Assimilating synthetic GOES-R radiances in cloudy conditions using an ensemble-based method

Article *in* International Journal of Remote Sensing · December 2011
DOI: 10.1080/01431161.2011.572094

8 authors, including:

Dusanka Zupanski
Spire Global
63 PUBLICATIONS **1,482** CITATIONS
SEE PROFILE

Milija Zupanski
Colorado State University
100 PUBLICATIONS **1,458** CITATIONS
SEE PROFILE

Renate Brummer
Colorado State University
12 PUBLICATIONS **133** CITATIONS
SEE PROFILE

Isidora Jankov
National Oceanic and Atmospheric Administration
29 PUBLICATIONS **409** CITATIONS
SEE PROFILE

International Journal of Remote Sensing

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

Assimilating synthetic GOES-R radiances in cloudy conditions using an ensemble-based method

Dusanka Zupanski [a], Milija Zupanski [a], Lewis D. Grasso [a], Renate Brummer [a], Isidora Jankov [a], Daniel Lindsey [b], Manajit Sengupta [a] & Mark Demaria [b]

[a] Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA

[b] NOAA/NESDIS/StAR, Fort Collins, CO, USA

Available online: 02 Aug 2011

To cite this article: Dusanka Zupanski, Milija Zupanski, Lewis D. Grasso, Renate Brummer, Isidora Jankov, Daniel Lindsey, Manajit Sengupta & Mark Demaria (2011): Assimilating synthetic GOES-R radiances in cloudy conditions using an ensemble-based method, International Journal of Remote Sensing, DOI:10.1080/01431161.2011.572094

To link to this article: http://dx.doi.org/10.1080/01431161.2011.572094

International Journal of Remote Sensing
iFirst, 2011, 1–23

Taylor & Francis
Taylor & Francis Group

Assimilating synthetic GOES-R radiances in cloudy conditions using an ensemble-based method

DUSANKA ZUPANSKI*†, MILIJA ZUPANSKI†, LEWIS D. GRASSO†, RENATE BRUMMER†, ISIDORA JANKOV†, DANIEL LINDSEY‡, MANAJIT SENGUPTA† and MARK DEMARIA‡

†Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA

‡NOAA/NESDIS/StAR, Fort Collins, CO, USA

(*Received 10 February 2009; in final form 16 February 2011*)

The weather research and forecasting (WRF) model and the maximum likelihood ensemble filter (MLEF) data assimilation approach are used to examine the potential impact of observations from the future Geostationary Operational Environmental Satellite, generation R (GOES-R) on improving our knowledge about clouds. Synthetic radiances are assimilated from the 10.35 μm channel of the GOES-R advanced baseline imager (ABI) employing a 'non-identical twins' experimental setup. The experimental results are examined for an extratropical cyclone named Kyrill that produced unusually strong winds, widespread damage and fatalities in Western Europe in January 2007. The data assimilation problem is especially challenging for this case, as there is a large error in the model-simulated radiances resulting from incorrect cloud location. Although this problem is difficult to eliminate, data assimilation results indicate the potential of GOES-R data to significantly reduce these errors.

1. Introduction

Current and new satellite missions provide a wealth of information to improve our understanding of weather, climate, ocean and the Earth system in general. This study is motivated by the needs of new satellite missions to undergo all necessary preparations well before the satellite launch in order to make sure that the observations will be successfully used as soon as they become available. Data assimilation studies, such as this one, are being employed to prepare methodologies capable of addressing the challenges of the new satellite missions. These methodologies should be able to effectively assimilate the new satellite observations and quantify the information content of the assimilated data. This study focuses on the data assimilation needs of the next-generation series of Geostationary Operational Environmental Satellite, R (GOES-R) mission, currently scheduled for launch in the year 2015. Our objective is to explore the information content and further improve capabilities of the current state-of-the-art data assimilation methods in order to extract maximum information from the GOES-R data, especially in cloudy scenes.

*Corresponding author. Email: dzupanski@precisionwind.com

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2011 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/01431161.2011.572094

Data assimilation methods have been successfully used to assimilate various satellite observations in order to improve weather, climate, ocean, hydrological and ecological forecasts over several decades (L'Ecuyer *et al*. (2006), Reichle (2007), Carton and Giese (2008), Hollingsworth *et al*. (2008), Jung *et al*. (2008), Keppenne *et al*. (2008) and Migliorini *et al*. (2008) are some recent examples). It is commonly accepted that satellite and other remote sensing observations are a major source of information for today's geophysical models due to wide spatial and high temporal coverage as compared to the non-remote sensing (the so-called conventional) observations. Current state-of-the-art data assimilation methods typically assimilate satellite radiances or brightness temperatures, rather than derived model state parameters (retrievals). This is mainly because the observation errors of the radiances/brightness temperatures are better known (and are typically less biased) than the observation errors of the retrievals.

One of the most difficult challenges of satellite data assimilation is assimilation of satellite radiances in cloudy conditions. Some of the major difficulties arise from the non-linear and often discontinuous character of modelled cloud-microphysical processes and from largely unknown forecast error covariances of these processes. Because of these and many other difficulties, the operational weather centres have been assimilating clear sky (or cloud cleared) radiances for decades, thus discarding important information about clouds and precipitation, as well as other atmospheric variables, contained in the cloudy visible, infrared and microwave radiances (more about the importance of assimilation of cloudy radiances can be found, e.g. in Andersson *et al*. (2005) and Errico *et al*. (2007)). Nevertheless, cloudy satellite retrievals can also bring important information about precipitation and clouds, as demonstrated in the recent studies by Hou and Zhang (2007) and Lin *et al*. (2007), where the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) and Special Sensor Microwave Imager (SSM/I) tropical rainfall observations were assimilated.

Thanks to advancements in numerical modelling and data assimilation methods, assimilation of cloudy radiances has advanced significantly in the last decade, especially in the atmospheric data assimilation applications. In a study by Chevallier *et al*. (2004) cloud-affected satellite infrared radiances were successfully assimilated using a variational data assimilation method. They pointed out that for some cloud-affected channels of the METEOrological SATellite (METEOSAT) and atmospheric infrared sounder (AIRS) instruments, the linearity assumption might still be valid, thus these channels were easier to assimilate. In Vukicevic *et al*. (2004, 2006) assimilation of the GOES imager brightness temperatures into a cloud-resolving model was successfully performed, indicating clear benefits in improved cloud analyses and short-term forecasts. It is important to note that the four-dimensional variational (4D-Var) data assimilation approach used in Vukicevic *et al*. (2004, 2006) involved an iterative minimization and non-linear updates of the cloud state variables in fine spatial and temporal resolution, which were helpful in alleviating some of the difficulties due to non-linearities of the cloud-microphysical processes. In Bauer *et al*. (2006a,b) the approach called 1D-Var + 4D-Var was introduced for assimilation of precipitation-affected microwave radiances, which was also adopted for operational application at the European Centre for Medium-Range Weather Forecasts (ECMWF). The two-step approach, where satellite radiances are assimilated by the non-linear 1D-Var step to produce increments of total column water vapour, and then these increments are assimilated

by the linear (so-called incremental) 4D-Var step, has proven better in handling non-linearities than the incremental 4D-Var approach alone. In Weng *et al*. (2007) rain-affected satellite microwave radiances from the Advanced Microwave Sounding Unit (AMSU) and the Advanced Microwave Scanning Radiometer (AMSR-E) are assimilated to improve hurricane vortex analysis. They used an approach called hybrid variational (HVAR) scheme, which is similar to the ECMWF 1D-Var + 4D-Var approach; however, they employed a different model, the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5) and its adjoint (Zou *et al*. 1998, Zou and Xiao 2000). The data assimilation results indicated improved, more detailed, structures for the hurricane warm core at the upper troposphere and enhanced lower-level wind speed and upper-level divergence, thus highlighting the importance of assimilation of cloudy satellite radiances.

Novel, ensemble-based, data assimilation methods hold a potential to overcome some of the difficulties of cloud and precipitation assimilation, especially because of the use of flow-dependent forecast error covariances and an improved treatment of non-linearities (due to not using tangent linear and adjoint models). However, applications of the ensemble-based data assimilation methods to remote sensing observations in general are still rare, and the experience with assimilation of cloudy satellite radiances is, to our knowledge, non-existent. Nevertheless, the potential of these methods to further improve the analysis and forecast of clouds and precipitation is evident from the currently available studies (e.g. Liu *et al*. 2008, Meng and Zhang 2008, Whitaker *et al*. 2008, Aksoy *et al*. 2009), which were performed using conventional and/or some remote sensing observations. Therefore, further exploring the ensemble data assimilation methods in cloud and precipitation assimilation should be well worth the effort.

