Received: 25 November 2017 | Revised: 18 February 2018 | Accepted: 04 April 2018

DOI: 10.1002/qj.3315

RESEARCH ARTICLE

Quarterly Journal of the Royal Meteorological Society RMetS

Towards more realistic hypotheses for the information content analysis of cloudy/precipitating situations – Application to a hyperspectral instrument in the microwave

Filipe Aires[1,2] | Catherine Prigent[1,2] | Stefan A. Buehler[3] | Patrick Eriksson[4] | Mathias Milz[5] | Susanne Crewell[6]

[1]LERMA, Observatoire de Paris, Paris, France
[2]Estellus, Paris, France
[3]Universität Hamburg, Hamburg, Germany
[4]Chalmers University of Technology, Gothenburg, Sweden
[5]LTU, Lulea, Sweeden
[6]Cologne University, Cologne, Germany

Correspondence
F. Aires, LERMA, Observatoire de Paris, 77 av. Denfert-Rochereau, 75014 Paris, France.
Email: filipe.aires@estellus.fr

Funding information
European Space Agency contract N 4000105721/12/NL/AF,

Information Content (IC) analysis can be used before an instrument is built to estimate its retrieval uncertainties and analyse their sensitivity to several factors. It is a very useful method to define/optimize satellite instruments. IC has shown its potential to compare instrument concepts in the infrared or the microwave. IC is based on some hypotheses such as the the Gaussian character of the radiative transfer (RT) and instrument errors, the first-guess errors (Gaussian character, std and correlation structure), or the linearization of the RT around a first guess. These hypotheses are easier to define for simple atmospheric situations. However, even in the clear-sky case, their complexity has never ceased to increase towards more realism, to optimize the assimilation of satellite measurements in numerical weather prediction (NWP) systems. In the cloudy/precipitating case, these hypotheses are even more difficult to define in a realistic way as many factors are still very difficult to quantify. In this study, several tools are introduced to specify more realistic IC hypotheses than the current practice. We focus on microwave observations as they are more pertinent for clouds and precipitation. Although not perfect, the proposed solutions are a new step towards more realistic IC assumptions of cloudy/precipitating scenes. A state-dependence of the RT errors is introduced, the first-guess errors have a more complex vertical structure, the IC is performed simultaneously on all the hydrometeors to take into account the contamination effect of the RT input uncertainties, and the IC is performed on a diversified set of cloudy/precipitating scenes with well-defined hydrometeor assumptions. The method presented in this study is illustrated using the HYperspectral Microwave Sensor (HYMS) instrument concept with channels between 6.9 and 874 GHz (millimetre and sub-millimetre regions). HYMS is considered as a potential next generation microwave sounder.

KEYWORDS
clouds and precipitation, hyperspectral, hydrometeors, microwave remote sensing

1 | INTRODUCTION

Information Content (IC) has been developed to theoretically estimate the uncertainty of atmospheric retrievals from satellite sounders (Rodgers, 1976; 1990). It has become a very important methodology to define and optimize new instruments (Boukabara and Garrett, 2011). IC allows comparison of instrument configurations by varying the number and location of channels and their associated instrumental noise levels (Lipton *et al.*, 2009; Blackwell *et al.*, 2011; 2012). It is also used to estimate the impact of new satellite measures in numerical weather prediction (NWP) forecasts (Bauer *et al.*, 2006), or to help develop the satellite retrieval algorithms before observations are available (Pellet and Aires, 2016).

Three pieces of information are required for IC: The observation-error specifications (covariance matrix **R**), the sensitivity of the channels to the parameters to be retrieved (Jacobian matrix **H**), and the *a priori* information available before the use of the satellite observations (e.g. weather forecasts at NWP centre). These three components are then combined:

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H} \quad (1)$$

to determine the retrieval error covariance matrix **A**. This quantity characterises the retrieval error estimates. **B** is covariance matrix of background error.

The IC of Equation 1 is based on some hypotheses such as the Gaussian probability character of the sources of uncertainty (radiative transfer (RT), instrument or background errors) and the linearization of the RT around a first guess (Rodgers, 2000). Even for clear-sky scenes, the complexity of these hypotheses has not ceased to increase: the NWP centres started by using constant background errors, then latitudinal variations were introduced, and nowadays more complex state-dependence on the atmospheric situation is added. This increase in complexity intends to improve the realism of the hypotheses in order to obtain more robust and reliable retrieval uncertainty estimates. For the cloudy and precipitating scenes, this need for realism is even stronger as the hypotheses have a strong impact on the obtained results. Unfortunately, much required information is difficult to quantify, or is still not well controlled by the scientific community, as we will see in the following.

- *Instrument errors* are generally provided by the instrument constructor. It is important to know if the channel errors are correlated to each other or not (Bormann *et al.*, 2017). A second point is to know if the instrument errors are state-dependent or not. This is very important for infrared instruments (Aires *et al.*, 2002), but they are generally considered constant for microwave instruments.
- *Radiative transfer errors* are not yet well characterized. For precipitating scenes, RT errors are much more important than the instrumental errors (Di Michele and Bauer, 2006). Several types of uncertainty sources can be identified: First, the RT model can have "structural" errors due to some practical approximations. For instance, the vertical discretization scheme introduces vertical resolution constraints, or the way a Monte-Carlo is implemented can result in approximations (e.g. related to the number of samples; Wang *et al.*, 2012). This type of error is not generally the major issue. Second, in the RT errors, it is also very important to take into account all the geophysical variables that might potentially impact the RT. For instance, the uncertainty related to the surface emissivity needs to be considered when using surface-sensitive channels, even if the emissivity is not the variable of interest. This is called the "geophysical" noise in Bauer *et al.* (2005). In Aires *et al.* (2016), it is shown that the geophysical noise, or contamination effect, can be solved by performing the IC simultaneously on all the variables. But this is often neglected in common practice. Lastly, some physical phenomena are still not well quantified. For instance, liquid and frozen hydrometeors can scatter the microwave radiation, especially at high microwave frequencies. The scattering effect is particularly difficult to quantify as it depends upon a large range of parameters that are highly variable and hard to estimate. The size distribution of the particles, liquid or frozen, the density or the shape of the frozen hydrometeors have a strong impact on the RT (Eriksson *et al.*, 2015; Galligani *et al.*, 2015; Wang *et al.*, 2017). These parameters are poorly known; cloud mesoscale models do not provide them and aircraft *in situ* measurements have proved their large variability. Under cloudy conditions at high microwave frequencies, the scattering effect is a major source of uncertainty that is very difficult to quantify. Uncertainties related to these hydrometeor parameters would require their quantification and then propagation using the sensitivity. This is still not obtainable to the scientific community because hydrometeor uncertainties are not yet well controlled.
- *First guesses and associated background errors* for precipitating scenes are also difficult to define. A climatology of hydrometeors would not provide a first guess of enough quality (due to the large variability of these parameters and the highly nonlinear character of the RT for these scenes). Therefore, the first guess of precipitating scenes is often obtained from atmospheric models (global circulation models or cloud-resolving models) by using perturbations on the state variables and analysing the resulting variability in the forecasts (Bauer *et al.*, 2006). The covariance of these forecast departures is used as a proxy covariance matrix for the first-guess errors, but it is difficult to know *a priori* what is the quality of this approximation (due to the quality of the model, the quality of the initial state variables, or the quality of the forecast itself). It is likely that this procedure will not take into account physical uncertainties such as: the type of hydrometeor (liquid or ice particles) for the situation (a neural network classification scheme such as in Wang *et al.* (2017) has always some uncertainty, the location of layers in the vertical that can be highly ambiguous in terms of RT, and the poor representation in the model of the correlations of the hydrometeors.
- *The linearization of the RT* around a first guess is more complex for precipitating scenes because the first guess is less precise and is difficult to obtain (as we have just noted), and the RT for cloudy/precipitating scenes is more nonlinear. So the reliability of the hydrometeor Jacobians used in the IC (Bauer *et al.*, 2005) is weaker than in the clear-sky case.

The goal of our study is to propose an IC framework for microwave instruments for cloudy/precipitating scenes which is more realistic than that generally used in the literature. Considering all the difficulties that were just listed, it will

not be possible here to be exhaustive and solve all the issues, but we propose some improvements to the current practice – a new step towards more realistic IC assumptions. This framework could be the basis for discussions in the community so that the community converges to a more realistic IC practice. This will strengthen the confidence we can obtain from IC analysis on cloudy/precipitating scenes, with potential improvements on the definition of new satellite missions, or the preparation of the NWP assimilation for next generation microwave observations.

In order to illustrate these ideas, we will use a complex mission project. HYMS is an HYperspectral Microwave Sensor that has been proposed to increase in the future the number of microwave channels for satellite remote sensing of the Earth (Lipton, 2003; Bauer and Di Michele, 2007; Blackwell *et al.*, 2011; Boukabara and Garrett, 2011). Several studies have been performed in order to select an optimal combination of channels (Mahfouf *et al.*, 2014; Pellet and Aires, 2016; Birman *et al.*, 2017) for the definition of the instrument. In Aires *et al.* (2015), the potential of such an instrument was analysed for clear-sky conditions; we will here consider the more complex cloudy/precipitating case. All the results presented are over ocean. The sensitivity has been assessed on all the oxygen and water vapour absorption bands separately for temperature, water vapour, and hydrometeor profiling. The bands are also considered together. The necessary ingredients for the IC analysis are first prepared and presented. A representative database of cloudy atmospheric profiles is selected, along with their background-error covariance matrix. RT simulations are performed on this dataset (the radiances as well as the Jacobians). Different hypotheses have been suggested for the estimation of the instrument and radiative transfer noises. The IC analysis is then presented, for temperature (T), humidity (Q), cloud liquid water content (CLW), cloud ice water content (CIW), rain water content (RWC), and snow water content (SWC) profile inversions, under various assumptions.

The HYMS instrument is briefly presented in section 2 (Aires *et al.*, 2015 give more details), together with the tools used in this study. The IC framework proposed in this paper is described in section 3. Section 4 provides the results. Conclusions and perspectives are proposed in section 5.

