve: Carbon monoxide sensitivity analysis. For display purposes, a factor multiplier equal to 10 has been used. Light purple: Carbon dioxide sensitivity analysis. For display purposes, a factor multiplier equal to 10 has been used. Orange curve: HNO_3 sensitivity analysis. For display purposes, a factor multiplier equal to 10 has been used. Yellow curve: N_2O sensitivity analysis. For display purposes, a factor multiplier equal to 10^3 has been used. Dark purple: SO_2 sensitivity analysis.

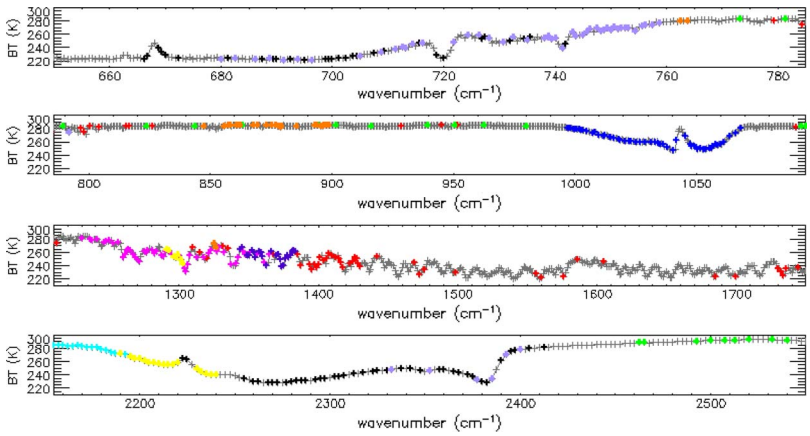

Fig. 2. Final channel selection. Gray cross symbols indicate the location of all 1305 channels present in CrIS original spectra. Superimposed colored cross symbols indicate the 10 channel subsets forming our final channel selection. The final selection is composed of (green) 24 surface temperature and emissivity, (black) 87 temperature, (red) 62 water vapor, (blue) 53 ozone, (cyan) 27 carbon monoxide, (magenta) 54 methane, (light purple) 53 carbon dioxide, (yellow) 24 N_2O, (orange) 28 HNO_3, and (dark purple) 24 SO_2 sounding channels. The total number of channels is 399.

selected over multiple-peaking and broadly structured ones, in the attempt of maximizing the vertical resolution of the retrieval product.

It needs to be said that, given the large number of channels carried by a hyperspectral instrument (1305 for CrIS, 2371 for AIRS, and 8741 for IASI), the preliminary channel screening described earlier is necessary in providing a considerable initial thinning of the original spectrum. Nonetheless, because of the large structure in the interference patterns of the concurrent atmospheric sources across the spectrum, it is not feasible to perform this initial thinning by means of fixed sensitivity thresholds. Indeed, the whole selection is computed manually, by means of multiple iterations covering the steps described earlier, during which channel properties are studied in subsets and gradually screened based on spectral purity and vertical sensitivity.

In summary, the NOAA/NESDIS methodology is a physically based procedure where channels are selected solely upon their spectral properties. Redundancy is removed by selecting channels with the highest spectral purity and whose sensitivity function structure maximizes vertical sounding coverage and resolution.

The example shown in Fig. 1 refers to a midlatitude profile. In reality, we have performed this sensitivity study on multiple atmospheric test cases, representative of different surface emissivity and air mass conditions, such as ocean/land, desert, polar, midlatitude, and tropical regimes, not shown here for brevity. A dedicated channel selection was performed under each of these regimes, following the directions discussed earlier. Notwithstanding the large variety of geophysical regimes under which they were taken, all these channel selections were seen in the end to share a considerable number of common channels. In order to maximize robustness and global applicability, the final channel selection was taken as the union of all these different selections. Test results on robustness and global applicability of the final channel selection will be presented in Section IV.

This final channel selection is composed of a total of 399 channels consisting of 24 surface temperature and emissivity,

87 temperature, 62 water vapor, 53 ozone, 27 carbon monoxide, 54 methane, 53 carbon dioxide, 24 N_2O, 28 HNO_3, and 24 SO_2 sounding channels. Gray cross symbols on Fig. 2 indicate the location of all 1305 channels present in the original spectrum of the CrIS instrument. Superimposed colored cross symbols indicate the 10 channel subsets forming our final channel selection (see online color figure). Here, we have adopted the same color code as that of Fig. 1 to indicate the different atmospheric species to which the selected channels are mainly sensitive.

The main spectral and noise characteristics of this channel selection are summarized in the paragraphs hereinafter. The reader is referred to [15] for a complete table list of all selected channel frequencies and associated spectral transmittance values (table omitted here for brevity).

A. Selection of Temperature Sounding Channels

Temperature sounding requires channels sensitive to well mixed and slowly varying gases. Infrared sounders normally employ channels sensitive to carbon dioxide which has two strong fundamentals in the infrared spectrum, specifically in the 670- and 2300-cm^{-1} regions (commonly expressed in micrometers as 15- and 4.3-μm bands). We follow previous theoretical studies by Kaplan *et al.* [16] and preferentially select channels along the wings of CO_2 transition lines, whose temperature weighting functions were shown to have half the width at half peak as that of core channel weighting functions, due to the rapid increase of the absorption coefficient with pressure.