In this study, we report the results of a pilot study, performed to evaluate the potential of the ensemble data assimilation methods to extract maximum information from the future GOES-R radiance observations in cloudy scenes. Through the use of information measures based on the flow-dependent forecast error covariance matrix, we define when and where the observed information is needed the most. This is in the areas where the flow-dependent forecast uncertainty is the largest. We focus on the impact of the observations from the advanced baseline imager (ABI), an instrument that will have significant improvements upon the current GOES imager with more spectral bands, higher spatial and temporal resolution, better navigation and more accurate calibration (Schmit *et al*. 2005). It will also have improved temporal and spatial resolution relative to those of the GOES-13/O/P sounders (Schmit *et al*. 2008). Before using observed radiances, synthetic ABI radiances at 10.35 μm were assimilated as a first step in this study. The 10.35 μm channel was selected because it is a clean window channel, expected to be sensitive to the hydrometeors at the cloud top, such as cloud ice and snow (Smith *et al*. 1992, Grasso and Greenwald 2004, Grasso *et al*. 2008). Therefore, this channel is capable of providing information about cloud-microphysical processes at the cloud tops.

2. Data assimilation approach

2.1 *Basic equations*

We employ a variant of the maximum likelihood ensemble filter (MLEF) data assimilation approach, developed at Colorado State University (Zupanski 2005, Zupanski

and Zupanski 2006, Zupanski *et al*. 2008). The MLEF seeks a maximum likelihood solution of the posterior probability density function (PDF), which is equivalent to seeking a minimum of the following cost function (under the standard assumption of Gaussian PDFs for the observation and forecast errors):

$$J(\boldsymbol{x}) = \frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{f}}\right)^{\mathrm{T}}\mathbf{P}_{\mathrm{f}}^{-1}\left(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{f}}\right)+\frac{1}{2}\left[\boldsymbol{y}-H(\boldsymbol{x})\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\boldsymbol{y}-H(\boldsymbol{x})\right]. \tag{1}$$

The cost function (equation (1)) measures the differences between the model and the observations, where vector $\boldsymbol{y}$ of dimension N_{obs} (number of observations) is the observation vector, vector $\boldsymbol{x}$ of dimension N_{S} (model state dimension) is the model state vector, non-linear operator H is an observation operator, matrix $\mathbf{R}$ is the observation error covariance and matrix $\mathbf{P}_{\mathrm{f}}$ is the forecast error covariance. The index 'f' refers to the forecast (used as a first guess). Superscript 'T' denotes transpose. The matrix $\mathbf{P}_{\mathrm{f}}$ is defined in a subspace spanned by ensemble forecast perturbations as

$$\mathbf{P}_{\mathrm{f}}^{1/2} = \left(\boldsymbol{p}_1^{\mathrm{f}} \quad \boldsymbol{p}_2^{\mathrm{f}} \quad \ldots \quad \boldsymbol{p}_{\mathrm{N_E}}^{\mathrm{f}}\right), \qquad \boldsymbol{p}_i^{\mathrm{f}} = M(\boldsymbol{x}^{\mathrm{a}}+\boldsymbol{p}_i^{\mathrm{a}}) - M(\boldsymbol{x}^{\mathrm{a}}), \tag{2}$$

where M denotes the non-linear forecast model, the superscript 'a' refers to the analysis and N_{E} is the number of ensembles. The vectors $\mathbf{p}_i^{\mathrm{a}}$ and $\mathbf{p}_i^{\mathrm{f}}$ represent columns of the square roots of the analysis and forecast error covariances, respectively. The square root of the analysis error covariance is defined at the analysis solution $\boldsymbol{x}^{\mathrm{a}}$ as

$$\mathbf{P}_{\mathrm{a}}^{1/2} = \mathbf{P}_{\mathrm{f}}^{1/2}\left[\mathbf{I}_{\mathrm{N_E}} + \left(\mathbf{Z}(\boldsymbol{x}^{\mathrm{a}})\right)^{\mathrm{T}}\mathbf{Z}(\boldsymbol{x}^{\mathrm{a}})\right]^{-1/2}, \tag{3}$$

where $\mathbf{I}_{\mathrm{N_E}}$ is an $N_{\mathrm{E}} \times N_{\mathrm{E}}$ identity matrix and the matrix $\mathbf{Z}(\boldsymbol{x}^{\mathrm{a}})$ is the observation perturbation matrix at the analysis solution, defined by the following equation:

$$\begin{aligned}\mathbf{Z}(\boldsymbol{x}^{\mathrm{a}}) &= \left[\boldsymbol{z}_1(\boldsymbol{x}^{\mathrm{a}}) \quad \boldsymbol{z}_2(\boldsymbol{x}^{\mathrm{a}}) \quad \ldots \quad \boldsymbol{z}_{\mathrm{N_E}}(\boldsymbol{x}^{\mathrm{a}})\right],\\ \boldsymbol{z}_i(\boldsymbol{x}^{\mathrm{a}}) &= \mathbf{R}^{-1/2}\left[H(\boldsymbol{x}^{\mathrm{a}}+p_i^{\mathrm{f}}) - H(\boldsymbol{x}^{\mathrm{a}})\right].\end{aligned} \tag{4}$$

The inverse square root calculation in equation (3) is obtained via eigenvalue decomposition of the matrix $\mathbf{I}_{\mathrm{N_E}} + \mathbf{Z}(\boldsymbol{x}^{\mathrm{a}})^{\mathrm{T}}\mathbf{Z}(\boldsymbol{x}^{\mathrm{a}})$. It is calculated as a symmetric square root, which is unique (e.g. Wang *et al*. 2004, Zupanski 2005, Wei *et al*. 2006).

We also calculate, as a diagnostic, the so-called information matrix $\mathbf{C}$, of dimensions $N_{\mathrm{E}} \times N_{\mathrm{E}}$, defined in ensemble subspace as

$$\mathbf{C} = \mathbf{Z}(\boldsymbol{x}^{\mathrm{a}})^{\mathrm{T}}\mathbf{Z}(\boldsymbol{x}^{\mathrm{a}}), \tag{5}$$

which we use to calculate information measures, such as degrees of freedom for signal (DFS) defined as d (e.g. Shannon and Weaver 1949, Rodgers 2000, Zupanski *et al*. 2007)

$$d = \sum_{i=1}^{\mathrm{N_E}} \frac{\lambda_i^2}{(1+\lambda_i^2)}, \tag{6}$$

where index 'i' denotes an ensemble member and λ_i^2 are eigenvalues of the information matrix $\mathbf{C}$.

As an additional diagnostic, we also calculate analysis uncertainty in observation locations and in terms of the observed variable (10.35 μm radiance in the experiments presented) to be able to compare it with the actual analysis uncertainty, defined by the differences between the analysis and the 'truth' (the 'truth' being defined by a different forecast model in our 'non-identical twins' experimental setup). The analysis uncertainty, defined as the standard deviation of the analysis errors ($\sigma^{\mathrm{a}}_{\mathrm{obs}}$), is calculated using the following formula

$$\sigma^{\mathrm{a}}_{\mathrm{obs}} = \left\{ \mathrm{diag}\left[\mathbf{R}^{1/2} \mathbf{G}(\boldsymbol{x}^{\mathrm{a}}) \mathbf{G}(\boldsymbol{x}^{\mathrm{a}})^{\mathrm{T}} \mathbf{R}^{\mathrm{T}/2} \right] \right\}^{1/2}, \tag{7}$$

where 'diag' stands for diagonal of a matrix and the $N_{\mathrm{obs}} \times N_{\mathrm{E}}$ matrix $\mathbf{G}$ consists of column perturbation vectors $\mathbf{g}_i(\boldsymbol{x}^{\mathrm{a}})$ defined by

$$\begin{aligned} \mathbf{G}(\boldsymbol{x}^{\mathrm{a}}) &= [\, \mathbf{g}_1(\boldsymbol{x}^{\mathrm{a}}) \quad \mathbf{g}_2(\boldsymbol{x}^{\mathrm{a}}) \quad \ldots \quad \mathbf{g}_{\mathrm{N_E}}(\boldsymbol{x}^{\mathrm{a}}) \,], \\ \mathbf{g}_i(\boldsymbol{x}^{\mathrm{a}}) &= \mathbf{R}^{-1/2} \left[H(\boldsymbol{x}^{\mathrm{a}} + \mathbf{p}^{\mathrm{a}}_i) - H(\boldsymbol{x}^{\mathrm{a}}) \right]. \end{aligned} \tag{8}$$

Note similarity between equations (8) and (4): the only difference is that for calculating matrix $\mathbf{G}$ the analysis perturbations $\mathbf{p}^{\mathrm{a}}_i$ are used, while for matrix $\mathbf{Z}$ the forecast perturbations $\mathbf{p}^{\mathrm{f}}_i$ are employed. Note also that both the forecast perturbation vectors $[\boldsymbol{z}_i(\boldsymbol{x}^{\mathrm{a}})]$ and the analysis perturbation vectors $[\mathbf{g}_i(\boldsymbol{x}^{\mathrm{a}})]$ are calculated using non-linear observation operators H, in accordance with the non-linear character of the MLEF algorithm.