2 | THE HYMS INSTRUMENT AND TOOLS USED IN THIS STUDY

2.1 | The HYMS instrument

Several studies have already explored the potential of the higher spectral sampling in the microwave (Lipton, 2003; Bauer and Di Michele, 2007; Blackwell *et al.*, 2011), and the concept of a HYMS was discussed by Boukabara and Garrett (2011). Blackwell *et al.* (2011) analysed the performances of a hyperspectral instrument deployable on polar or geostationary satellites. They showed that with 64 channels around the 118.75 GHz oxygen absorption line and 16 around the 183.31 GHz water vapour absorption line, the root-mean-square (RMS) errors in temperature profiling could be reduced by 0.5 K on average, compared to a traditional instrument such as the Advanced Microwave Sounding Unit (AMSU-A and B), and that the RMS error in relative humidity is also reduced by about 5% with respect to the *a priori* RMS. The improvements to the retrieval of the temperature and water vapour profiles are attributable partly to the high density of the weighting functions. Boukabara and Garrett (2011) performed a pilot study highlighting the benefits of a HYMS for temperature profiling, especially under cloudy conditions. They simulated the performance of an instrument over the ocean, with channels spanning the 1–330 GHz portion of the microwave spectrum at a spectral resolution of 100 MHz. Reduction in the temperature profile error is evident when 100 channels are used in the retrieval, with a strong benefit under cloudy conditions where IR sounders are error prone at best and useless for opaque clouds. In this paper, we use the mission concept introduced in Aires *et al.* (2015).

TABLE 1 Overview of the 276 frequency bands. The spectral resolutions together with the number of channels are given for each band.

Spectral band (GHz)	Type	No. of channels (resolution 1)	Resolution (MHz)
Around 60	O_2 – Temperature	94	100
Around 118	O_2 – Temperature	51	200
Around 183	H_2O – Water vapour	51	400
Around 325	H_2O – Water vapour	21	1,000
Around 425	O_2 – Temperature	17	1,000
Around 448	H_2O – Water vapour	17	1,000
From 6.9 to 874	Window channels	25	—

Two different channel sets were defined: (a) a set of sounding channels, in the vicinity of prominent spectral absorption lines to allow the retrieval of vertical profiles of atmospheric state parameters, and (b) a set of window channels, located between spectral lines, to quantify cloud and precipitation, to sense lower in the atmosphere, and to account for the surface contribution. The characteristics of the bands are presented in Table 1, with the two spectral resolutions that will be tested in this study. The lowest spectral resolutions for each frequency band are selected to provide comparisons with the instruments planed for MetOp Second Generation (Kangas *et al.* 2012). This is to keep reasonable instrument noise figures with the present technology. The 100 MHz resolution for the 50–60 GHz oxygen band is below the spectral resolution of some of the MicroWave Sounder (MWS) channels (Kangas *et al.* 2012) for high-altitude sounding, but our study concentrates on the tropospheric retrievals for NWP. Then a factor of 10 is applied on the initial bandwidth for the second resolution (Resolution 2, with 10 MHz bandwidth for the 50–60 GHz oxygen band).

2.2 | The database of atmospheric profiles

The cloudy atmospheric profiles are extracted from a ECMWF database. The need for RT simulations of the

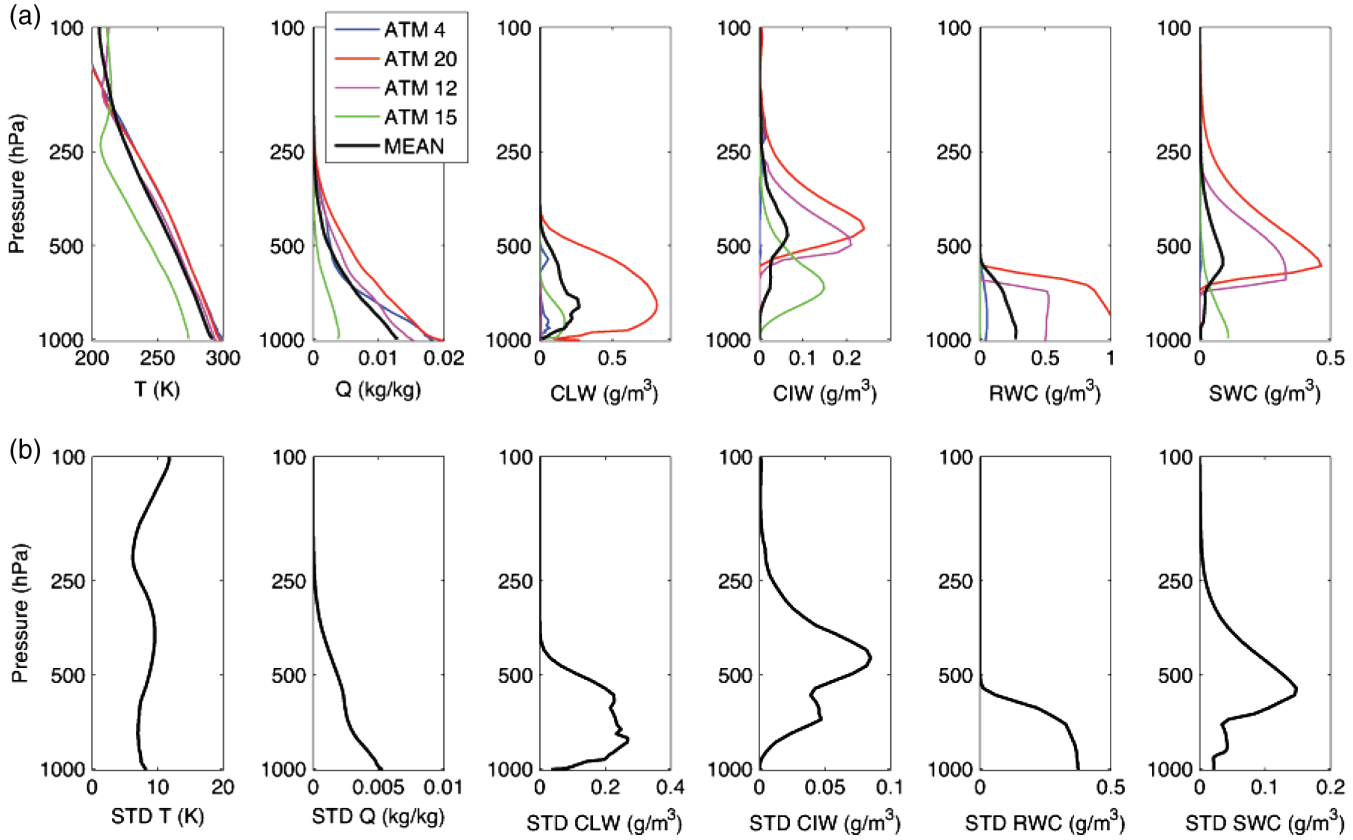

FIGURE 1 (a) Four representative atmospheric profiles (ATM 4, 20, 12, 15) for the six atmospheric variables (a) T, (b) Q, (c) CLW, (d) CIW, (e) RWC, and (f) SWC. The average over the 25 atmospheric situations is shown in black. (g)–(l) show the corresponding standard deviations [Colour figure can be viewed at wileyonlinelibrary.com]

full Jacobians for these atmospheres (including the Jacobians in each hydrometeor species) is a very strong constraint, due to the computation time of these Jacobians. For each cloudy profile, the background covariance matrices **B** also have to be available. To satisfy these two requirements, 25 situations have been selected by Météo-France from the ECMWF database (Birman *et al.*, 2017 gives the spatial location of these situations). Four hydrometeor species are considered: CLW, CIW, RWC, and SWC. The selected atmospheres are representative of a large panel of T, Q and hydrometeor profile (CLW, CIW, RWC, and SWC) variations.

The atmospheric profiles are taken from short-range forecasts from the ECMWF model (CY32R3) which has been run with a T799 spectral truncation (25 km) and 91 vertical levels from July 2006 to June 2007. Forecasts are over 42, 48, 54 and 60 h from days 1, 10 and 20 of every month. The database has been kindly provided by Sabatino Di Michele (ECMWF). This database is rather similar to the one from Chevallier *et al.* (2006), but it contains additional information necessary to perform this study.

Figure 1a–f represent four representative atmospheric profiles of the six quantities (T, Q, CLW, CIW, RWC, and SWC) selected over the 25 atmospheres, with the average over the 25 atmospheres shown in black. Figure 1g–l provide the corresponding standard deviations. As expected from models, the hydrometeor quantities are rather different, from cloud ice to rain. Cloud ice in models are representative of small ice particles (cirrus) in the upper atmosphere, associated with rather low maximum quantities. The other frozen species are included in the snow variable (in some models there is the addition of graupel and/or hail species); related particles can be much larger. Remember that snow here does not only represent frozen precipitation but also frozen particles in clouds. The liquid quantities (cloud and rain) are significantly larger than the frozen contribution.

2.3 | The radiative transfer simulations

The RT simulations were performed by version 2.2 of the Atmospheric Radiative Transfer System (ARTS; Eriksson *et al.*, 2011), with the Discrete Ordinate Iterative (DOIT; Emde *et al.*, 2004) method as the scattering solver applied. The general settings for the cloudy-sky simulations correspond to a satellite instrument operating in a Sun-synchronous orbit and looking toward the Earth. The radiative transfer is performed at nadir (180° zenith angle in ARTS) with a constant value of surface emissivity of 0.6 for ocean surfaces. To calculate the emission and scattering properties of cloud and precipitation, the decision was made to follow as much as possible the hypotheses suggested by the recent work of Geer and Baordo (2014).