As shown in [17], the most pronounced sensitivity functions are found in the 2300-cm^{-1} band. This is due to the rapid increase of the absorption coefficient with tropospheric temperature. Aside from the higher instrumental noise, the only complication of the short-wave band is the high solar reflectivity term affecting the surface and low troposphere sounding channels of the CO_2 R-branch and the nonlocal thermodynamic equilibrium affecting the upper and stratospheric sensitive channels of the P-branch. These factors, both occurring during daily conditions, are affected by large uncertainties during forward model calculations.

The complementary 670-cm^{-1} channels become important for sounding under the conditions of partial cloudiness [18]. Long-wave channels have good sensitivity to temperature from the surface to the stratosphere, up to 1 mb. The only complication affecting this band rests in the interference of water vapor and ozone in contrast to the head band of the 2300-cm^{-1} spectral region which instead shows low sensitivity to these two interfering species.

The temperature subset is also applicable for cloud filtering techniques [3]. For cloud height retrievals based on CO_2 slicing techniques [18], only the long-wave subset of this selection is used in order to avoid solar reflectance effects, as described earlier, and because of the uncertainty in the radiative properties of low cloud emissivity.

B. Selection of Surface Temperature and Emissivity Sounding Channels

The selection of surface temperature and emissivity sounding channels requires high transmittance from all atmospheric components. A total of 24 channels have been selected in the 750–1100-cm^{-1} and 2150–2540-cm^{-1} window regions. Attention is required when using channels in the 1000–1275-cm^{-1} band, due to existing difficulties in the parametrization of desert emissivity.

C. Selection of Atmospheric Constituent Sounding Channels

For atmospheric constituents, the core spectral line has been favored over wing line channels in order to maximize the signal-to-noise ratio, with equivalent attention to spectral purity and sensitivity function sharpness properties. Complementary wing line channels have been added to the selection in the attempt of improving the vertical coverage of the channel selection sensitivity, enabling vertical profile retrievals.

D. Instrumental and Apodization Noise Effects

A large effort of the selection methodology also involves a thorough scrutiny of the channel noise properties. The selection generally tends to avoid channels with high instrumental noise, such as those in the short-wave region, and channels affected by poor calibration.

Careful attention must also be paid to the noise correlation of adjacent channels resulting from apodization. Apodization is typically employed in forward models to allow for efficient computations of radiances [19]. For a full spectrum, apodization is a reversible process and therefore does not alter the information content of the spectrum. The downside of apodization though is the spectral correlation of the noise [20]. The expansion of an apodization function into an infinite cosine series enables the determination of the apodization noise reduction factor and spectral correlation among adjacent channels.

Apodized radiances sampled at the Nyquist sampling can be written as a $2J-1$ point running mean of the raw instrument radiances with weights w_k, given by

$$R_{apod}(v') = \sum_{k=-J+1}^{J-1} w_k R(v+k). \tag{3}$$

We can define the weights w_k in terms of the cosine expansion coefficients a_k as follows:

$$w_k = a_{|k|}; \qquad k = -(J-1), \ldots, J-1. \tag{4}$$

The apodized noise is reduced by a factor f given by

$$f = \left[\sum_{k=-J+1}^{J-1} w_k^2 \right]^{-1/2} \tag{5}$$

and the channel correlation factor CF_n for two channels separated by n Nyquist channels is given by

$$CF_n = f^2 \sum_{k=-J+1}^{J-1-n} w_k w_{k+n} \tag{6}$$

where $n = 1, \ldots, 2J-2$.

For NESDIS applications, CrIS radiances are apodized using a Hamming apodization function. The Hamming apodization

has $J = 2$, $a_0 = 0.54$, and $a_1 = 0.23$ so that $w_1 = 0.23$, $w_0 = 0.54$, $w_1 = 0.23$, and $\sum(w_k) = 1$. The Hamming apodization is able to reduce CrIS's noise by a factor $f = 1.5862$. In turn, adjacent channels are now correlated by a correlation factor $CF_1 = 62.5\%$, and alternate channels are correlated by $CF_2 = 13.3\%$. Channels separated by more than two indexes have zero correlation. In the attempt of maximizing the vertical sensitivity coverage, the proposed channel selection does contain, at times, few groups of adjacent channels. Users are advised then to pick every other third channel from the proposed selection, if in need of eliminating apodization correlation noise effects. In doing so though, attention must be paid to still retain uniform sensitivity coverage along the vertical atmospheric column.

It needs to be said that the Joint Polar Satellite System (JPSS) apodization code employs a Blackman–Harris apodization function [20], for which $J = 3$. Channels selected herein should perform adequately in this application as well.

E. Applications

The issue of spectral impurity affecting the accuracy of the retrieval product can be circumvented by exploiting the distinctive nonlinearity of the radiative interference sources shown in Fig. 1. During the retrieval of a given species of interest, the interference signals of undesired species (held constant at that step) can be treated as radiance error sources that can be used as off-diagonal terms of the observed radiance error covariance matrix. In doing so, the least square minimization is made insensitive to the interfering species, and the inversion problem can be solved by parts. In this scheme, the perturbation on species X_j, from which we compute the associated error source $\Delta R_j(\nu)$, is derived from an error estimate on the geophysical species, δX_j. The corresponding computational uncertainty $\Delta R_j(\nu)$ is a finite difference expressed by

$$\Delta R_j^{s,i}(v) = \left(R\left(X_o^{s,i-1} + \delta X_j * Q_j\right) - R\left(X_o^{s,i-1}\right)\right) \quad (7)$$

where X_o is the current estimate of the geophysical state at a given step s and iteration i. δX_j is the perturbation on the X_j component.