Equations (1)–(6) are solved employing an iterative minimization (Zupanski 2005). In addition, as demonstrated in Zupanski *et al*. (2008), assumptions of differentiability, used in the standard gradient-based minimization methods, were not necessary, thus this approach is considered adequate for non-linear and discontinuous cloud-microphysical processes, which play an important role in the experiments of this study.

2.2 *Covariance localization*

Covariance localization (e.g. Houtekamer and Mitchell 2001, Whitaker and Hamill 2002, Ott *et al*. 2004) is an effective way to account for the 'missing degrees of freedom' in ensemble-based data assimilation systems. By 'missing degrees of freedom' we mean that the number of degrees of freedom in the model state variable is much larger than the affordable ensemble size on a given computer. This often happens in applications to complex weather forecast models where the size of the model state variable could easily reach the order of 10^7–10^8, while the computationally feasible ensemble size can hardly be increased beyond the order of 10^2. In our experiments the size of the model state vector $\boldsymbol{x}$ is $N_{\mathrm{S}} \approx 3 \times 10^7$, thus it is several orders of magnitude larger than the ensemble sizes employed (16 and 48 members).

We adopted the covariance localization approach based on the so-called 'local domains', first proposed by Ott *et al*. (2004). In this approach, the entire model domain is partitioned into smaller local domains and the analysis solution is defined independently for each local domain. Due to the use of the globally defined forecast error covariance ($\mathbf{P}_{\mathrm{f}}$), overlapping local domains and/or some kind of smoothing, the assumption of 'independent local domains' is only partially enforced (e.g. Ott *et al*. 2004, Yang *et al*. 2009). The size of the local domains typically reflects the spatial

scales of the processes being analysed (e.g. extratropical or tropical cyclones, ocean currents, carbon transport), thus the assumption that the local domains are, to a degree, independent is considered appropriate. This covariance localization approach was successfully used, in slightly different variants, in many applications (e.g. Hunt *et al*. 2007, Miyoshi and Yamane 2007, Yang *et al*. 2009, Zupanski 2009a,b). We use the variant explained in Zupanski (2009a,b). Unlike in the original Ott *et al*. (2004) approach, we use non-overlapping local domains, which ensure a well-posed minimization problem in each local domain and provide a straightforward definition of information measures, since each observation belongs to a single local domain and thus contributes to the information measures uniquely (Zupanski 2009a). A disadvantage of using non-overlapping local domains is in possible creations of discontinuities at the boundaries between local domains (which could appear visible in the analysis fields). To eliminate/reduce these discontinuities, smoothing of the analysis weights is applied (e.g. Yang *et al*. 2009, Zupanski 2009a,b). The use of non-overlapping local domains, in conjunction with the smoothing, provided a satisfactory solution to the two contradictory requirements: to define a well-posed minimization problem over each local domain and to reduce discontinuous transitions from one local domain to another.

3. Model and data

In the data assimilation experiments of this study, we employ, as a forecast model, the non-hydrostatic weather research and forecasting (WRF) model, which uses the advanced research WRF (ARW) dynamical core (Wicker and Skamarock 2002, Skamarock *et al*. 2005). The WRF-ARW model was configured with 15 km grid spacing and 50 vertical levels, covering a domain of approximately 4500 km $\times$ 4500 km centred over Germany. For initial and boundary conditions the National Centers for Environmental Prediction (NCEP) Global Forecasts System (GFS) analyses were used.

In terms of physics, the Betts–Miller–Janjic (BMJ) convective parameterization (Betts 1986, Betts and Miller 1986, Janjic 1994) was utilized, and the Mellor–Yamada–Janjic (Janjic 1994) planetary boundary layer (PBL) scheme was chosen. The effects of radiative transfer for long- and short-wave radiation were treated by the rapid radiative transfer model (Mlawer *et al*. 1997) and the Dudhia (1989) scheme, respectively. The Mlawer scheme accounts for multiple bands, trace gasses and microphysics species, while the Dudhia scheme accounts for simple downward integration and includes an efficient cloudy and clear sky absorption and scattering. For microphysical processes a single-moment, five-species, cloud-microphysical scheme (Schultz 1995) was used. The scheme of Schultz (1995) has been modified to use the saturation adjustment method of Asai (1965), to slow the melting rate of snow in air slightly warmer than freezing and to allow for the formation of cloud liquid water in unsaturated grid volumes with lapse rates approaching convective instability.

Regarding the observations, we focus our attention on synthetic radiances from the infrared 10.35 μm channel of the future ABI instrument. We assimilate synthetic observations because the real ABI 10.35 μm radiances are not yet available. The synthetic observations were generated using a different forecast model, the Colorado State University/Regional Atmospheric Modeling System (CSU/RAMS; Cotton *et al*. 2003). As RAMS includes an advanced two-moment microphysical scheme (Meyers *et al*. 1997, Saleeby and Cotton 2004), it was considered adequate to

generate realistic cloud-microphysical variables and, consequently, realistic synthetic 10.35 μm radiances. In Grasso *et al*. (2008) it was shown that synthetic GOES-R ABI radiances based on the RAMS microphysical variables were quite reasonable. As in Grasso *et al*. (2008) we create synthetic radiances by applying the satellite observation operator (developed by Greenwald *et al*. 2002, Grasso and Greenwald 2004) to the RAMS microphysical variables. The observation operator includes a radiative transfer model at infrared wavelengths based on the delta-Eddington 2-stream method (Deeter and Evans 1998) and cloud optical property models at all non-visible bands, based on modified anomalous diffraction theory (Mitchell 2000, 2002, Greenwald *et al*. 2002), applied to both liquid and ice particles. It also includes a gas extinction model: Optical Path TRANsmittance (OPTRAN; McMillin *et al*. 1995). The same observation operator is applied to both RAMS and WRF outputs.

4. Synoptic case

The synoptic case chosen for this study is the extratropical cyclone named Kyrill, which lasted during the period 15–19 January 2007. Due to unusually strong winds, Kyrill caused widespread damage and fatalities in Western Europe, especially in the United Kingdom and in Germany. In figure 1(*a*) and (*b*) we show infrared METEOSAT imagery of Kyrill. The imagery in figure 1(*a*), valid at 1212 UTC 18 January 2007, corresponds to the beginning of the first data assimilation cycle and the imagery in figure 1(*b*), valid at 1912 UTC 18 January 2007, corresponds to the end of the seventh data assimilation cycle of the experiments presented. As the figures indicate, there is a well-developed fast-moving cloud system associated with Kyrill. Note that the METEOSAT data plotted in figure 1(*a*) and (*b*) are not assimilated in this study. At present, we only use the METEOSAT data to illustrate the location and the extent of the Kyrill cloud system. Assimilation of the infrared METEOSAT radiances will be performed for the same synoptic case in the next stage of this research and reported in a follow-up manuscript.

In the experiments of this study, we are focusing on the clouds associated with Kyrill, since our goal is to assimilate cloudy ABI infrared radiance observations. The strong winds of this system make the data assimilation problem more difficult since the clouds are moving quickly and their exact locations are difficult to predict. We anticipate that the use of flow-dependent forecast error covariance would be important for this challenging problem, since it could assign larger uncertainties to the cloud-affected areas.

5. Experimental design

The experimental design corresponds to the so-called 'non-identical twins' setup because two different models are used: one (WRF) as a part of data assimilation, and another (RAMS) to create observations. Unlike identical twins, the non-identical twins imply that the forecast models are not perfect, thus achieving experimental conditions resembling assimilation of the real ABI observations.