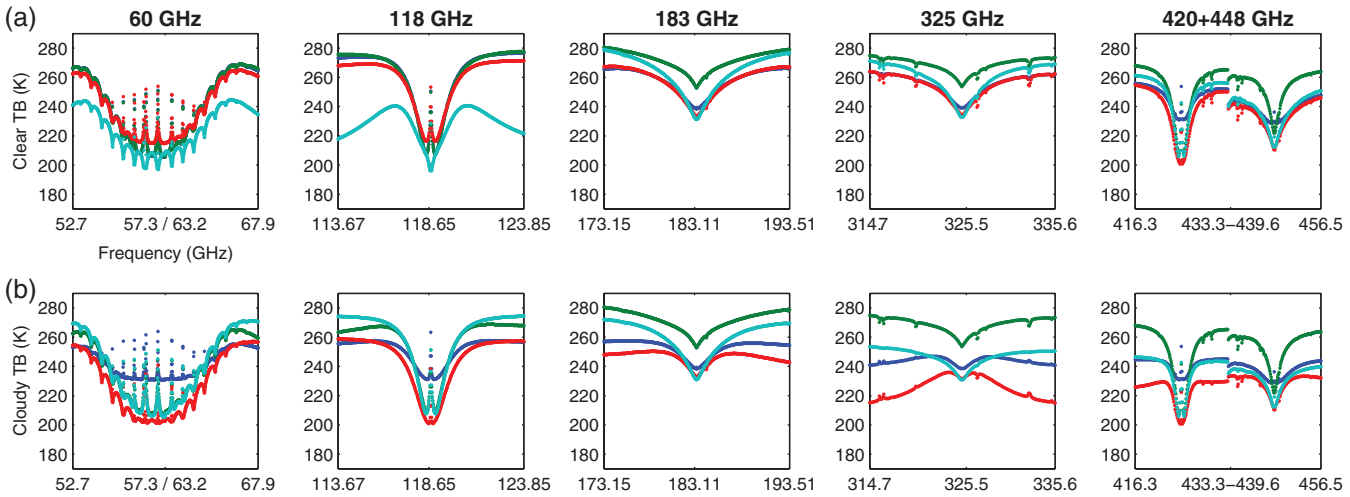

FIGURE 2 Brightness temperatures for the four atmospheric situations shown in Figure (1) with no hydrometeors, for bands (a) 60 GHz, (b) 118 GHz, (c) 183 GHz, (d) 325 GHz, and (e) 420+448 Ghz. (f)–(j) are as (a)–(e), but including all hydrometeors [Colour figure can be viewed at wileyonlinelibrary.com]

For cloud liquid water content, the size distribution is calculated using a modified gamma distribution for cloud water after Geer and Baordo (2014). Mie scattering is calculated for liquid water spheres. The shape of the particle size distribution is independent of temperature and cloud liquid water content.

For cloud ice, a modified gamma distribution is selected. The density of the ice particles is 900 kg m^{-3}. The optical properties are calculated following the Mie theory (Mätzler, 2002) based on dielectric properties from Mätzler (2006). For the snow, the Fields *et al.* (2005) size distribution is chosen. For each particle size in the distribution, the scattering properties are calculated using the sector-like snow flakes from Liu (2008) up to 360 GHz and the aggregates from Hong *et al.* (2004) above 400 GHz (as the Liu estimates are not available at these frequencies).

The resulting simulations are presented in Figure 2a–e for the same four atmospheres as in Figure 1 for the clear-sky case, and in Figure 2f–j for the same atmospheres but with all hydrometeor quantities included.

The O_2 band around 60 GHz is not affected much by the hydrometeors, although a brightness temperature (TB) increase is noticed in the right wing of the lines (around 70 GHz) for some atmospheres, related to the emission by liquid particles. Around 183 GHz, some significant decrease in the TBs is observed associated with scattering over frozen clouds. The same colours are always used for the same atmospheres: the atmosphere in green in Figure 1 containing a significant amount of snow produces a strong decrease of the TBs when including the hydrometeors. With increasing frequencies, the scattering effect increases dramatically, as expected. Some instability in the simulations can be observed for the higher frequencies, in the case of strong scattering; for extreme cases, the simulation set-up can encounter some difficulties but this is very localized and will not affect the analysis.

It is important to note that, with these hydrometeor definitions, cirrus clouds and other high ice clouds contain a much larger mass of snow particles (SWC) than ice particles (CIW). As a consequence, as will be shown below, CIW is hard to detect, but high ice clouds are nevertheless clearly detectably by their SWC.

2.4 | Radiative transfer Jacobians

For ARTS radiative transfer simulations including cloud scattering, no analytical Jacobians are available at the time of this study. The calculation of the Jacobians is therefore performed by applying perturbations on every vertical level in the relevant atmospheric state profiles. The simulated radiances for the perturbed profiles are then subtracted from the simulation based on the unperturbed profiles. The difference in TB has then been rescaled using the applied perturbations.

For calculating these Jacobians, the following perturbations have been applied:

1 K for temperature;
1 ppmv (1.0×10^{-6}) for humidity;
1×10^{-6} (kg m^{-2}) for rain (but Jacobians are converted to RWC in kg m^{-3});
and 1×10^{-6} kg m^{-3} for CLW, CIW and SWC.

For reasons of consistency, the Jacobians for CLW, CIW, RWC and SWC are provided in terms of mass content (kg m^{-3}). As the treatment of rain uses rain-rates as input, the perturbations were applied in rain-rates, but for the calculation of the Jacobians, the perturbation in rain-rates was converted to rain water content. Note that perturbations for certain parameters, and thus Jacobians, were not calculated for certain altitudes as the calculation was unphysical (e.g. frozen quantities at temperatures above 280 K or liquid quantities below 250 K). For these levels we interpolated the Jacobians linearly between the levels with valid Jacobians. For cases where these were the lowest or the highest levels,

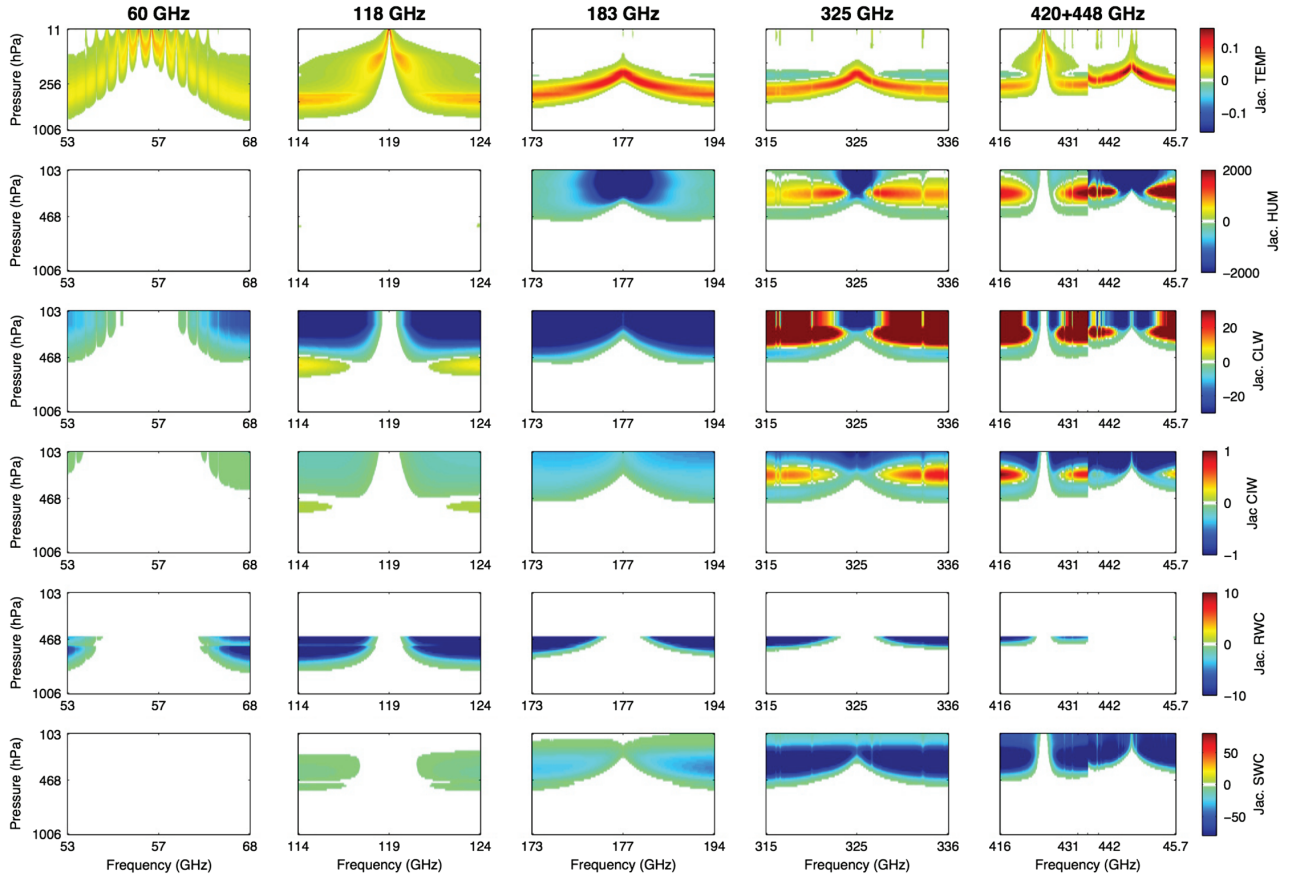

FIGURE 3 Jacobians for (top to bottom) T, Q, CLW, CIW, RWC and SWC, for bands (from left to right) at 60, 118, 183, 325 and 420+448 GHz [Colour figure can be viewed at wileyonlinelibrary.com]

we extrapolated the Jacobians using the value of the last valid level.

Figure 3 represents the T, Q, CLW, CIW, RWC and SWC Jacobians simulated by ARTS, on an atmospheric situation (situation number 1). The Jacobians for the channels intended for atmospheric sounding show the expected behaviour. The Jacobians for water vapour are sensitive in the troposphere, mainly for the channels situated around prominent water lines. The further the channel from the line centre, the larger is its sensitivity at low altitude. The Jacobians in temperature are significant for most sounding channels. This is expected, since the Planck function, which governs the thermal emission, is a function of temperature.

The values of the Jacobians at a given altitude will obviously depend upon the presence of the different hydrometeor layers. Little information is expected from the lower atmospheric layers, especially for the high frequencies with large atmospheric opacities. The magnitude of the Jacobians are very different for the different hydrometeor species. One has to remember that the Jacobians for each layer are calculated around the profile value of each constituant. A nonlinear behaviour is expected for the scattering by the frozen phase for instance, and the Jacobians can change drastically with the ice content for a given layer. In addition, the simulation hypotheses also have to be kept in mind. For the cloud ice, the particles are expected to be small with the selected size distribution (of the order of 10 to 20 μm) and their scattering efficiency will be very limited, even at high frequency. On the contrary, the snow particles can be large and strongly affect the radiation down to 183 GHz.