Since δX_j is an residual sum of squares (RSS) error estimate, it can be vertically cross-correlated and spectrally correlated with respect to other parameters (e.g., surface spectral emissivity error can be correlated with skin temperature). A scaling factor Q_j is needed to compensate for assumed correlations in these error estimates.

During the retrieval step s and iteration i of a given species of interest, the off-diagonal terms of the error covariance matrix will be obtained from the summation of all radiance interference terms from undesired geophysical species (held constant during that retrieval step), as in

$$\Delta R(\nu, \nu') = \sum_{j=0}^{N-1} \Delta R_j^{s,i}(\nu)\left(\Delta R_j^{s,i}(\nu')\right)^T. \quad (8)$$

The accuracy of the solution will strongly depend on the accuracy of the radiative transfer model used to compute these off-diagonal noise terms.

The approach described earlier forms the basics of the current NASA operational hyperspectral inversion algorithm for the AIRS [3] and the NOAA/NESDIS IASI [21], whose operational channel selections were performed following the same methodology as that described in this paper [9]. The same inversion scheme will be employed for the NOAA/NESDIS operational processing of CrIS data.

On the other hand, this channel selection is also suitable for alternative variational inversion techniques, which can resort to the interference of multiple species affecting a given channel, as described by the curves of Fig. 1, to simultaneously solve for multiple species of interest. At present, NCEP employs a Three-Dimensional Variational (3-DVAR) inversion technique to operationally assimilate the aforementioned AIRS and IASI channel selections, along with the CrIS channel selection discussed in this paper.

III. Information Content Analysis

A. SVD Analysis

We have used a subsample of the global CrIS proxy data from October 19, 2007 [22] where we have considered only the first two scan lines of each granule, for a total of roughly 14 000 radiance spectra, uniformly distributed, including ocean/land/day/night cases. We use this data sample to perform a singular value decomposition (SVD) analysis of the noise weighted radiance covariance matrix, after mean centering of the data. We then presort the derived eigenvectors based on their eigenvalues, λ's, from the highest to the lowest eigenvalue. The first five eigenvalues alone were found to account for more than 99% of the total variance contained in the full ensemble (not shown). We only focus our analysis then on the first five eigenfunctions of this analysis, $U_{n,k}$, as a function of frequency $\nu(n)$, where n is the frequency index and k is the eigenfunction index (see Fig. 3 and online color figure).

The scope of this analysis is to analyze the contribution of each of the 399 selected channels to these most representative top eigenfunctions, as a qualitative measure of their effective representativeness of the total variance contained in the radiance sample.

While eigenfunctions are representative of the variation of combined atmospheric species, by comparison with Fig. 1, we can distinctly see the first eigenfunction being mainly representative of the surface temperature signal. We notice that the 24 selected surface temperature and emissivity channels are among the most important contributors to this first eigenvector (green cross symbols).

The highest components of the second eigenvector are found in the 1200–1500-cm^{-1} region where methane, water vapor, N_2O, and SO_2 are normally active. We have indeed selected the majority of our methane (magenta crosses), water vapor (red crosses), N_2O (yellow crosses), and SO_2 (dark purple crosses) sounding channels among the highest peaking channels of this region. The highest components of the third and the fourth eigenvector fall in the 1000–1100-cm^{-1} range which is dominated by the ozone signal. We can confirm that our 53 ozone sounding channels (blue crosses) correspond