The WRF model is run over Europe, to capture the extratropical cyclone Kyrill. The horizontal grid spacing of the model is 15 km and there are 50 vertical levels. The synthetic 10.35 μm radiances are created with a footprint of 15 km and assimilated into the system every hour (i.e. data assimilation interval was 1 hour). Note that real ABI observations will be available in higher resolution (0.5–2 km), thus this experiment simulates the conditions of assimilation of observations that were thinned to

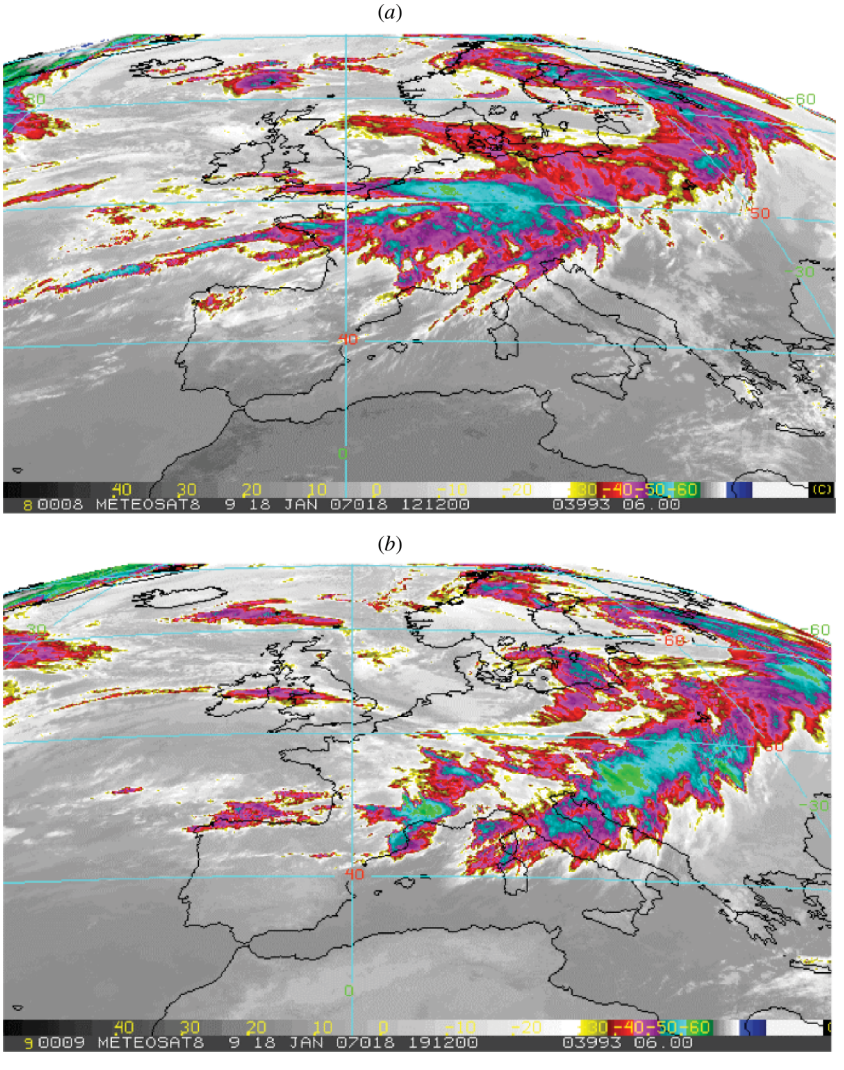

Figure 1. METEOSAT infrared satellite imagery of Kyrill, from 18 January 2007 at (*a*) 1212 UTC and (*b*) 1912 UTC. Note that negative longitude numbers refer to the east longitude (e.g. −60 means 60° E).

approximately match the resolution of the forecast model. Data thinning is often done in routine assimilation of satellite observations at operational meteorological centres to reduce the computational time and storage.

In the experiments presented we used two different ensemble sizes, 16 and 48, and employed two different sets of control variables: with and without cloud ice.

6. Results

6.1 *Selecting control variables*

We selected the following model state variables as control variables of data assimilation: potential temperature, specific humidity and five hydrometeors (cloud water,

cloud ice, rain, snow and graupel). We selected these control variables because we expected that they would play significant roles in the cloud-microphysical processes (described by the Schultz microphysical scheme). Then, we performed sensitivity experiments by excluding one control variable at a time and by examining the impact of the absence of this control variable on the data assimilation results. The purpose of these experiments was to evaluate how important is the choice of the control variables for maximizing information from the assimilated observations.

Our sensitivity experiments indicated that cloud ice and snow had more significant impacts, as measured by improvements to both the data assimilation and the first guess forecast, than the remaining three hydrometeors (cloud water, rain and graupel). As an example, in figures 2–4, we summarize the impact of including/excluding cloud ice into the control variable. The impact of snow was similar, but less pronounced. The results shown in figures 2–4 are produced using all other initially selected control variables (potential temperature, specific humidity, cloud water, rain, snow and graupel), with only cloud ice switching on and off. In figure 2 we show the first guess and the analysis of potential temperature at 850 hPa and the corresponding analysis increments (i.e. differences between the analysis and the first guess), obtained in the experiments without and with using cloud ice as a component of the control variable. In figure 2(*a*) the first guess (1-hour forecast) is given, in figure 2(*b*) and (*c*) the analyses obtained without and with the cloud ice in the control variable are presented, and in figure 2(*d*) and (*e*) the corresponding analysis increments are plotted. By comparing figure 2(*a*) and (*b*) we can see that the potential temperature analysis, obtained in the experiment without cloud ice in the control variable, is dramatically different from the background in the area extending from the Alpine region, through northern Mediterranean and southeastern Europe, towards the Scandinavian Peninsula. These differences exceed 25 K in some points, as shown in figure 2(*d*). By examining figure 2(*a*) and (*c*) we can see that the potential temperature analysis, obtained in the experiment with the cloud ice in the control variable (figure 2(*c*)), is not dramatically different from the background (figure 2(*a*)). The analysis increments in figure 2(*c*) are much smaller, with a maximum magnitude of 10 K.

The next question we pose is as follows. Are the large potential temperature analysis increments, obtained in the experiment without including cloud ice in the control variable, effective in making the forecast model closer to the observations? To answer this question, we examine the results shown in figure 3. As seen in figure 3(*a*), the differences between the first guess and the verification ('observed' 10.35 μm radiances) are large, exceeding -0.06 and $+0.05$ W m^{-2} sr^{-1} cm in the cloudy areas, which roughly corresponds to the errors between -60 and $+50$ K in brightness temperature. The errors of such large magnitudes are due to modelled clouds being shifted with respect to the 'observed' clouds. Because the radiative transfer model detects a warm surface instead of a cold cloud top, and vice versa, the brightness temperature differences are large. The first guess and the analysis errors in figure 3 indicate that the WRF-produced clouds are shifted to southeast compared to the RAMS-produced (i.e. 'observed') clouds. Also note (figure 1(*a*) and (*b*)) that the real clouds of the extratropical cyclone Kyrill are generally in this area; however, there is a shift with respect to the clouds obtained by any of the two models. The large differences in figure 3 (especially the negative differences) were reduced due to data assimilation; in the experiment with cloud ice in the control variable (figure 3(*c*)) the model is generally in better agreement with the 'observations' than in the experiment without cloud ice in the control variable (figure 3(*b*)). There is, however, a relatively large white area in central Europe (around 51° N, 15° E) in figure 3(*b*), which has smaller errors than in the corresponding