Jacobians are important information because they prove the sensitivity of each channel to each geophysical variable and for each vertical layer. Jacobians are also the quantity used in retrieval schemes (e.g. 1D-Var) or IC analysis. It is therefore important to verify that the RT code estimates reliable Jacobians. However, the interpretation of Jacobians can be misleading; e.g. a same physical unit is used for all the vertical layers but the range of variability of humidity in the lower or upper layers differs by orders of magnitude, so a Jacobian value can be very important in upper layers but be completely unrealistic because the humidity variation is too big for that layer.

2.5 | Perturbation impact on TBs

To better evaluate the sensitivity of a channel to the geophysical variables considered in this study, it is often simpler to monitor the TB impact of a perturbation for a profile than to analyze the Jacobians. The TB perturbations ΔTB_x for the change of variable x are defined as $JAC_x \cdot \Delta x$. Four profiles within the 25 have been selected to represent the large variability of situations. Figure 4 represents, from top to bottom,

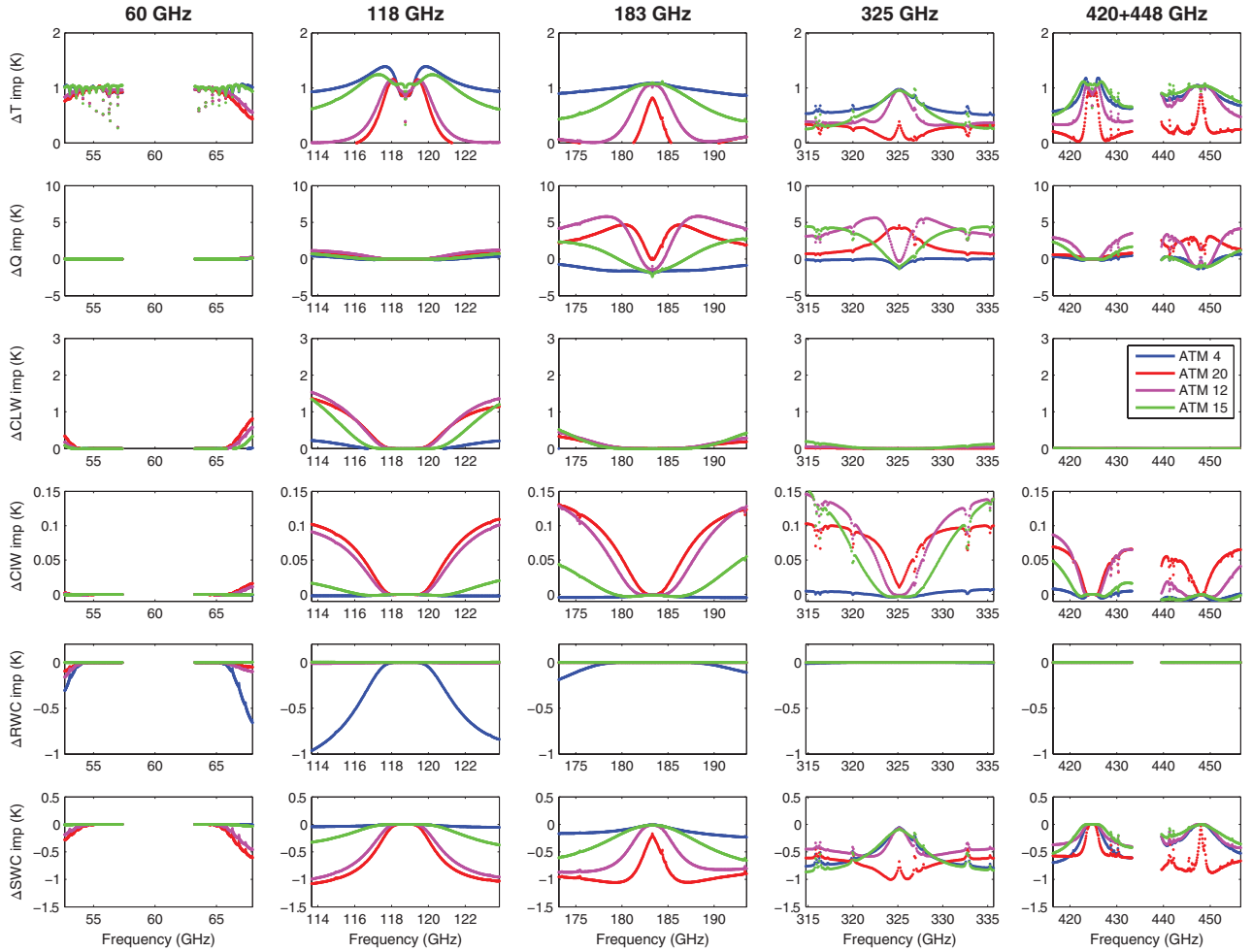

FIGURE 4 TB impact from (top to bottom) T, Q, CLW, CIW, RWC and SWC perturbations (1 K for T, 20% for Q, and 40% for the hydrometeor quantities), for (left to right) bands at 60, 118, 183, 325 and 420+448 GHz. Results are provided for the four chosen atmospheric situations [Colour figure can be viewed at wileyonlinelibrary.com]

the TB perturbations due an increase of 1 K of the temperature profile, 20% of the humidity (but with a constraint to be lower than 100%), and 40% of the cloud liquid, cloud ice, rain, or snow profiles, for the four atmospheric situations previously selected. The difference with the Jacobians here is that all the layers of the profile are changed. Furthermore, the change of each variable layer is not the same for each layer. The Jacobians are more related to retrieval of profiles, the perturbation impact on TBs are more linked to the retrieval of vertically integrated quantities.

3 | HYPOTHESES PROPOSED FOR THE INFORMATION CONTENT

In this section, the hypotheses that we propose to feed the IC Equation 1 are presented.

3.1 | Instrument noise, $\mathbf{R}_\mathrm{I}$

In the IC analysis approach, the observation noise **R** needs to be specified. This is important because the available information for the retrieval is related to the signal-to-noise ratio of the satellite observations. As mentioned in the Introduction, there are two components in these observation uncertainties: the instrumentation and RT uncertainties.

The instrument error, R_ϵ is the minimum detectable change in brightness temperature. It is given by the classical radiometric equation:

$$R_\epsilon = (T_\mathrm{rec} + T_\mathrm{ant})/\sqrt{Bandwidth \cdot \tau}, \tag{2}$$

where T_rec is the equivalent noise temperature of the receiver, T_ant is the antenna temperature that is approximately equal to the scene brightness temperature, *Bandwidth* is the bandwidth of the receiver, and τ is the integration time. For simplicity's sake, this radiometric equation neglects the fluctuation of the receiver gain as well as the calibration errors (Hersman and Poe, 1981). The instrument noise levels used in this study for the HYMS concept rely on recent receiver noise specifications. From a collection of current state-of-the-art receivers, the following formula has been derived (Aires *et al.*, 2014):

$$T_\mathrm{rec}(K) = 4.5 \times F(GHz) + 30. \tag{3}$$

An average T_ant of 270 K is assumed. The selected integration time (τ) is 20 ms for all channels, corresponding to the integration time for MWS on board MetOp-SG.

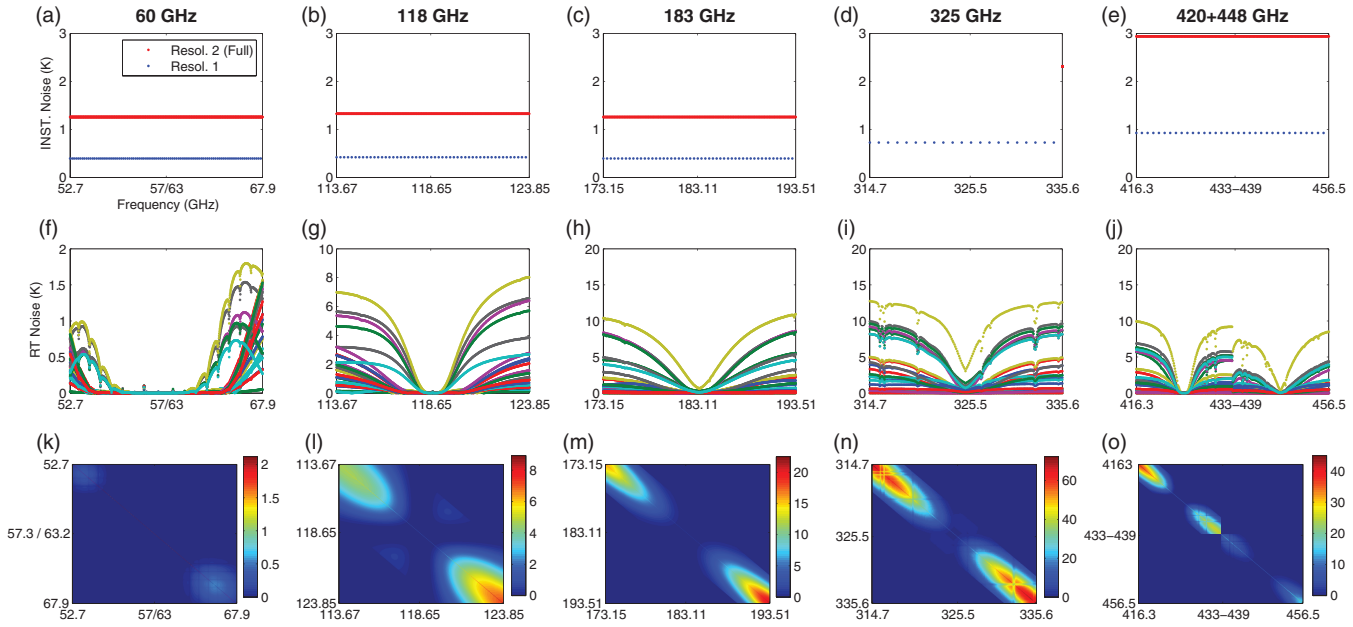

FIGURE 5 (a–e) Instrument noise for all the bands considered in this study. (f–j) The radiative transfer errors specified as 20% of the clear minus cloudy brightness temperature differences. (k–o) Covariance matrices for the observation errors (instrument + radiative transfer) [Colour figure can be viewed at wileyonlinelibrary.com]

Figure 5a–e represents these noise specifications for all the bands. There is no significant change expected in the instrumentation error between the clear and cloudy scenes; the TB range might be different, but we will assume here that the instrument behaviour is linear over the full dynamic of the possible signal.