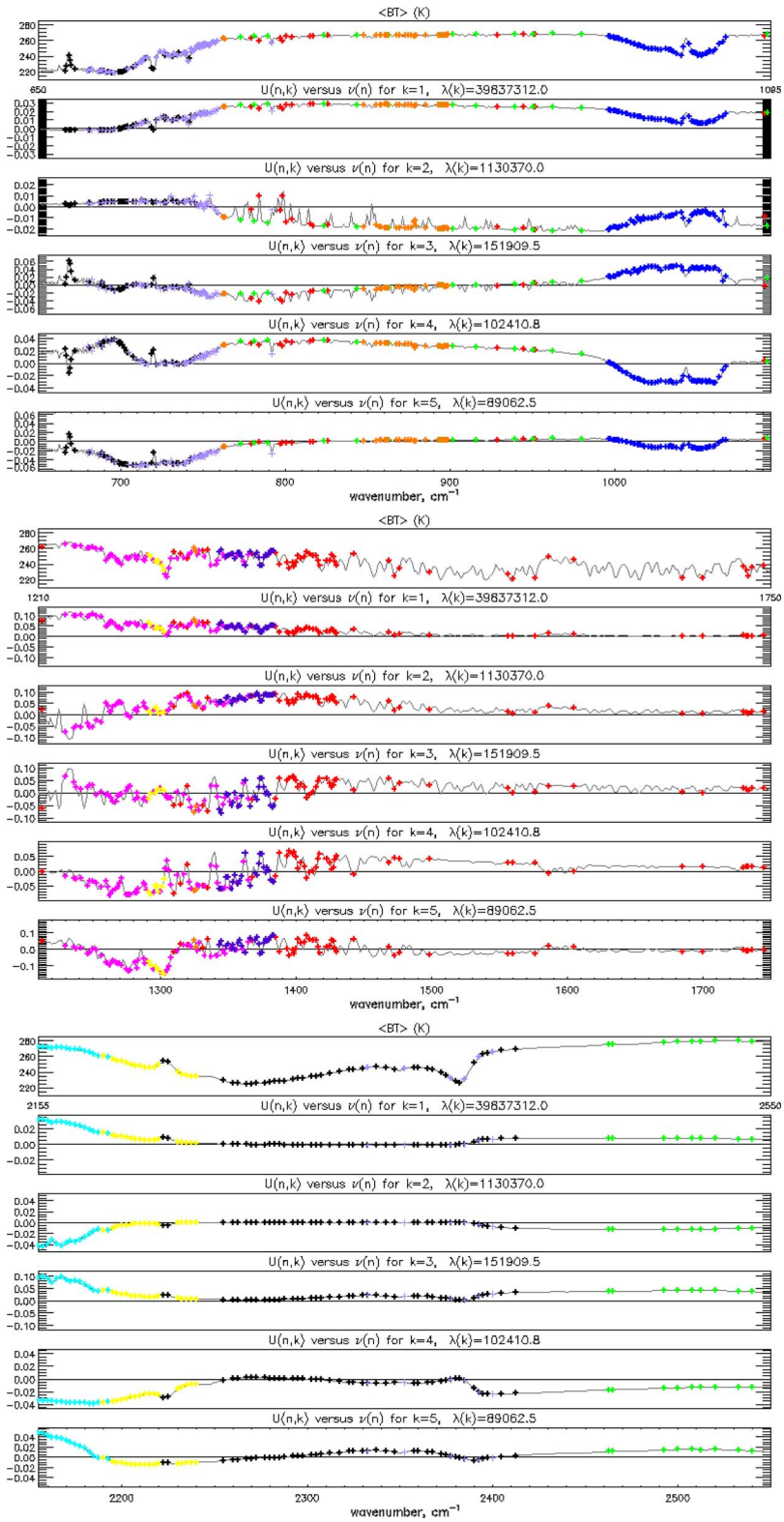

Fig. 3. SVD analysis of a global sample of CrIS proxy data. Top figure: Band 1. Middle figure: Band 2. Bottom figure: Band 3. On each subfigure, the top part is the average brightness temperature of the sample, followed by the first five eigenfunctions of the SVD analysis. Cross symbols indicate the location of the selected channels, color coded according to their main atmospheric sensitivity as in Fig. 1.

to the highest peaking channels of both eigenvectors. Eigenvectors 4 and 5 are dominated by the carbon dioxide band signal. The selected temperature and carbon dioxide channels (black and light purple crosses) are among the highest components of these two eigenvectors. The selected carbon monoxide (cyan crosses) and N_2O (yellow crosses) channels are among the highest peaking channels of the first and third eigenvectors.

The analysis just shown indicates a high representativeness of the proposed channel selection in the linear combinations of the orthonormal basis eigenvectors used to describe the main directions of variability contained in the full radiance spectra.

B. Variance Explained

As anticipated in the introduction, the scope of this channel selection is to optimize the system computational efficiency by selecting channels carrying unique pieces of information such that the internal variability contained in the original spectrum is still fully represented by the selection, up to instrumental noise. We now address the issue of quantitatively computing the actual fraction of the full spectrum variance, explained by the proposed channel selection.

We use a granule sample of full CrIS proxy radiance spectra (1305 channels) and compute a polynomial linear fit to the correspondent 399-channel selection spectra, as described by

$$\Delta R_{N_f,N_s} = A_{N_f,N_{sub}} \Delta R_{N_{sub},N_s} \tag{9}$$

where N_f is the size of the full spectrum (1305 channels), N_s is the number of spectra contained in one granule (5940), N_{sub} is the size of the channel selection, $\Delta R_{N_f,N_s}$ is the independent variable matrix of mean centered radiance values, $\Delta R_{N_{sub},N_s}$ is the corresponding predictor matrix, and $A_{N_f,N_{sub}}$ is the regression coefficient matrix, given by

$$A_{N_f,N_{sub}} = \left[\Delta R_{N_f,N_s} \Delta R^T_{N_s,N_{sub}}\right] * \left[\Delta R_{N_{sub},N_s} \Delta R^T_{N_s,N_{sub}}\right]^{-1}. \tag{10}$$

Once $A_{N_f,N_{sub}}$ is computed, we can derive the calculated $N_f x N_s$ matrix, as in

$$\Delta R^{calc}_{N_f,N_s} = A_{N_f,N_{sub}} \Delta R_{N_{sub},N_s}. \tag{11}$$

We then compare the variance contained in the original spectra, Var^{tot}, with the residual sum of squares, Var^{err}. Specifically, from the ratio of these two terms, we derive what in some texts is referred to as the *coefficient of determination* D [23], defined as

$$D = 1 - Var^{err}/Var^{tot}. \tag{12}$$

In statistical methods, D is a metric used to determine the goodness of a fit to a model. In regression, D is a statistical measure of how close the fit is to the original data points. The closer D is to unity, the better is the fit, i.e., the closer are the predictors (in our case, the actual 399 selected channels) to explain the total variance of the original data sample.

The coefficient of determination is a useful tool for testing the impact of new channels that are either used as a replacement of old ones or that have been incrementally added to a preexisting selection. Test channels are kept in the selection if they contribute at increasing the total variance explained. In least squares regressions though, like in our case, D increases with the number of predictors. As such, D is not meaningful to compare selections differing by the actual number of channels. For a meaningful comparison then, we resort to an *adjusted* coefficient of determination, D^{adj}, defined as

$$D^{adj} = \left(Var^{err}/Var^{tot}\right)\left(dof^{tot}/dof^{err}\right) \tag{13}$$

which represents a statistically unbiased version of D. Here, dof^{tot} is the degree of freedom of the population variance of the dependent variable and is given by $N_s^{tot} - 1$, where N_s^{tot} is the sample size of the dependent variable population. dof^{err} is the degree of freedom of the predictor population variance, $N_s^{tot} - N_{sub} - 1$, where N_{sub} is again the number of predictors (selected channels).