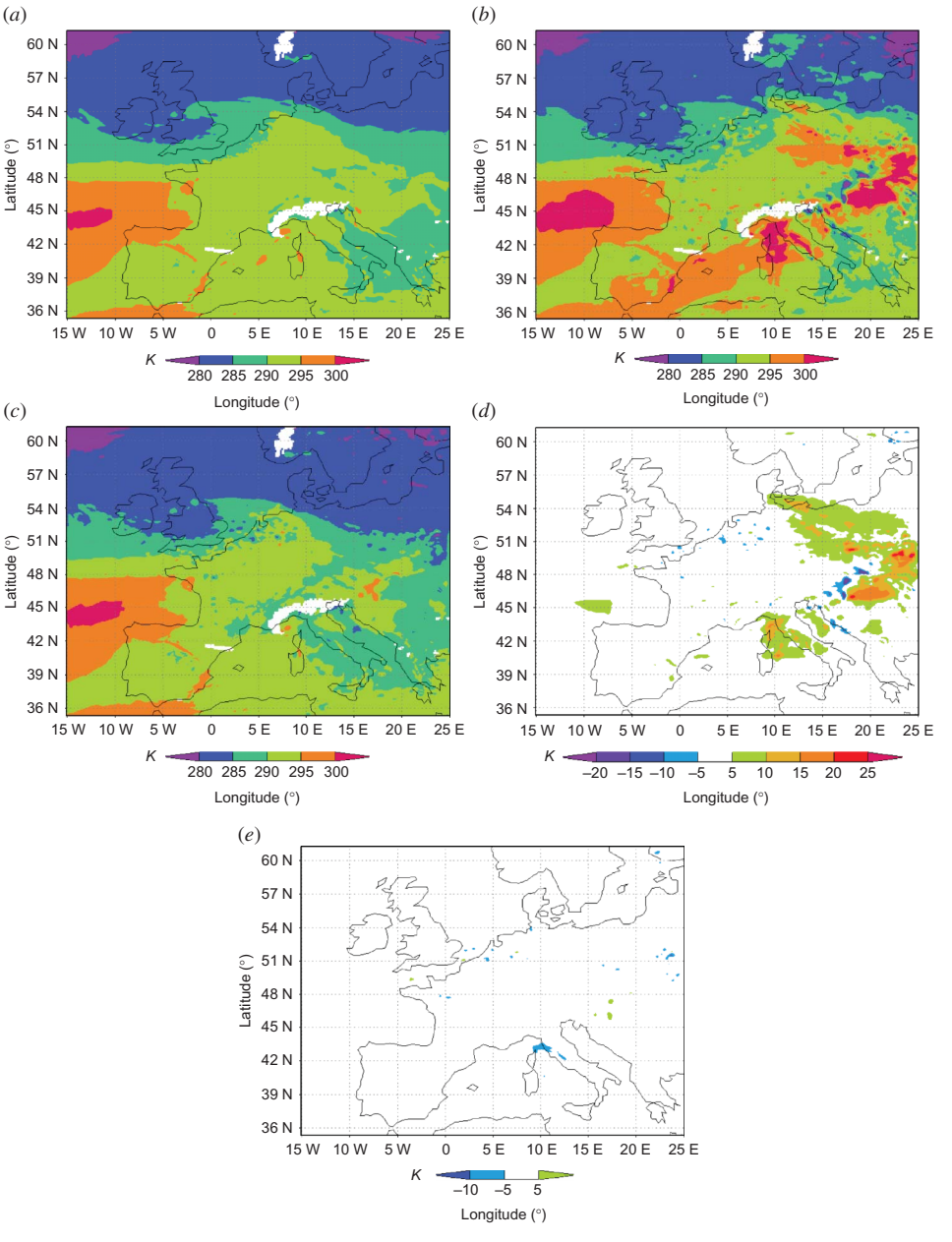

Figure 2. Potential temperature (K) at 850 hPa and the corresponding analysis increments, obtained in the experiments without and with cloud ice in the control variable. The first guess (1-hour forecast) is shown in (*a*), the analysis without cloud ice in the control variable is plotted in (*b*) and the analysis with cloud ice in the control variable is given in (*c*). The corresponding analysis increments are shown in (*d*) and (*e*) for the experiments without and with cloud ice in the control variable, respectively.

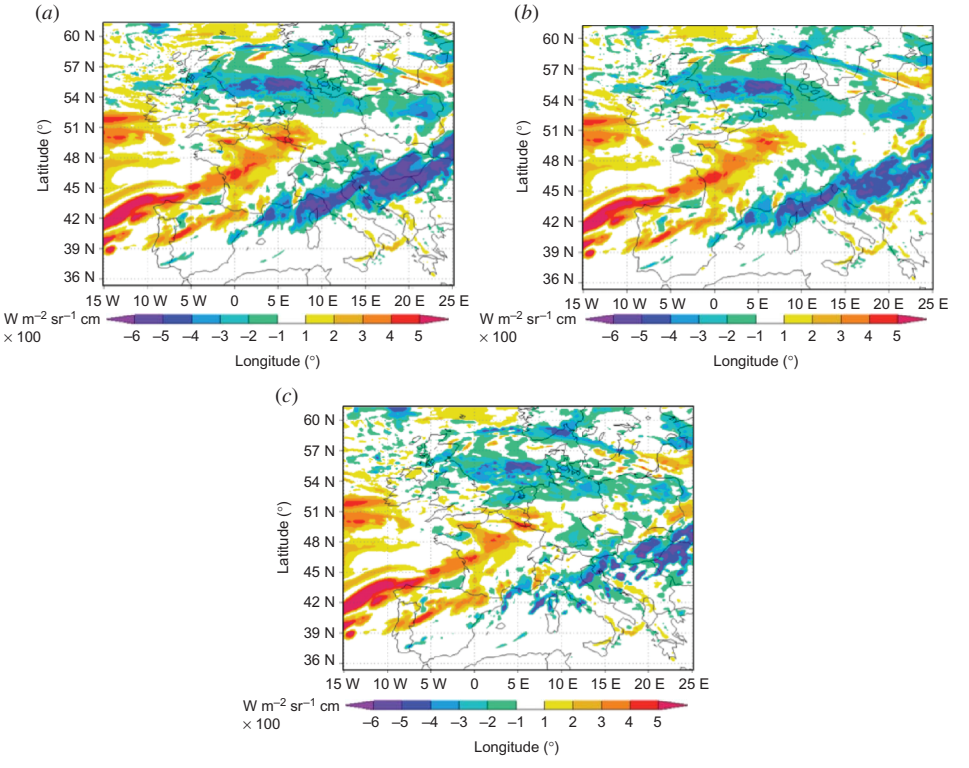

Figure 3. Differences calculated between the model-simulated and the 'observed' 10.35 μm radiances for (*a*) the first guess forecast, (*b*) the analysis without cloud ice in the control variable and (*c*) the analysis with cloud ice in the control variable. The differences correspond to the end of the first data assimilation cycle (1300 UTC 18 January 2007). Units for radiances are W m^{-2} sr^{-1} cm. The values of radiances are scaled by 100.

area in figure 3(*c*), thus indicating that inclusion of cloud ice into the control variable resulted in increased errors in the radiance field in this area. A potential explanation for this could be that the forecast error covariance for cloud ice was not perfect in this area, since inadequate forecast error covariance could result in negative impact of data assimilation (e.g. Morss and Emanuel 2002). Nevertheless, overall results indicate that cloud ice had a more effective impact on improving the analysis than potential temperature in this case, since the unrealistically large changes in potential temperature were eliminated when the cloud ice was included into the control variable.

Finally, we show in figure 4(*a*) the cloud ice analysis increments at 600 hPa obtained in the experiment with the cloud ice in the control variable, and in figure 4(*b*) and (*c*), a quantitative measure of the amount of information of the assimilated observations, called DFS (equation (6)), obtained in the experiments without and with cloud ice in the control variable. We present the 600 hPa level as an example of cloud ice increments. Similar increments were obtained at other higher altitude levels, where the amount of cloud ice was non-negligible. We can see, in figure 4(*a*), that the analysis increments, ranging from −0.1 to 1.5 g kg^{-1}, are present in central and southeastern Europe, in the area of large disagreement between the model and observations. These analysis increments are of rather large magnitude. The maximum value of 1.5 g kg^{-1}

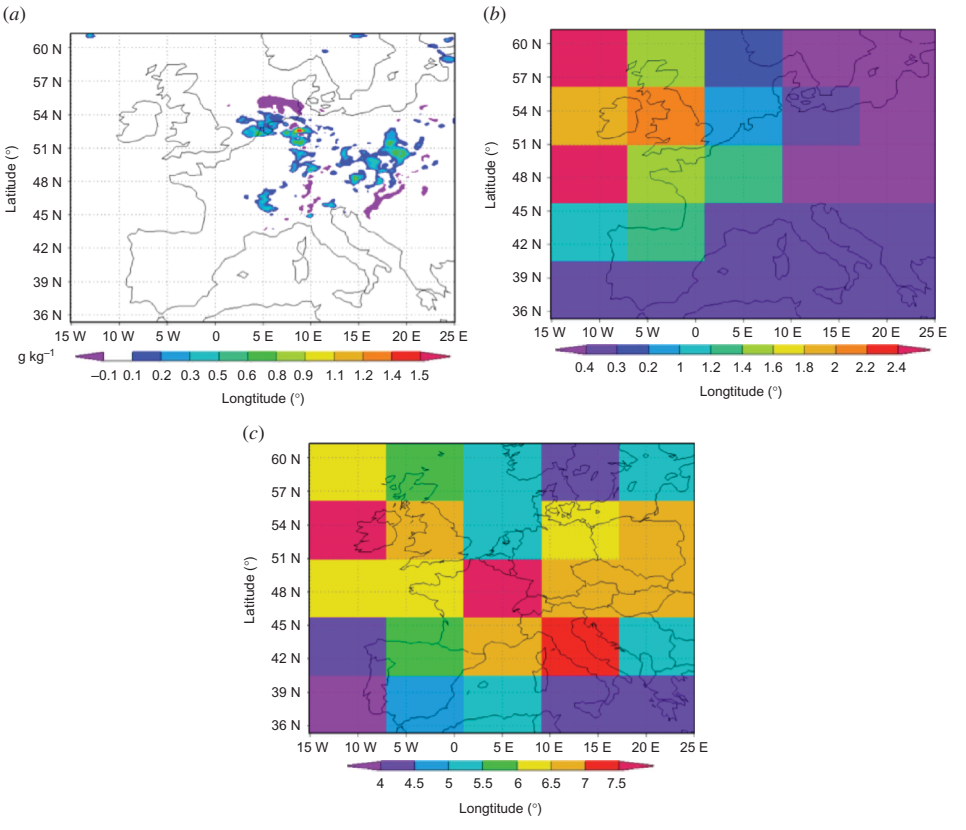

Figure 4. (*a*) Cloud ice analysis increments (in units of g kg^{-1}) at 600 hPa level obtained in the experiment with cloud ice in the control variable in the first data assimilation cycle (1300 UTC 18 January 2007). Information content of the assimilated observations (degrees of freedom for signal (DFS), equation (6)) calculated in the same data assimilation cycle is shown for the experiment without cloud ice in the control variable in (*b*) and for the experiment with cloud ice in the control variable in (*c*). The DFS are non-dimensional quantities.