3.2 | Radiative transfer noise, $\mathbf{R}_{\mathrm{RT}}$

It is extremely difficult to estimate the errors made by a RT code, especially under cloudy situations. Simulation versus observation errors include:

- Numerical errors in the RT calculations;
- Uncertainties in the inputs (i.e. geophysical variables) including those to be retrieved and those that are not;
- Errors in the hypothesis in the radiative properties of the parameters (especially the scatterers).

No RT model yet provides estimation on its uncertainties. In this study, we decided to link the RT errors to the RT cloud impact since this will be the dominant factor in the simulation errors. The larger the cloud impact, the larger the potential error. It was decided to estimate the RT errors for all channels as 20% of the (clear–cloudy) brightness temperature differences. Figure 5(f–j) represents these errors. The error increases with the scattering effect of the clouds, and with frequency, which is expected.

We decided to introduce correlations between the RT errors among the channels. In fact, errors for cloud assumptions for one channel are likely also to impact the neighbouring channels. Indeed, taking no correlation among channels close to each other would not be realistic. This would be artificially very beneficial to high spectral resolution: n multiple narrow channels close to each other would result in a RT error divided by $\sqrt{n}$ and this would be incorrect since each channel would have highly correlated systematic RT errors. As a consequence, instead of considering a diagonal covariance matrix $\mathbf{R}_f$, a non-diagonal matrix needs to be used to account for the correlation of errors among close channels. The correlation matrix is built using a Gaussian-shape truncated at two standard deviations to model the correlation dependence with frequency. By construction, these inter-channel noise correlations are identical for the various spectral resolutions considered in this study. For one particular channel, the range of correlation is 2 GHz on each side of the channel. The final RT error covariance matrix results from the standard deviations (Figure 5(f–j)) and this correlation matrix.

The total observation uncertainty is the sum of the instrumental and RT errors. These two errors are supposed to be independent, meaning that the observation errors can be calculated as the sum of the two covariance matrices: $\mathbf{R} = \mathbf{R}_{\mathrm{I}} + \mathbf{R}_{\mathrm{RT}}$. Figure 5k–o represents the observation-error covariance matrix for all the bands considered in these experiments, for one atmosphere.

3.3 | The background-error covariance matrices, B

Hydrometeors are generally not part of the control variable in NWP centres, so their corresponding background-error statistics are not available. Therefore, a methodology to obtain background-error covariance matrices such as in Hólm and Kral (2012) for temperature and specific humidity cannot be used. However, some attempts have been made to obtain the hydrometeor background-error covariance matrices $\mathbf{B}_{\mathrm{h}}$

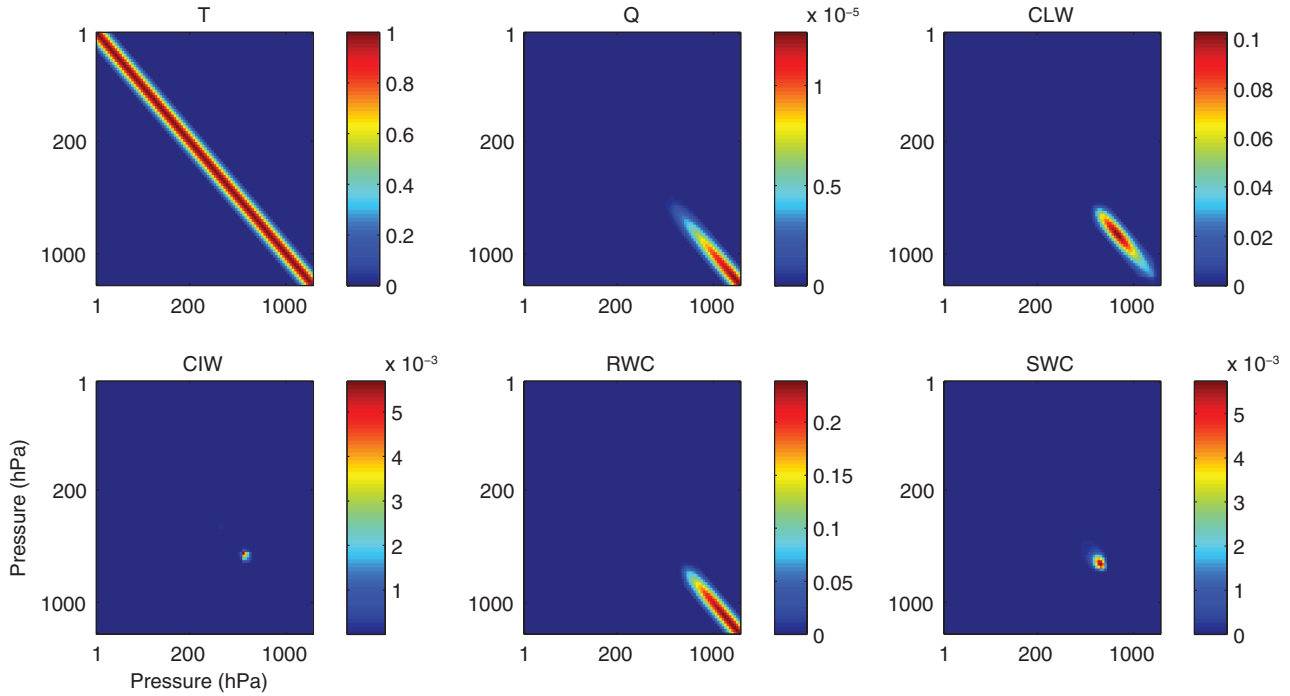

FIGURE 6 Background-error covariance matrix **B** for the six geophysical variables (a) T, (b) Q, (c) CLW, (d) CIW, (e) RWC and (f) IWC, for a particular atmospheric situation [Colour figure can be viewed at wileyonlinelibrary.com]

from the background-error covariance matrices $\mathbf{B}_{(T,Q)}$ for temperature T and specific humidity Q. Indeed, moist physical parametrization schemes for large-scale condensation and deep and shallow convections can be linearized (i.e. Jacobian **H**) to link in a simple way the two covariance matrices:

$$\mathbf{B}_{\mathrm{h}} = \mathbf{H} \cdot \mathbf{B}_{(T,Q)} \cdot \mathbf{H}^{\mathrm{T}}. \tag{4}$$

Such linearized physical processes have been developed at ECMWF by Lopez and Moreau (2005) for the moist convection and by Tompkins and Janisková (2004) for stratiform precipitation and cloud cover. The linearization **H** is used in the operational ECMWF four-dimensional variational assimilation system and allows for the assimilation of cloudy microwave radiances and surface precipitation rates. Some experiments have been performed in this study, but it was found that these background errors were quite irregular, noisy, and unstable, with rather surprising values. As a consequence, it was decided that the ECMWF background-error covariance matrices would not be used, and that a simpler yet reasonably realistic solution should be found instead.

The background errors of the atmospheric profiles are specified using covariance matrices. This is the standard practice and it is what is needed for the IC analysis. Covariance matrices result from two types of information: (a) the variance of error for each variable (this is the diagonal part of the covariance matrix), and (b) the correlation matrix among the background errors that describes the off-diagonal terms of the covariance matrix.

1. The standard deviation of the background errors are chosen to be state-dependent. For temperature T, 1 K is kept from the top to the bottom of the atmosphere. Humidity (Q) background uncertainty is 20% of the humidity profile. For CLW, CIW, RWC and SWC, the uncertainty is set to 40% of the actual real profile.
2. In the background covariance matrices **B**, we only take correlation among atmospheric levels for the same variable, not the correlation among the variables themselves. This would be more realistic but first we do not know how to model these relationships, and second we want assumptions simple enough so that we can understand the results of the forthcoming analysis. The background-error correlations for a variable are related to the proximity of the atmospheric layers. The correlation among two layers decreases with their vertical distance, following a Gaussian shape (truncated at 2 standard deviations). The spread of this correlation in the vertical is about ±10 atmospheric layers. This vertical spread of correlation is the same for the six atmospheric variables.

Figure 6 represents the resulting error covariance matrices for the six atmospheric variables, for one particular atmospheric situation. The structure of correlation in the vertical can be noted. It can also be seen that uncertainty for cloud hydrometer parameters are only considered for the vertical layers where they are present for this atmospheric situation.

3.4 | The combined versus independent information content

Equation 1 is often used independently for each variable to retrieve T, Q, CLW, CIW, RWC, and SWC. This implies that for each estimation, the uncertainty on the other variables is not taken into account, resulting in an underestimation of the

retrieval uncertainties. This is a simplification that is not necessary as it is quite easy technically to perform the uncertainty characterization simultaneously with each variable of interest.

In this study, we tested this effect and decided to combine the retrieval of all the variables at once. This is more realistic, however it provides higher errors than the independent estimations. The combination is easy, since it just consists of incorporating in a single Jacobian matrix **H** the Jacobians for the six variables, and similarly to incorporate the six background uncertainty matrices into a single block-diagonal matrix **B**. All the results in the following will use this "combined" approach.

3.5 | External factors and contamination effect

Our state vector $\mathbf{x} = (T, Q, \text{CLW}, \text{CIW}, \text{RWC}, \text{SWC})$ is quite complete, but if other geophysical variables called here "external factors" (surface or atmosphere) would impact the satellite observations, it would be interesting to include them in the analysis. For instance, surface temperature and emissivities will have an important effect on window channels. An uncertainty on them would introduce an additional uncertainty in the observations. In Bauer *et al.* (2005), the uncertainties related to the emissivity contamination are called the "geophysical" noise. Since we are concerned mostly here by the sounding channels, this contamination effect by the surface temperature and emissivity will be neglected.