We first compute the coefficient of determination for each channel. The variance contained in the original sample, $Var^{tot}(n)$, and the residual sum of squares, $Var^{err}(n)$, corresponding to channel ν of index n, are defined by (14) and (15), respectively

$$Var^{tot}(n) = \sum_{s=1}^{N_s}\left[\Delta R_s(n) - (1/N_s)\left(\sum_{s=1}^{N_s}\Delta R_s(n)\right)\right]^2 \tag{14}$$

$$Var^{err}(n) = \sum_{s=1}^{N_s}\left[\Delta R_s(n) - \Delta R_s^{calc}(n)\right]^2. \tag{15}$$

We then have, for each channel

$$D^{adj}(n) = -\left[Var^{err}(n)/Var^{tot}(n)\right] * \left[(N_s - 1)/(N_s - N_{sub} - 1)\right]. \tag{16}$$

For completeness, we have plotted the full sequence of improvements obtained by incrementally adding each of the ten channel subsets described in the previous sessions, each including channels mainly sensitive to a given atmospheric species of interest. Results are plotted in Figs. 4 and 5 (see online color figures). For comparison, the results of a channel subset addition are plotted on top of the results of the previous partial selection. We use the same color code as that of the previous figures, and for brevity, we only display results from a midlatitude test granule. Multiple test granule results will be summarized in Fig. 6.

We start by considering the sole 24 selected surface temperature sensitive channels, and we perform the regression as in (9)–(11). We then compute the coefficient of determination for the full spectrum, according to (14)–(16). Here, $N_{sub} = 24$. Green symbols on Fig. 4 indicate the percentage variance explained by the selected 24 surface temperature and emissivity channels. Regions of the spectrum from where these 24 channels were selected are almost completely explained. This was expected due to the redundancy present across the spectrum.

When we add the 87 temperature channels (670- and 2300-cm^{-1} regions), we considerably improve the variance explained across most of the long-wave and short-wave regions. The addition of the 62 selected water vapor channels (which mainly fall in the 1200–1700-cm^{-1} region) brings the explained variance in the full midwave band to more than 99% (red cross symbols). A few percentage more is gained in the 1000–1100-cm^{-1} region with the addition of the 53 ozone sounding channels (blue cross symbols), and a few percentage

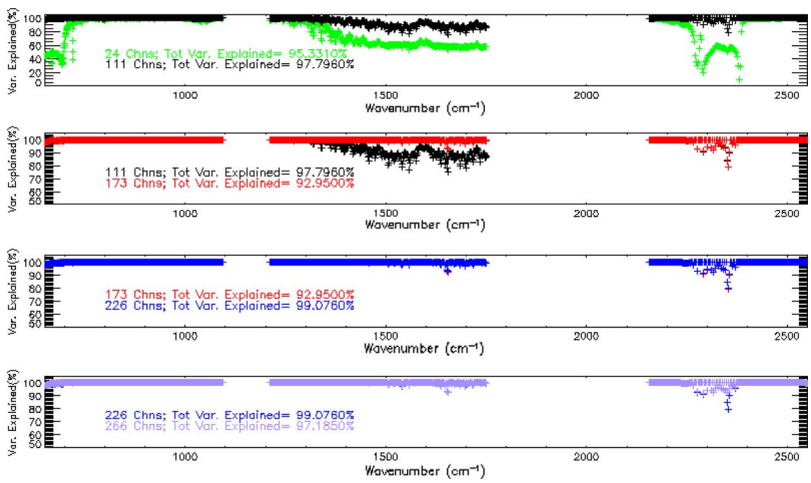

Fig. 4. Percentage variance explained. Different colors indicate different channel subsets used as predictors. Green curve: Surface temperature channel selection only (24 channels). Black curve: Previous subset plus 87 temperature sensitive channels. Red curve: Like black curve, with 62 additional water vapor sounding channels. Blue: With 53 additional ozone sounding channels. Light purple: With 53 additional carbon dioxide sounding channels. Captions on the figure report the total variance explained by the channel subset. See text for explanation.

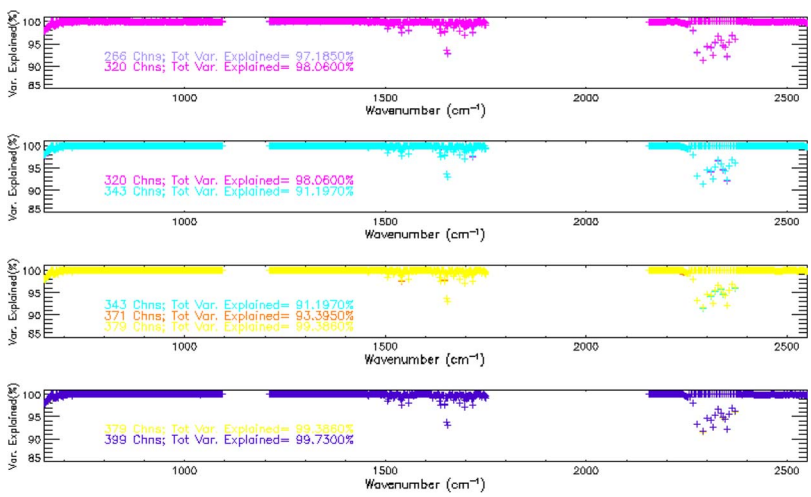

Fig. 5. Same as Fig. 4. The magenta curve is due to the addition of 54 methane sounding channels. The cyan curve represents the impact of 27 additional carbon monoxide sounding channels. The orange curve refers to the addition of 28 HNO_3 sounding channels, the yellow to 24 N_2O, and the dark purple to 24 SO_2. For comparison, the results of a channel subset addition are plotted on top of the results of the previous partial selection. Captions on the figure report the total population variance explained by the channel subset. See text for explanation.