is about 3 times larger than the maximum value of cloud ice in the background field. A possible explanation for large analysis increments is that there are still some important components of the control variable missing (e.g. $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ wind components, model error) which causes the existing control variable components to change by a greater amount in order to fit the data. Nevertheless, these analysis increments do not result in the ever-increasing amount of cloud ice, since the total amount of cloud ice remains similar throughout all data assimilation cycles. On the other hand, when excluding cloud ice from the control variable, larger changes in other hydrometeors are obtained. For example, the results indicated that analysis increments for snow mixing ratio become 2 to 3 times larger than in the experiment with cloud ice (figure not shown). This is an indication that exclusion of radiatively active hydrometeors (such as cloud ice) from the control variable might force other hydrometeors to account for the missing impact of the cloud ice on the radiance field. We can also see, by comparing figure 4(*b*) and (*c*), that introduction of cloud ice to the control variable results in significantly increased information content of the assimilated observations, even though

the same observations were assimilated in both experiments. This is an indication that the appropriate control variable must be chosen in order to extract more information from the same observations.

Note that the DFS, shown in figure 4(*b*) and (*c*), have blocky structures. The DFS are calculated as total numbers of DFS over each of 25 local domains (blocks), which are the same local domains used for covariance localization (explained in §2.2). Even though the information measures could be calculated for each model grid point (and the blocky structures would disappear), we calculated them over the local domains, thus to measure the information content contained in each local domain.

In summary, the results presented in figures 2–4 demonstrate that adjusting cloud ice is of substantial importance for assimilation of the 10.35 μm radiances, in cases when cloud ice is present at the cloud tops. These results also indicate that excluding cloud ice from the control variable results in unrealistic changes to the remaining components of the control variable (e.g. potential temperature), since these remaining components could never account, in a physically correct way, for the effect of the cloud ice. This finding confirms indications from earlier studies (e.g. Smith *et al*. 1992, Grasso and Greenwald 2004, Grasso *et al*. 2008) that the 10.35 μm channel should be sensitive to cloud ice. For example, in Grasso and Greenwald (2004), it was demonstrated that the 10.35 μm channel should be most sensitive to the hydrometeors at the top of the thunderstorm. Since the top of the thunderstorm mostly contains the cloud ice particles (also called pristine ice), the 10.35 μm channel basically measures the cloud ice in this case. More generally, the results shown in figures 2–4 indicate that it is imperative to include all radiatively active hydrometeors into the control variable for maximizing the benefits of assimilated cloudy radiances and to avoid obtaining degraded data assimilation results due to neglecting some of the important hydrometeors. On the other hand, including microphysical variables to which the radiances have little sensitivity (e.g. rain or graupel) had negligible impact on data assimilation, thus these variables could be either included or excluded from the control variable. In the experiments presented in the remainder of this article, we keep all initially selected microphysical variables (potential temperature, specific humidity, cloud water, cloud ice, rain, snow and graupel) as components of the control variable.

6.2 *Data assimilation experiments over multiple data assimilation cycles*

In figure 5, we present the root mean square (RMS) errors of the analysis and the first guess, calculated with respect to the RAMS-simulated 10.35 μm radiances, as functions of data assimilation cycles. For reference, the RMS errors of the experiment without data assimilation are also included in figure 5. As seen in figure 5, both the analysis and the first guess are in better agreement with the 'truth' than the experiment without data assimilation. Furthermore, the analysis indicates clear improvements with respect to the first guess, and the ensemble size of 48 members has smaller errors than the ensemble size of 16 members. The errors in both data assimilation experiments (with 16 and 48 members) generally decrease with time until around cycle 12, when they saturate at a certain level (with the exceptions of increasing errors from cycle 8 to 12 in some experiments). Note that the errors also decrease, at a slower rate, in the experiment without data assimilation. This is because the influence of the horizontal boundary conditions becomes more dominant with time. Since the same NCEP analyses (from the Gridpoint Statistical Interpolation (GSI) system (Wu *et al*. 2002)) were used to create boundary conditions for both WRF and RAMS, the differences

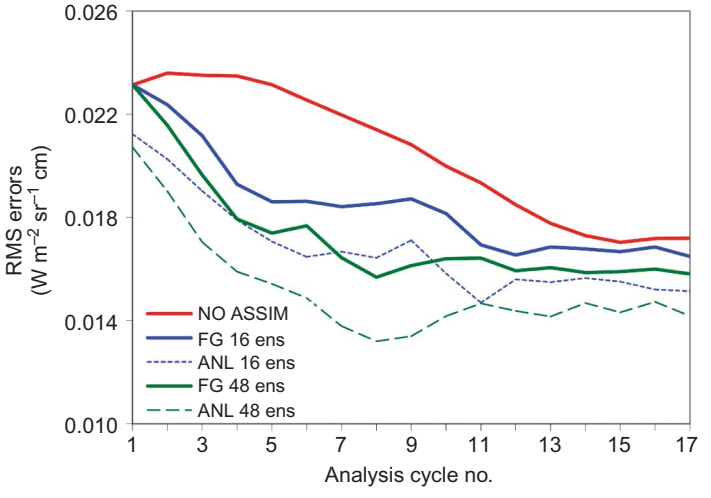

Figure 5. Root mean square (RMS) errors of the analysis (ANL, dashed lines) and the first guess (FG, solid lines) calculated with respect to the Regional Atmospheric Modeling System (RAMS)-simulated 10.35 μm radiances and plotted as functions of data assimilation cycles. Results from the experiments with 16 and 48 ensemble members are shown (e.g. 'FG 16 ens' means 'first guess from the experiment using 16 ensembles'). For reference, the RMS errors of the experiment without data assimilation (NO ASSIM, solid line) are also included. Units for radiances are W m^{-2} sr^{-1} cm.

between the two models decrease with time, and consequently the differences between the WRF-simulated and the RAMS-simulated (i.e. 'observed') radiances decrease too. In summary, figure 5 indicates a positive impact of data assimilation of the 10.35 μm synthetic radiances on the analysis and short-term forecast and a positive impact of the increased ensemble size, over 17 data assimilation cycles. We have not run data assimilation experiments beyond 17 data assimilation cycles since the analysis and forecast errors saturated around cycle 12 in all experiments.

We also show, in figure 6, the histogram of error distributions for the data assimilation experiment with 48 ensemble members and the experiment without data assimilation. By comparing the results from cycles 1 and 7, shown in panel 6(*a*) and (*b*), respectively, we can see that the errors are clustering around zero in the later cycle (i.e. cycle 7), indicating improvements in the first guess, the analysis and the no assimilation experiment. The errors from the analysis and the first guess cluster more around zero than the errors of the experiment without assimilation (note outliers in the positive errors for the no assimilation case). These results confirm a positive impact of data assimilation on the analysis and short-term forecast improvements.

One can also see in figure 6(*a*) and (*b*) that the errors follow a Gaussian distribution reasonably well. This is an indication that the commonly used assumption about Gaussian errors (and also used in this study) is reasonable in this case. Note, however, that cloudy satellite radiances could often depart from Gaussian distribution (e.g. could follow log-normal distribution: cf. Grasso *et al*. (2010)). In such cases a different cost function needs to be minimized as proposed in Fletcher and Zupanski (2006a,b).