It is important to keep in mind that, in general, all the variables impacting the observations should be considered in the IC analysis. In most IC analyses, the external variables are neglected and this can artificially overestimate the quality of the satellite observations. Collard (2007) attempted to take account of the effects of trace gases not included in the RT simulation by inflating the observation errors for channels that showed sensitivity to the missing species. A more complete approach was taken by Ventress and Dudhia (2013), who used climatological variability of atmospheric constituent species to model their effect on the radiances. They applied this technique not only to trace gases, but to humidity as well. In Pellet and Aires (2016), two ways are presented to consider the external factors. First, the external geophysical variables can just be added to the simultaneous IC analysis (including Jacobians and background errors) as in section 3.4. This is the cleaner way to consider the external factors. Another, simpler, way is to consider the impact of the external factors uncertainty on the satellite observations as an additional noise. Pellet and Aires (2016) show that these two approaches are equivalent if the variables to retrieve are independent of the external factors.

3.6 | Using multiple atmospheric situations

Hydrometeors have a larger diversity than variables such as temperature and humidity. Several orders of magnitude can be observed in nature, and a particular hydrometeor can have huge values or be completely absent (Figure 1 and section 2.2). This diversity in atmospheric situations translates into a large diversity in:

- Brightness temperature measurements (Figure 5);
- Jacobians (Figure 3 represents the averaged Jacobians); and
- Background errors (Figure 6 is for a particular situation).

The diversity of these three elements will have a strong impact on the retrieval uncertainty estimates of Equation 1. However, it is still a common practice to perform IC analyses on only one atmospheric situation. This is dangerous and potentially misleading for cloudy/precipitating scenes. In the experiments conducted in this article, the 25 atmospheric situations described in section 2.2 will be used, and the IC results will be averaged over these 25 carefully selected situations.

4 | APPLICATION TO THE HYMS INSTRUMENT

In this section, the goal is to measure the ability of a hyperspectral microwave instrument to retrieve geophysical profiles (T, Q, CLW, CIW, RWC, and SWC) under cloudy conditions. This will be performed using the IC and the hypothesis presented in the previous section.

4.1 | HYMS retrieval estimates

Figure 7 represents the retrieval improvements for T, Q, CLW, CIW, RWC and SWC with respect to the *a priori information*. Note that improvement for temperature would continue over the 100 hPa limit, but we were mostly interested here in tropospheric applications. The sensitivity of the different frequencies to a specific hydrometeor is modulated by the noise level that changes with the frequency. As described above, the observation noise is chosen proportional to the impact of the cloud on the TBs, and as a consequence, the noise tends to limit the effect of the most sensitive channels to a specific constituent. For instance, for the snow, the higher sensitivity of the 420 and 448 GHz bands to the snow is hampered by the increased noise, as compared to the 183 GHz band. Note nevertheless the good synergy between the different bands, with the lower frequencies providing more information on the lower atmospheric levels than the higher frequencies, as expected from the increasing attenuation with increasing frequency.

4.1.1 | Comparison with MetOp-SG

Figure 8 compares the retrieval improvements (with respect to the *a priori*) for the HYMS instrument and a MetOp-SG proxy (Ice Cloud Imager ICI, Microwave Imager MWI, and MWS), with two different levels of noise for the HYMS instrument. Results are averaged over the 25 atmospheres. First, the HYMS instrument with its initial noise level presents a significant improvement with respect to

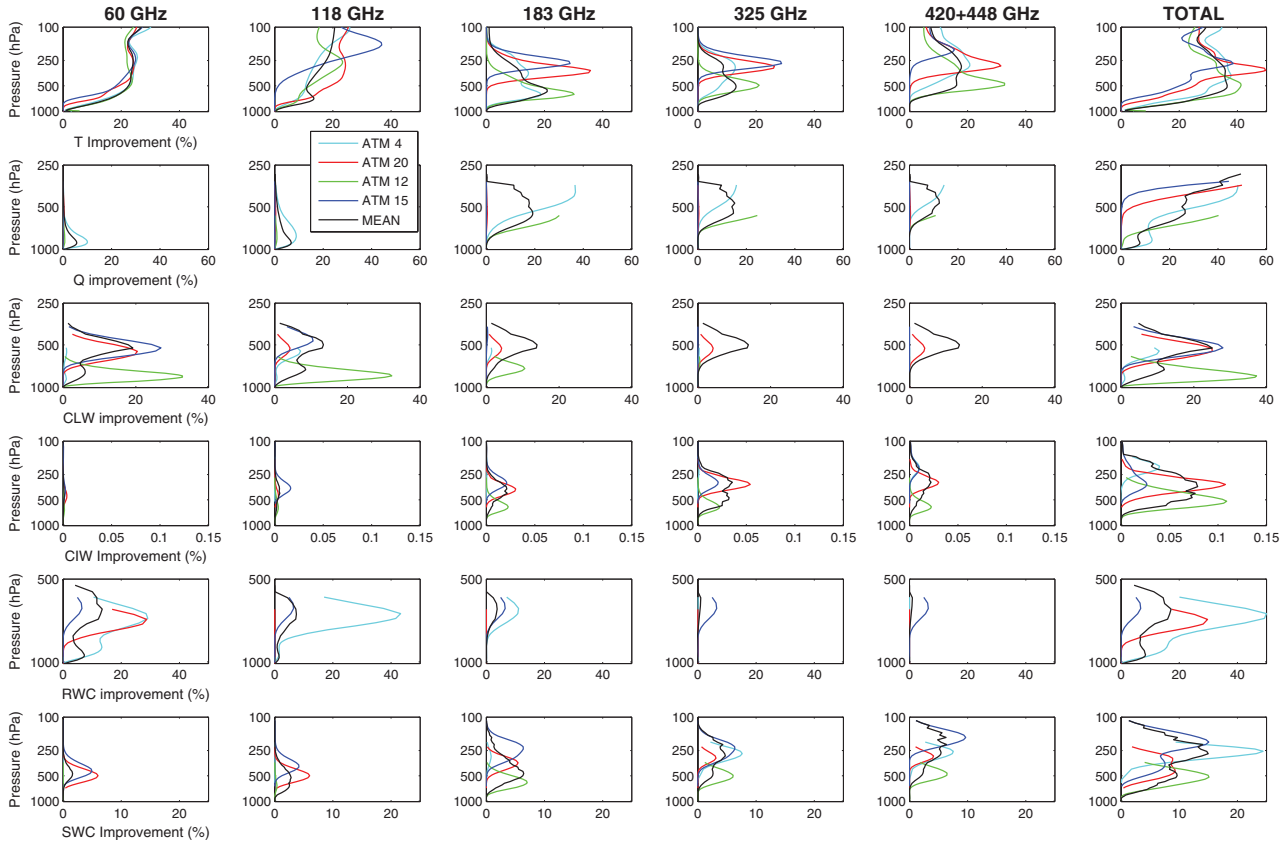

FIGURE 7 Retrieval improvements for (top to bottom) T, Q, CLW, CIW, RWC and SWC, for bands (left to right) at 60, 118, 183, 325, 420+448 GHz and TOTAL. Statistics are represented for the four selected situations and, in black, for the average of the 25 atmospheric situations [Colour figure can be viewed at wileyonlinelibrary.com]

MetOp-SG, especially for temperature. We extend the interest of the hyperspectral information to the cloudy case, as compared to the previous analysis under clear-sky conditions (Aires *et al.*, 2015). The improvement of temperature profiles in cloudy atmospheres is really a huge benefit; in fact, the World Meteorological Organization has such temperatures as a high priority for the improvement of weather forecasts. The improvement is also visible for the humidity retrieval. For the hydrometeor profiling, the conclusions are not so clear, with large differences from one atmosphere to the next, and different results from one hydrometeor species to the next. The fact that we cannot improve results in rain could be due to the presence of clouds or snow above the rain. Changing the noise levels of the HYMS or the MetOp-SG instruments does not change the conclusion much. Increasing HYMS noise does not degrade its performance significantly, and the HYMS with degraded performance still surpasses the MetOp-SG microwave suite. This means that the differences between MetOp-SG and HYMS is not just due to instrumental noise differences, it is also due to the multiplicity of channels, in particular for the T sounding above cloud.

4.2 | Sensitivity to very high spectral resolution

In Figure 9, the temperature retrieval improvements with respect to the *a priori* are compared for the two spectral resolutions 1 and 2 (Table 1), for the simple configuration, for bands 60 GHz and TOTAL. It can be seen that there is a significative difference between the lower (dashed) and the higher (continuous) spectral resolutions. In the clear-sky case, an improvement was observed for temperature at higher altitudes as well, and the higher spectral resolution could help retrieve the temperature above the cloud. For the five other variables (Q, CLW, CIW, RWC and SWC) no significant impact was observed (not shown).

5 | CONCLUSION AND PERSPECTIVES

5.1 | Conclusion

In this paper, we presented tools and methods to improve the realism of the hypotheses used in the Information Content (IC) analysis of microwave observations for cloudy/precipitating scenes. In particular, we introduced:

- Observation noise including a well-defined instrument noise, and RT error characteristics that are state-dependent and realistic.
- Background covariance matrices that are state-dependent, take into account vertical correlations, with a cloud layer localization uncertainty.
- The IC results come from statistics performed on a diverse set of atmospheric situations which have been carefully