improvement is obtained with the addition of the selected 53 CO_2 channels in the short-wave region. The addition of the remaining trace gas subsets brings an improvement of the order of 10^{-3} across the spectrum, as shown in Fig. 5. The reader should notice the change of scale in the reported values of the y-axis. The magenta curve is due to the addition of 54 methane sounding channels. The cyan curve represents the impact of 27 additional carbon monoxide sounding channels. The orange curve refers to the addition of 28 HNO_3 sounding channels, the yellow to 24 N_2O, and the dark purple to 24 SO_2.

In conclusion, the full 399-channel selection is seen to account for more than 99% of the total variance across the whole spectrum, except for the 600–700-cm^{-1} and 1700-cm^{-1} regions, where the variance explained is about 95%, and for the 2200–2300-cm^{-1} region, where it oscillates between 85% and 99%. This result was expected and originates from the lower representativeness of these spectral bands in our channel selection. This choice rested in the high instrumental noise of the 600–700-cm^{-1} region, in the lower spectral purity of the 1700-cm^{-1} region with respect to the 1400-cm^{-1} band (from where we have preferentially selected most of the water vapor channels), and in the high instrumental and calibration noise of the short-wave band. For the reasons just explained, retrieval users generally tend to avoid these low signal-to-noise spectral regions. Accordingly, we do not expect the lower representativeness of these regions to be penalizing in terms of retrieval application performance.

We now summarize the previous discussion by investigating the variance of the full population in the sample. We define

$$VAR^{TOT} = \sum_{n=1}^{N_f} Var^{tot}(n) \tag{17}$$

$$VAR^{ERR} = \sum_{n=1}^{N_f} Var^{tot}(n). \tag{18}$$

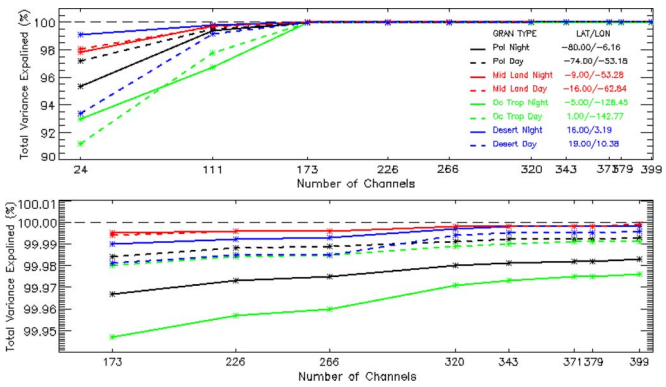

Fig. 6. Top: Total population variance explained by the channel selection after the incremental addition of each channel subset. The full 399-channel selection explains 99.99% of the total variance contained in the original 1305-channel spectrum. Bottom: Enlargement of the right side of top figure to show the improvement in the explained variance resulting from the incremental addition of the remaining trace gas sensitive channel subsets. See online color figure. Solid and dashed curves refer to nighttime and daytime test granules, respectively. Black, red, green, and blue curves refer to polar, midlatitude, tropical ocean, and desert granules, respectively, with average coordinates listed on the right side of the legend.

We then have, for the total population variance explained by each subset

$$D^{adj^{TOT}} = 1 - \left[\frac{VAR^{ERR}}{VAR^{TOT}}\right] * \left[\frac{(N_f * N_s - 1)}{(N_f * N_s - N_{sub} - 1)}\right]. \quad (19)$$

Results for the total variance explained, $D^{adj^{TOT}}$, upon the incremental addition of each subset are listed as captions on Figs. 4 and 5 and are now summarized in Fig. 6 (see online color figure). For this analysis, we have resorted to eight sample granules, differing by surface type and air mass, specifically nighttime (solid) and daytime (dashed) polar (black), midlatitude land (red), ocean tropical (green), and desert (blue) granules. Average latitude and longitude coordinates for each granule are reported on the right side of the legend on the top part of Fig. 6. We can see that the first 24 surface temperature and emissivity sounding channels account already for more than 90% of the full spectrum variability. The lowest representativeness occurs over the tropical ocean and daytime desert regions. This result was expected since, over these areas, most of the information arises from the large air mass variability. For the tropical case in particular, this large variability is attributable to clouds and water vapor. The addition of the 87 selected temperature sounding channels consistently improves the total variance explained, which now ranges between 95% and 99%. By adding the 62 selected water vapor channels, we can see that these first 173 channels alone (surface, profile temperature, and water vapor sensitive channels only) are already capable to account for more than 99% of the total variability contained in the full sample, over all test granules. The bottom part of Fig. 6 is an enlargement of the right side of the top figure and provides a better display of the improvement in the variance explained resulting from the incremental addition of the remaining trace gas sensitive channel subsets. We can notice that, over all geophysical regions, the total variance explained by the complete 399-channel selection amounts to about 99.99%.