We now examine whether the actual analysis errors are in agreement with the estimated analysis uncertainty. Examples of difference fields calculated between the

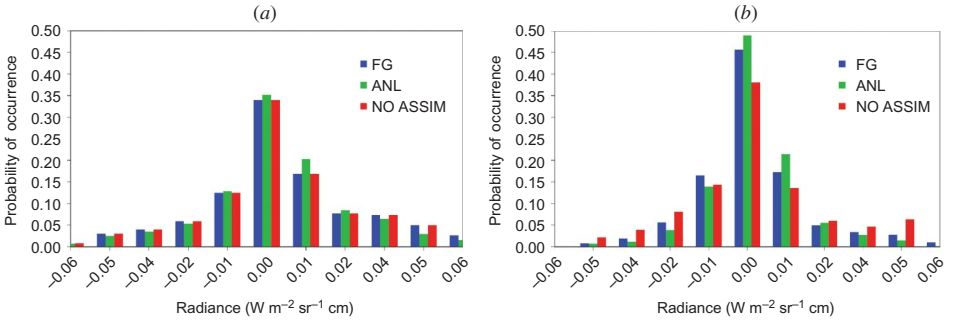

Figure 6. Probability histogram of the errors of the 10.35 μm radiances, calculated with respect to Regional Atmospheric Modeling System (RAMS)-simulated radiances, for the first guess (FG), the analysis (ANL) and the experiment without assimilation (NO ASSIM). Data assimilation experiments using ensemble size of 48 members are shown. Results from first (*a*) and seventh (*b*) data assimilation cycles are presented. Units for radiances are W m^{-2} sr^{-1} cm.

model-simulated radiances and the 'observed' radiances at 10.35 μm and the corresponding analysis uncertainty σ^{a}_{obs}(equation (7)) are shown in figures 7 and 8. The differences are plotted for data assimilation cycle 7, valid at 1900 UTC 18 January 2007, and they correspond to the experiment without assimilation (i.e. old forecast, figure 7(*a*)), the first guess (1-hour forecast after data assimilation, figure 7(*b*)) and the analysis (figure 7(*c*)). As indicated before, there are large discrepancies between the model and the 'observations' in areas over the Alps, the northern Mediterranean and southeastern Europe. These discrepancies remain relatively large throughout the entire period (17 data assimilation cycles); however, they are decreasing with time in both experiments (with and without data assimilation), as the RMS errors in figure 5 indicate. Nevertheless, we can still see a pronounced dipole of positive–negative differences (with maxima exceeding −0.06 or +0.05 W m^{-2} sr^{-1} cm) extending over the Alps, the northern Mediterranean and southeastern Europe in figure 7(*a*) (experiment without data assimilation).

The impact of data assimilation is to significantly reduce the magnitudes of the large differences for the first guess (figure 7(*b*)) and to a lesser extent the analysis (figure 7(*c*)). Note, however, that the large differences still remain in this area and they appear in some additional areas (e.g. over the northern part of the Pyrenean peninsula). Due to the non-identical twins setup we do not expect that the differences should eventually become negligible throughout the entire model domain, as they would in the identical twins setup. We do expect, however, that the method should be able to assign realistic (i.e. large) uncertainties to the points where these differences are large.

As we can see in figure 8(*a*), the analysis uncertainty is generally larger in the areas where the analysis-minus-observations differences are larger (in figure 7(*c*)). We can also see that the magnitudes of the analysis errors are in good agreement: for example, there are areas of large analysis errors, exceeding ±0.04 W m^{-2} sr^{-1} cm, in figure 7(*c*) (e.g. Alpine region, northern Mediterranean, southeastern and northeastern Europe), and the estimated analysis uncertainty is generally in the range 3.5–5 W m^{-2} sr^{-1} cm in these areas. These results confirm that the MLEF calculated analysis uncertainty is a good estimate of the actual analysis errors in terms of both the magnitude

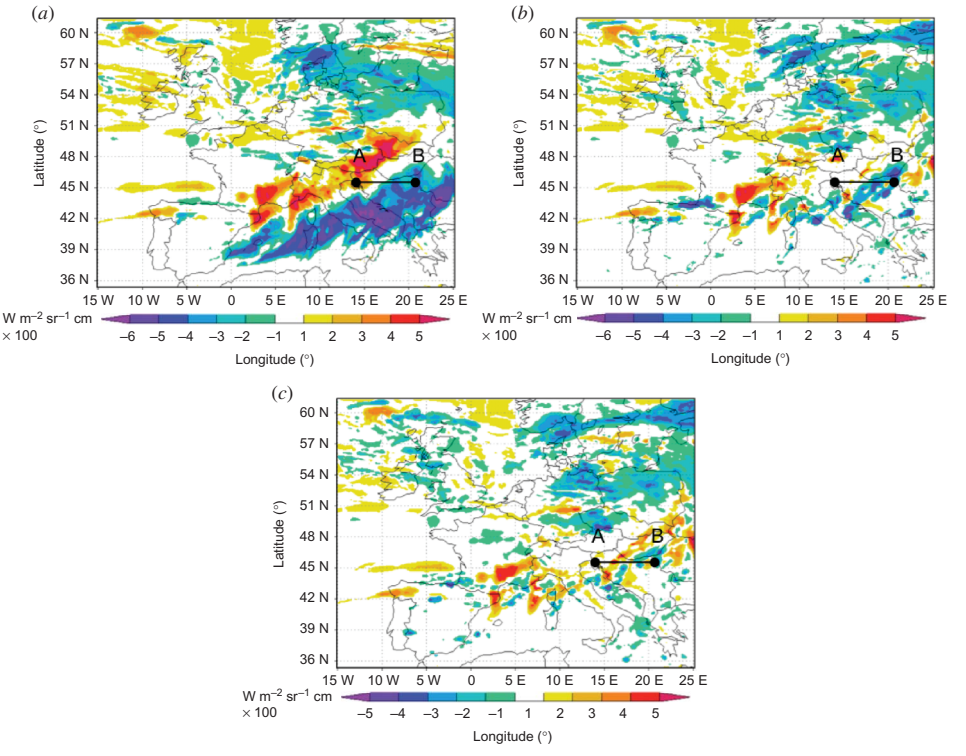

Figure 7. Differences, corresponding to the end of the seventh data assimilation cycle (1900 UTC 18 January 2007), calculated between the model-simulated and the 'observed' 10.35 μm radiances for (*a*) the experiment without data assimilation (i.e. old forecast), (*b*) 1-hour forecast used as a first guess in data assimilation and (*c*) the analysis. The units for radiances are W m^{-2} sr^{-1} cm. The values of radiances are scaled by 100. Line AB indicates location of the vertical cross section shown in figure 9.

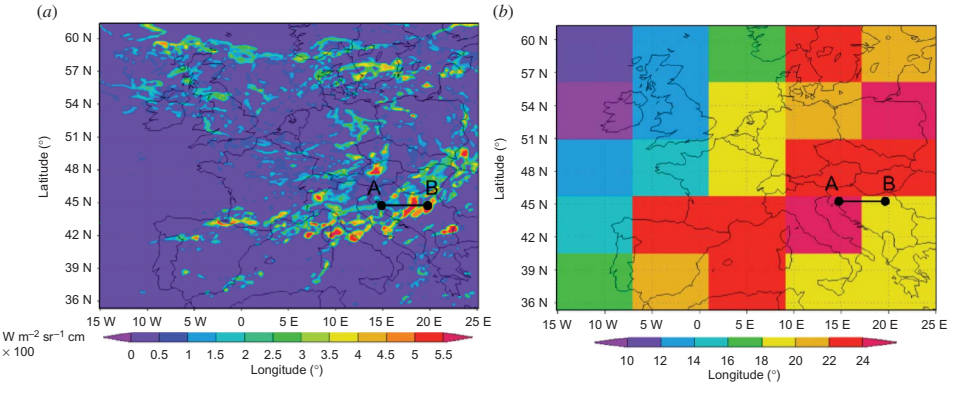

Figure 8. Analysis uncertainty for the 10.35 μm radiances (σ^{a}_{obs}, equation (7)) is plotted in (*a*) and the corresponding DFS (equation (6)) is shown in (*b*). The results are given for the seventh data assimilation cycle, thus are comparable to the results in figure 7. The units for radiances are W m^{-2} sr^{-1} cm; however, the values are scaled by 100.

and structure of the errors. In figure 8(*b*) we show an information measure, DFS (equation (6)), calculated in data assimilation cycle 7. As the figure indicates, areas (i.e. blocks) with large values of DFS generally coincide with the large analysis uncertainty in figures 7(*c*) and 8(*a*). The results shown in figure 8(*a*) and (*b*) confirm the flow-dependent character of the analysis error covariance and the information measures, because the analysis errors and the information measures follow flow characteristics of the Kyrill cloud system. Flow-dependent analysis and forecast error covariance matrices are considered important ingredients of advanced data assimilation systems.