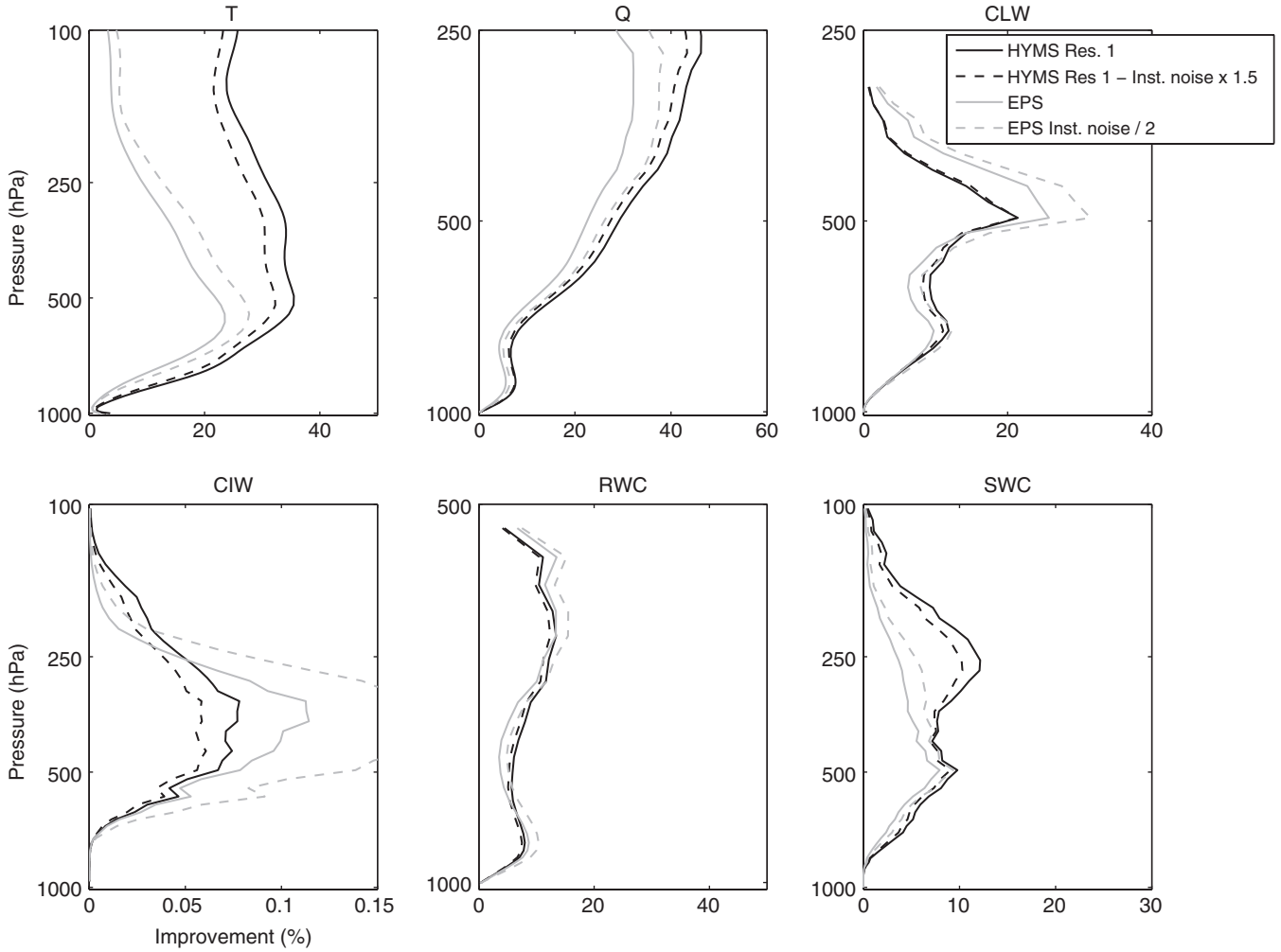

FIGURE 8 Improvement of the performances compared to the *a priori*, for the retrieval of (a) T, (b) Q, (c) CLW, (d) CIW, (e) RWC, and (f) SWC. Results are presented for HYMS with the initial noise, for HYMS with a 50% noise increase, and for the MetOp-SG microwave suite (MWS, MWI, and ICI) with regular or divided-by-two noise levels

simulated using an up-to-date RT model with realistic hydrometeor assumptions.

- The IC is performed simultaneously on all the important variables (T, Q, CLW, CIW, RWC and SWC) to be more reliable since it takes into account all sources of uncertainty.

We obtained realistic retrieval estimates for the HYMS instrument. The potential improvements of such a hyperspectral instrument has already been shown under clear-sky conditions (Aires *et al.*, 2015). Here, we also showed that such an instrument can be valuable for both temperature and water vapour profiling, under cloudy-sky conditions, when compared to the MetOp-SG microwave suite. In particular, the improvement in terms of temperature profile is significant even under cloudy conditions, mainly above the cloud. We analysed the sensitivity of the retrievals to the spectral resolution of the channels. The use of very high resolution (a factor of 10 with respect to the initial high spectral resolution) does not improve the results much. This can be explained by the fact that the Jacobians are by nature rather wide vertically, even for monochromatic radiation, which is a consequence of the radiative transfer equation. The high spectral resolution does not improve the hydrometeor retrievals but efforts still need to be made to obtain reliable hydrometeor properties.

5.2 | Perspectives

The tools proposed in this study can and need to be improved further by the scientific community since much work remains to be done in a coordinated manner. Three main issues are still a limiting factor in what can be done so far with IC for cloudy/precipitating scenes.

- *Databases.* As we showed here, the IC results are really dependent on the considered atmospheric situations. In order to obtain a realistic picture of the observing system, it is important to use a large and diverse set of atmospheric situations. Such databases need to be built using sophisticated sampling techniques in high-dimensional space. This database needs to be simulated using realistic hydrometeor assumptions, which need to be described in a statistical manner, using distributions and physically coherent correlation structures (e.g. from cloud-resolving models). Of

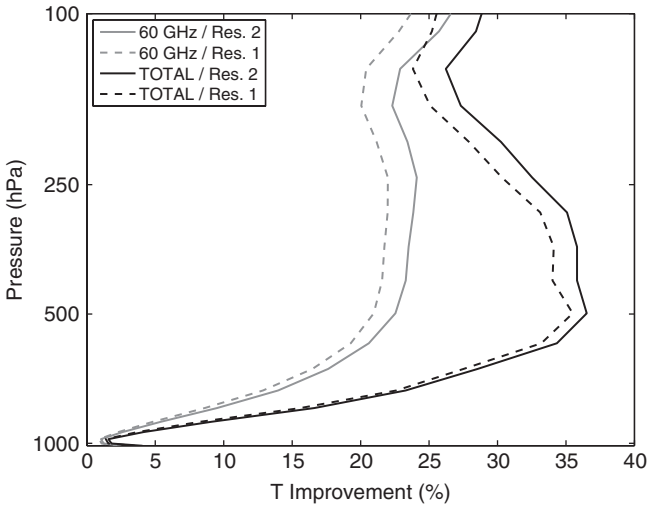

FIGURE 9 Temperature retrieval improvement (with respect to the *a priori* estimations) for the spectral resolutions 1 (dashed line) and 2 (continuous line) (Table 1), when using only the 60 GHz band (grey), and when using all the bands (black)

course, these hydrometeor assumptions need to be compared to *in situ* measurements to be confirmed by reality. The fact that each group uses their own assumptions does not facilitate inter-comparison exercises. On the other hand, it reflects the large uncertainty in these assumptions.

- *RT errors.* A big effort needs to be undertaken by the radiative transfer community. Characterizing the RT uncertainties on cloudy/precipitating scenes is needed to advance our understanding of the physics, to build performant retrieval algorithms, and to improve the assimilation of these observations in NWP centres. The RT model developers should focus more on the estimation of their model uncertainties (structural, due to uncertain parameters, or related to the geophysical noise). This can be done by comparing simulations to observations, but this is not sufficient. Error propagation tools could also help perform this task.
- *Background errors.* Better background matrices should ideally consider all the ambiguities of the physical inverse problem, e.g. localization of the layers, type of hydrometeors. We proposed a state-dependent background-error scheme, but no correlations among the state variables (T, Q, CLW, CIW, RWC and SWC) were considered, which is not satisfactory. It took the NWP community several years to improve the temperature and humidity background errors, and these efforts now need to be continued for the hydrometeor errors.

Statistical retrievals are relatively easy to implement, and they can offer good estimates of the retrieval accuracies (Defer *et al.*, 2008; Wang *et al.*, 2017) as an alternative to IC. Both approaches are legitimate and have their pros and cons. Fewer assumptions are required for statistical retrieval tests, so there is no need to linearize the radiative transfer, and the sources of uncertainty do not need to follow Gaussian distributions. However, the instrumental and RT errors still need to be provided since the retrieval is based on the signal-to-noise ratio. Correlations (among the observations and among the geophysical parameters) are used by the statistical retrieval, but only implicitly (described in the database), not explicitly as in the IC technique. The reliability of these statistical retrievals strongly depends on the quality and diversity of the atmospheric situations in the database. As for the geophysical noise in the RT uncertainties, each geophysical variable potentially impacting the satellite observations (e.g. surface temperature or emissivity) needs to be represented and sampled in the database so that they are well considered/exploited by the statistical retrieval. This requires sophisticated sampling techniques in high-dimensional space (Aires and Prigent, 2007; Paul and Aires, 2014).

ACKNOWLEDGEMENTS

The authors are grateful to Sabatino Di Michele (ECMWF) for providing to the project the atmospheric profiles described in this report. This study was supported by an ESA contract N 4000105721/12/NL/AF entitled "Use of spectral information at microwave region for numerical weather prediction". Authors would like to thank also Jean-Francois Mahfouf, Steve English, Bill Bell and Sid Boukabara for interesting and fruitful discussions. We also would like to thank those who helped the RT simulations, in particular Jana Mendrok from the University of Hamburg in Germany. The dataset used in this study is available upon request to filipe.aires@obspm.fr.

REFERENCES

Aires, F. and Prigent, C. (2007) Sampling techniques in high-dimensional spaces for the development of satellite remote-sensing database. *Journal of Geophysical Research: Atmospheres*, 112(D20). https://doi.org/1029/2007JD008391.

Aires, F., ChéDin, A., Scott, N.A. and Rossow, W.B. (2002) A regularized neural net approach for retrieval of atmospheric and surface temperatures with the IASI instrument. *Journal of Applied Meteorology*, 41(2), 144–159. https://doi.org/10.1175/1520-0450(2002)041<0144:arnnaf>2.0.CO;2.

Aires, F., Prigent, C., Mahfouf, J.-F., Orlandi, E. and Milz, M. (2014) Work-Package 2 - Clear sky analysis. Technical Report, Number 2.0. ESTEC, ESA, Noordwijk, Netherlands

Aires, F., Prigent, C., Orlandi, E., Milz, M., Eriksson, P., Crewell, S., Lin, C.-C. and Kangas, V. (2015) Microwave hyperspectral measurements for temperature and humidity atmospheric profiling from satellite: the clear-sky case. *Journal of Geophysical Research: Atmospheres*, 120(21), 10. https://doi.org/1002/2015JD023331.

Aires, F., Pellet, V., Prigent, C. and Moncet, J.L. (2016) Dimension reduction of satellite observations for remote sensing. Part 1: a comparison of compression, channel selection and bottleneck channel approaches. *Quarterly Journal of the Royal Meteorological Society*, 142, 2658–2669. https://doi.org/10.1002/qj.2855.