The results just shown confirm that the internal variability contained in the original spectrum is fully explained by the proposed channel selection up to instrumental noise and is globally applicable. This implies that the replacement of the full 1305-channel list in favor of the proposed 399-channel selection will have no detrimental effects on data assimilation and retrieval performance, uniformly over the globe.

IV. Conclusion and Future Work

We have described the NOAA/NESDIS physically based methodology for the near-real-time channel selection of the CrIS. This channel selection is composed of 24 surface temperature and emissivity, 87 temperature, 62 water vapor, 53 ozone, 27 carbon monoxide, 54 methane, 53 carbon dioxide, 24 N_2O, 28 HNO_3, and 24 SO_2 sounding channels.

By means of a detailed SVD analysis, the proposed channel selection was found to carry a high representativeness of the atmospheric variability contained in the full radiance spectrum. Furthermore, by means of a regression analysis, we computed the total variance explained by the selection and found that the total internal variability contained in the original spectrum is fully explained by the proposed channel selection up to instrumental noise (99.99% variance explained). This analysis was performed over multiple geophysical regimes, demonstrating the global applicability of the proposed channel selection. These results ensure that the replacement of the full 1305-channel spectrum in favor of the proposed 399-channel selection will have no detrimental effects in terms of data assimilation and retrieval accuracy.

Test applications have shown this channel selection to be significantly advantageous in terms of computational efficiency. For example, the computational time of routine operations, such as the computation of radiance covariance matrices, has been observed to improve by a factor of 10 when going from assimilating full spectra down to the 399-channel selection.

Existing methodologies, as described in [5]–[7], follow a statistical iterative methodology [8] where successive tests are performed on the incremental addition of each channels. At each addition, the channel is retained if the information content (based on the entropy of the system) is observed to increase. For near-real-time operational applications, a *constant* channel selection is normally used, which is derived as an average from multiple optimal selections computed over the globe.

Differently from these statistical techniques, the NOAA/NESDIS methodology described in this paper employs a physically based procedure where channels are selected solely upon their spectral properties: High priority is given to spectral purity, avoidance of redundancy, and vertical sensitivity properties, along with low instrumental noise and global optimality.

The methodology described in this paper has been previously adopted for the operational channel selection of the AIRS and IASI sensors, [3], [9]. National Weather Prediction users have found great advantage in employing these AIRS and IASI near-real-time channel selections to expedite data assimilation processing, while guaranteeing good quality in the retrieval results [10]–[12]. Fourrie and Thepaut [6] have compared the global optimality of the AIRS NOAA/NESDIS and the statistically derived ECMWF *constant* channel selection. They found the

performances of the two selections to be comparable, with the NOAA/NESDIS selection providing good global optimality with respect to different background error specifications.

We can expect the proposed near-real-time CrIS channel selection to carry on AIRS and IASI heritage with similar robustness and global optimality. A cross-comparison of the information content contained in the AIRS, IASI, and CrIS operational channel selections is under exam. Vertical resolution and degrees of freedom of the signal as in [8] will also be part of a separate study, once the CrIS NOAA/NESDIS retrieval algorithm is fully optimized and retrieval products become operationally available.

Acknowledgment

The authors would like to thank W. Wolf for his useful suggestions on the topic of this paper, M. Divakarla for providing the Cross-track Infrared Sounder (CrIS) proxy data set, and E. Maddy for installing the CrIS SARTA radiative transfer model version 10a into the NOAA system.

References

[1] NASA Goddard Space Flight Center Greenbelt, MD, "Cross-Track Infrared Sounder," 2011. [Online]. Availabel: http://www.ipo.noaa.gov

[2] M. Goldberg, L. Qu, Y. McMillin, W. Wolf, L. Zhou, and M. Divakarla, "AIRS near-real-time products and algorithms in support of operational weather prediction," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 379–389, Feb. 2003.

[3] J. Susskind, C. D. Barnet, and J. Blaisdell, "Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 390–409, Feb. 2003.

[4] T. August, D. Klaes, P. Schlussel, T. Hultberg, M. Crapeau, A. Arriaga, A. O'Carroll, D. Coppens, R. Munro, and X. Calbet, "IASI on Metop-A: Operational level 2 retrievals after five years in orbit," *J. Quant. Spectrosc. Radiat. Transf.*, vol. 113, no. 11, pp. 1340–1371, Jul. 2012.

[5] F. Rabier, N. Fourrie, D. Chafai, and P. Prunet, "Channel Selection Methods for Infrared Atmospheric Sounding Interferometer Radiances," IASI Sci., Paris, France, Tech. Rep., 2002.

[6] N. Fourrie and J. Thepaut, "Validation of the NESDIS Near Real Time AIRS Channel Selection," Eur. Centre Medium-Range Weather Forecasts, Reading, U.K., Tech. Rep., 2002.

[7] A. Collard, "Selection of IASI channels for use in numerical weather prediction," *Q. J. R. Meteorol. Soc.*, vol. 133, no. 629, pp. 1977–1991, Oct. 2007.

[8] C. Rodgers, "Information content and optimization of high spectral resolution measurements," in *Proc. SPIE*, 1996, vol. 2830, pp. 136–147.

[9] A. Gambacorta, C. Barnet, W. Wolf, and M. Goldberg, "The Near Real Time IASI Channel Selection for Data Assimilation and Retrieval Purposes," NOAA NESDIS, Washington, DC, Tech. Rep., 2007.

[10] A. McNally, P. Watts, J. Smith, R. Engelen, G. Kelly, J. Thepaut, and M. Matricardi, "The assimilation of AIRS radiance data at ECMWF," *Q. J. R. Meteorol. Soc.*, vol. 132, no. 616, pp. 935–957, Apr. 2006.

[11] J. Le Marshall, J. Jung, J. Derber, M. Chahine, R. Treadon, S. J. Lord, M. Goldberg, W. Wolf, H. C. Liu, J. Joiner, J. Woollen, R. Todling, P. Van Delst, and Y. Tahara, "Improving global analysis and forecasting with AIRS," *Bull. Amer. Meteorol. Soc.*, vol. 87, no. 7, pp. 891–894, Jul. 2006.

[12] J. Jung and J. Derber, "Assimilation of IASI Data, North America Perspective," Eur. Centre Medium-Range Weather Forecasts, Reading, U.K., 2009.

[13] *World Meteorological Organization*, WMO, Geneva, Switzerland, 2012.

[14] L. Strow and S. Hannon, "NOAA Implemented Version v10a, Internal Documentation," NOAA, Washington, DC, 2011.

[15] A. Gambacorta, C. Barnet, W. Wolf, T. King, E. Maddy, M. Divakarla, and M. Goldberg, "Methodology and Information Content of the NOAA/NESDIS Operational Channel Selection for the Cross-Track Infrared Sounder (CrIS)," NOAA NESDIS, Washington, DC, Tech. Rep. 133, 2011.

[16] L. D. Kaplan, M. T. Chahine, J. Susskind, and J. E. Searl, "Spectral band passes for a high precision satellite sounder," *Appl. Opt.*, vol. 16, no. 2, pp. 322–325, Feb. 1977.

[17] L. D. Kaplan, "The Use of High-Frequency Infrared Radiometry for Remote Atmospheric Probing With High Vertical Resolution," Dept. Meteorol., Massachusetts Inst. Technol., Cambridge, MA, Tech. Rep., 1969.

[18] M. T. Chahine, "Remote sounding of cloudy atmospheres in a single cloud layer," *J. Atmos. Sci.*, vol. 31, no. 1, pp. 233–243, Jan. 1974.

[19] C. Barnet, J. Blaisdell, and J. Susskind, "Practical methods for rapid and accurate computation of interferometric spectra for remote sensing applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 169–183, Jan. 2000.

[20] L. McMillin, M. Goldberg, H. Ding, J. Susskind, and C. Barnet, "A forward calculation for interferometers: Method and validation," *Appl. Opt.*, vol. 37, no. 14, pp. 3059–3068, May 1998.

[21] A. Gambacorta, "The IASI Algorithm Theoretical Basis Documentation, NOAA NESDIS, Washington, DC," 2011. [Online]. Available: http://www.star.nesdis.noaa.gov/smcd/spb/iosspdt/qadocs/

[22] C. Barnet, M. Divakarla, M. Goldberg, T. King, G. Guo, and N. Nalli, "The CrIS/ATMS Proxy Data Package for the CrIMSS EDR Generation and Evaluation (Release 1.0), IPO Report, Nat. Ocean. Atmos. Admin., Washington, DC," Tech. Rep. 2010. [Online]. Available: ftp://ftp2.orbit.nesdis.noaa.gov/smcd/tking/IPORELV1.0

[23] B. Everitt, *Cambridge Dictionary of Statistics*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2002.

Antonia Gambacorta received the Ph.D. and M.S. degrees in atmospheric physics from the University of Maryland, Baltimore County, MD, in 2004 and 2008, respectively, and the M.S. and B.S. degrees in applied physics from the Universita degli Studi di Bari, Bari, Italy, in 2001.

She has been with NOAA/NESDIS/Satellite Applications and Research (STAR) since 2004 and specializes in the field of microwave and infrared remote sensing. Since 2006, she has been responsible for the development and operational transition of the Infrared Atmospheric Sounding Instrument (IASI) system. She is currently in charge of the development, validation, and operational transition of the NOAA Unique Cross-Track Infrared Sounder/Advanced Technology Microwave Sounder Processing System (NUCAPS) retrieval algorithm. In parallel, she has been applying satellite-derived products to conduct research studies in the field of the hydrological cycle and climate change. She is currently chairing the vertical profile section of the Global Energy and Water Cycle Experiment Water Vapor Assessment and acting as point of contact for the NOAA Atmospheric Infrared Sounder, IASI, and CrIS water vapor retrieval products during this experiment.

Christopher D. Barnet received the B.S. degree in electronics technology and the M.S. degree in solid-state physics from Northern Illinois University, Dekalb, in 1976 and 1978, respectively. In 1990, he received the Ph.D. degree from New Mexico State University, Las Cruces, in remote sensing of planetary atmospheres using visible and infrared instruments aboard the Voyager spacecraft. His postdoctoral research at NASA and the Canadian Institute for Space and Terrestrial Science focused on ultraviolet, visible, and near-infrared observations of the planets using the Hubble Space Telescope.

Since 1995, he has worked on advanced algorithms for terrestrial infrared and microwave remote sensing and has actively supported NASA's Advanced Infrared Sounder science team and the National Polar-orbiting Operational Environmental Satellite System (NPOESS) Sounder Operational Algorithm Team. In June 2003, he joined the Office of Research and Applications of NOAA/NESDIS where he is exploiting operational sounder missions to provide the first global understanding of carbon monoxide, carbon dioxide, and methane in the free troposphere. These measurements will contribute to the understanding of the terrestrial carbon cycle and climate change.