Bauer, P. and Di Michele, S. (2007) Mission assumptions and technical requirements for post-EPS. Report EOP-SFP/2006-07-1223. ESTEC, ESA, Noordwijk, Netherlands [AUTHOR: Is this the intended document? If not, give FULL details.].

Bauer, P., Moreau, E. and Di Michele, S. (2005) Hydrometeor retrieval accuracy using microwave window and sounding channel observations. *Journal of Applied Meteorology*, 44(7), 1016–1032. https://doi.org/10.1175/JAM2257.1.

Bauer, P., Lopez, P., Salmond, D., Benedetti, A., Saarinen, S. and Bonazzola, M. (2006) Implementation of 1D+4D-Var assimilation of precipitation-affected

microwave radiances at ECMWF. II: 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 132, 2307–2332. https://doi.org/10.1256/qj.06.07.

Birman, C., Mahfouf, J.-F., Milz, M., Mendrok, J., Buehler, S.A. and Brath, M. (2017) Information content on hydrometeors from millimeter and sub-millimeter wavelengths. *Tellus A*, 69(1), 1–24. https://doi.org/10.1080/16000870.2016.1271562.

Blackwell, W.J., Bickmeier, L.J., Leslie, R.V., Pieper, M.L., Samra, J.E., Surussavadee, C. and Upham, C.A. (2011) Hyperspectral microwave atmospheric sounding. *IEEE Transactions on Geoscience and Remote Sensing*, 49(1), 128–142. https://doi.org/10.1109/TGRS.2010.2052260.

Blackwell, W.J., Galbraith, C., Hancock, T., Leslie, R., Osaretin, I., Shields, M., Racette, P. and Hilliard, L. (2012) Design and analysis of a hyperspectral microwave receiver subsystem, In *IEEE 12th Specialist Meeting on Microwave Radiometry and Remote Sensing of the Environment (MicroRad)*, Rome. https://doi.org/10.1109/MicroRad.2012.6185268.

Bormann, N., Collard, A.D. and Bauer, P. (2017) Observation errors and their correlations for satellite radiances. *ECMWF Newsletter*, 128, 17–22.

Boukabara, S.A. and Garrett, K. (2011) Benefits of a hyperspectral microwave sensor. In: *IEEE Sensors conference*, pp. 1881–1884, Limerick, Ireland. https://doi.org/10.1109/ICSENS.2011.6127357.

Chevallier, F., Di Michele, S. and McNally, A.P. (2006) Diverse profile datasets from the ECMWF 91-level short-range forecasts. NWP SAF Technical report 10. ECMWF, Reading, UK.

Collard, A.D. (2007) Selection of IASI channels for use in numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 133, 1977–1991. https://doi.org/10.1002/qj.178.

Defer, E., Prigent, C., Aires, F., Pardo, J.R., Walden, C.J., Zanifé, O.Z., Chaboureau, J.-P. and Pinty, J.-P. (2008) Development of precipitation retrievals at millimeter and sub-millimeter wavelengths for geostationary satellites. *Journal of Geophysical Research: Atmospheres*, 113(D8). https://doi.org/10.1029/2007JD008673.

Di Michele, S. and Bauer, P. (2006) Passive microwave radiometer channel selection based on cloud and precipitation information content. *Quarterly Journal of the Royal Meteorological Society*, 132, 1299–1323. https://doi.org/10.1256/qj.05.164.

Emde, C., Buehler, S.A., Davis, C., Eriksson, P., Sreerekha, T.R. and Teichmann, C. (2004) A polarized discrete ordinate scattering model for simulations of limb and nadir long-wave measurements in 1D/3D spherical atmospheres. *Journal of Geophysical Research: Atmospheres*, 109(D24). https://doi.org/1029/2004JD005140.

Eriksson, P., Buehler, S.A., Davis, C.P., Emde, C. and Lemke, O. (2011) ARTS, the atmospheric radiative transfer simulator, Version 2. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 112(10), 1551–1558. https://doi.org/10.1016/j.jqsrt.2011.03.001.

Eriksson, P., Jamali, M., Mendrok, J. and Buehler, S.A. (2015) On the microwave optical properties of randomly oriented ice hydrometeors. *Atmospheric Measurement Techniques*, 8(5), 1913–1933. https://doi.org/10.5194/amt-8-1913-2015.

Field, P., Hogan, R., Brown, P., Illingworth, A., Choularton, T. and Cotton, R. (2005) Parametrization of ice-particle size distributions for mid-latitude stratiform cloud. *Quarterly Journal of the Royal Meteorological Society*, 131, 1997–2017.

Galligani, V.S., Prigent, C., Defer, E., Jimenez, C., Eriksson, P., Pinty, J.-P. and Chaboureau, J.-P. (2015) Meso-scale modelling and radiative transfer simulations of a snowfall event over France at microwaves for passive and active modes and evaluation with satellite observations. *Atmospheric Measurement Techniques*, 8(3), 1605–1616. https://doi.org/10.5194/amt-8-1605-2015.

Geer, A.J. and Baordo, F. (2014) Improved scattering radiative transfer for frozen hydrometeors at microwave frequencies. *Atmospheric Measurement Techniques*, 7(6), 1839–1860. https://doi.org/10.5194/amt-7-1839-2014.

Hersman, M.S. and Poe, G.A. (1981) Sensitivity of the total power radiometer with periodic absolute calibration. *IEEE Transactions on Microwave Theory and Techniques*, 29(1), 32–40. https://doi.org/10.1109/TMTT.1981.1130283.

Hólm, E.V. and Kral, T. (2012) Flow-dependent, geographically varying background-error covariances for 1D-Var applications in MTG-IRS L2 processing. Technical Memo. 680, ECMWF, Reading, UK.

Hong, S.Y., Dudhia, J. and Chen, S.-H. (2004) A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Monthly Weather Review*, 132, 103–120.

Kangas, V., D'Addio, S., Betto, M., Barre, H. and Mason, G. (2012). MetOp Second Generation Microwave Radiometers. *12th Specialist meeting on microwave radiometry and remote sensing of the Environment (MicroRad)*, pp. 1–4.

Lipton, A.E. (2003) Satellite sounding channel optimization in the microwave spectrum. *IEEE Transactions on Geoscience and Remote Sensing*, 41(4), 761–781. https://doi.org/10.1109/TGRS.2003.810926.

Lipton, A.E., Moncet, J., Boukabara, S.A., Uymin, G. and Quinn, K.J. (2009) Fast and accurate radiative transfer in the microwave with optimum spectral sampling. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7), 1909–1917. https://doi.org/10.1109/TGRS.2008.2010933.

Liu, G. (2008) A database of microwave single-scattering properties for nonspherical ice particles. *Bulletin of the American Meteorological Society*, 89, 1563–1570. https://doi.org/10.1175/2008BAMS2486.1.

Lopez, P. and Moreau, E. (2005) A convection scheme for data assimilation: description and initial tests. *Journal of the Royal Meteorological Society*, 131, 409–436.

Mahfouf, J.-F., Birman, C., Aires, F., Prigent, C., Orlandi, E. and Milz, M. (2014) Information content on temperature and water vapour from a hyper-spectral microwave sensor. *Quarterly Journal of the Royal Meteorological Society*, 141, 3268–3284.

Mätzler, C. (2002) MATLAB functions for Mie scattering and absorption, version 2. Research Report 2002-11. IAP, University of Bern, Switzerland

Mätzler, C.(ed.). (2006) *Thermal Microwave Radiation: Applications for Remote Sensing*. IET, Stevenage, UK.

McFarquhar, G.M. and Heymsfield, A.J. (2010) Parameterization of tropical cirrus ice crystal size distributions and implications for radiative transfer: results from CEPEX. *Journal of the Atmospheric Sciences*, 54, 2187–2200. https://doi.org/10.1175/1520-0469(1997)054<2187:potcic>2.0.CO;2.

Paul, M. and Aires, F. (2014) Using Shannon's entropy to sample heterogene and high-dimensional atmospheric datasets. *Quarterly Journal of the Royal Meteorological Society*, 141, 469–476. https://doi.org/10.1002/qj.2367.

Pellet, V. and Aires, F. (2016) Dimension reduction of satellite observations for remote sensing, Part 2: Illustration using hyperspectral microwave observations. *Quarterly Journal of the Royal Meteorological Society*, 142, 2670–2678.

Rodgers, C.D. (1976) Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Reviews of Geophysics and Space Physics*, 14(4), 609–624. https://doi.org/1029/RG014i004p00609.

Rodgers, C.D. (1990) Characterization and error analysis of profiles retrieved from remote sounding measurements. *Journal of Geophysical Research: Atmospheres*, 95(D5), 5587–5595. https://doi.org/10.1029/JD095iD05p05587.

Rodgers, C.D. (2000) *Inverse Methods for Atmospheric Sounding – Theory and Practice*. World Scientific, Singapore. https://doi.org/10.1142/3171.

Tompkins, A.M. and Janisková, M. (2004) A cloud scheme for data assimilation: description and initial tests. *Quarterly Journal of the Royal Meteorological Society*, 130, 2495–2517. https://doi.org/10.1256/qj.03.162.

Ventress, L. and Dudhia, A. (2013) Improving the selection of IASI channels for use in numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 140, 2111–2118. https://doi.org/10.1002/qj.2280.

Wang, Z., Yang, S., Qiao, Y., Cui, S. and Zhao, Q. (2012) Monte Carlo simulations of radiative transfer in cloudy atmosphere over sea surfaces. *Terrestrial, Atmospheric and Oceanic Sciences*, 23(1), 59–70. https://doi.org/10.3319/TAO.2011.08.29.01(A).

Wang, D., Prigent, C., Aires, F. and Jiménez, C. (2017) A statistical retrieval of cloud parameters for the millimeter wave ice cloud imager on board MetOp-SG. *IEEE Access*, 5, 4057–4076. https://doi.org/10.1109/ACCESS.2016.2625742.

How to cite this article: Aires F, Prigent C, Buehler SA, Eriksson P, Milz M, Crewell S. Towards more realistic hypotheses for the information content analysis of cloudy/precipitating situations – Application to a hyperspectral instrument in the microwave. *Q J R Meteorol Soc.* 2018;1–14. https://doi.org/10.1002/qj.3315