Let us now examine the vertical cross section taken along the line AB (the location of the cross section is shown in figures 7 and 8). As seen in figure 7, the cross section is taken in the area where the differences between the model and the 'observations' are large and they also change sign within the cross section. The analysis uncertainty and the information measure (DFS) are also large in this area (see figure 8). We present the vertical cross section of potential temperature and cloud ice in figure 9, including

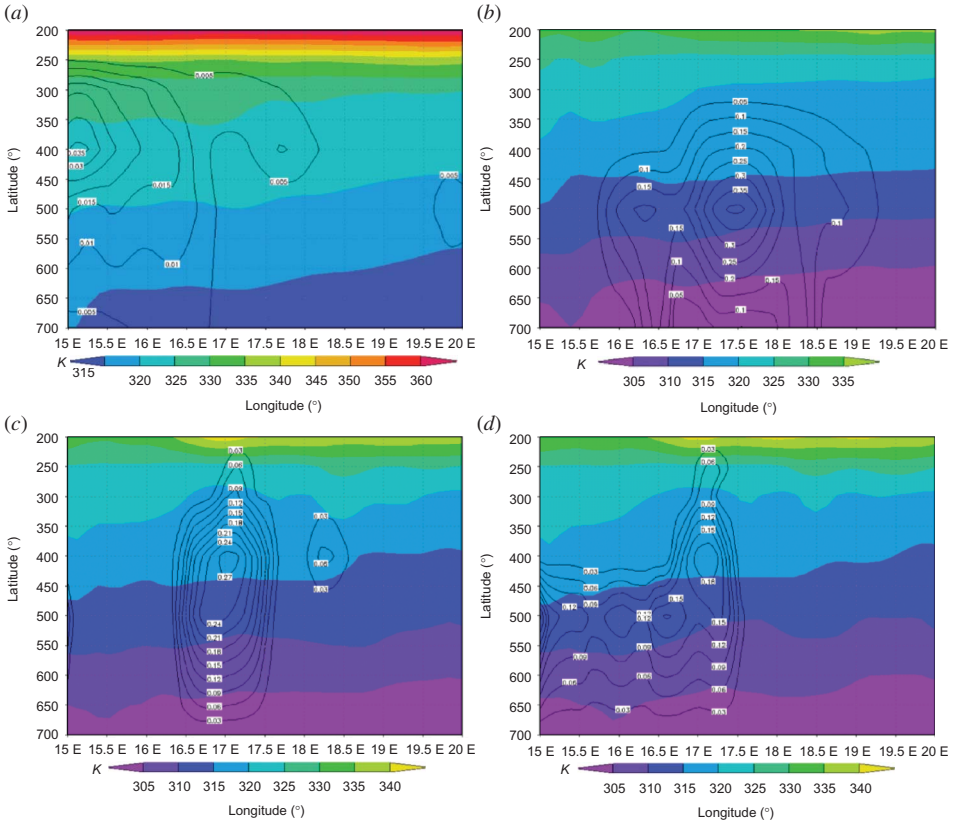

Figure 9. Vertical cross section taken along the line AB is shown in figures 7 and 8. Potential temperature (shading) and cloud ice (contours) are shown in the figure. The results are valid at the end of seventh data assimilation cycle (1900 UTC 18 January 2007). The 'observations' (RAMS forecasts) are given in panel (*a*), the forecast results from the experiment without assimilation are shown in (*b*), the 3-hour background in (*c*) and the analysis in (*d*). Unit for potential temperature is *K* and for cloud ice g kg^{-1}.

'observations' and model results. We can see that the model-produced potential temperature is generally colder in the lower levels compared to the 'observations', and the opposite is true for the upper levels. We also see large differences between the modelled and the 'observed' cloud ice in both magnitudes and locations of the maximum values. These differences got reduced due to data assimilation: modest changes in potential temperature and more pronounced changes in cloud ice can be seen. For example, the maximum of the cloud ice has moved westward in the first guess forecast (figure 9(*c*)) and the analysis (figure 9(*d*)), which is in better agreement with the 'observations' (figure 9(*a*)) than the maximum obtained in the experiment without data assimilation (figure 9(*b*)). Nevertheless, significant discrepancies between the analysis and the 'observations' still exist (the maximum should be 2° farther west and the maximum in the analysis is stronger than the observed (0.18 vs. 0.035 g kg^{-1})). These discrepancies could be due to significant differences between the two models in both dynamical cores and cloud-microphysical schemes, which are difficult to eliminate without altering the models' equations (e.g. without taking into account model errors).

Important to note is that the vertical structure of the cloud ice analysis (figure 9(*d*)) is generally in better agreement with the 'observations' than the first guess (figure 9(*c*)) and the experiment without assimilation (figure 9(*b*)). The figures indicate that the cloud ice maximum got lifted to the upper levels (from around 500 hPa in the experiment without assimilation to around 400 hPa in the experiment with assimilation). The changes in the potential temperature field, even though marginal, when combined with the more pronounced change in the vertical position of the cloud ice maximum, resulted in significant changes in the radiance field. Thus, modifying cloud ice at the upper, rather than lower, levels has larger impact on the radiance field since the energy being radiated by the cloud ice to the satellite is less impeded. Our calculations also indicate that shifting the cloud ice maximum to the higher levels resulted in the change of temperature in the area of the cloud ice maximum from (approximately) 257 to 243 K, which is closer to the 'correct' value of 249 K. This correction might be quite significant for the model microphysics. In conclusion, the impact of data assimilation is seen in more pronounced changes (which were also in the right direction) in the cloud ice than in the potential temperature. These results are reasonable, since the 10.35 μm radiance is expected to be more sensitive to the cloud ice than to any other microphysical variable under the conditions of this experiment (also confirmed in Grasso and Greenwald (2004) for a similar window channel of 10.7 μm).

7. Conclusions

In this study we assimilated synthetic GOES-R ABI radiances at 10.35 μm in cloudy conditions to evaluate the potential impact of these observations on improving model-simulated clouds. In particular, we were interested in extracting maximum information from such observations by taking into account when and where this information is needed the most. Our criteria for defining when and where the information is needed the most were based on information theory and the use of information measures (such as DFS). We have performed cycled data assimilation experiments for the case of the extratropical cyclone Kyrill. This case was especially challenging because of the large errors in model-simulated radiances due to spatially shifted clouds.

The experimental results indicated that the same GOES-R observations could bring varying amounts of information, depending on the choice for the control variable in

data assimilation. When cloud ice was included into the control variable, the information content of data, as measured by DFS, was significantly increased and the data assimilation results were much improved compared to the case without including cloud ice into the control variable. The impact of data assimilation was also seen in much more significant changes to the cloud ice than to other model state variables, such as potential temperature. This is not surprising, since the 10.35 μm channel is expected to be sensitive to the hydrometeors at the cloud top, but not sensitive to the air temperature or potential temperature (e.g. Smith *et al*. 1992, Grasso and Greenwald 2004). Therefore, our results indicated that it is imperative to include all radiatively active hydrometeors into the control variable for maximizing the benefits of assimilated cloudy radiances and to avoid obtaining degraded data assimilation results due to neglecting some of the important hydrometeors. On the other hand, microphysical variables to which the radiances have little sensitivity (e.g. rain or graupel) could be either included or excluded from the control variable.

The data assimilation and short-term forecast results over multiple data assimilation cycles have clearly indicated improvements due to assimilation of the GOES-R ABI radiance 'observations', compared to the experiment without data assimilation. The experimental results also indicated that the flow-dependent DFS used in this study realistically reflect the actual forecast uncertainty: the values of DFS were the largest in the areas of largest forecast errors. This is a confirmation that the data assimilation method used in this study has the capability to maximize information content of the assimilated observations.

In future work, we plan to further evaluate the capability of the MLEF approach, and the ensemble-based data assimilation methods in general, to extract maximum information from real satellite observations. For this purpose we plan to assimilate cloudy radiances similar to the future ABI radiances (e.g. from the METEOSAT Second Generation – MSG and/or the AIRS instruments). We will also apply the forecast model in finer spatial and temporal resolution (preferably using the grid spacing of less than 10 km). This will provide more realistic conditions for assimilating real satellite radiances in finer resolution and taking into account small-scale processes (e.g. non-hydrostatic effects and cloud-microphysical processes). This would potentially further improve assimilation of satellite radiances in cloudy conditions. Assimilation of real satellite radiances could additionally be improved by including model error into the control variable (e.g. by state augmentation as in Zupanski and Zupanski 2006), which is planned for the future.

Acknowledgements

The authors are grateful to two anonymous reviewers for their insightful comments. This work was supported by the National Oceanic and Atmospheric Administration/National Environmental Satellite, Data and Information Service (NOAA/NESDIS) GOES-R Risk Reduction Program, grant no. NA17RJ1228, and the National Science Foundation Collaboration in Mathematical Geosciences Program, grant no. ATM-0327651. Our gratitude is also extended to National Center for Atmospheric Research (NCAR) for the computational time on the IBM Bluefire supercomputer. The views, opinions and findings in this report are those of the authors and should not be construed as an official NOAA and or US government position, policy or decision.

References

Aksoy, A., Dowell, D.C. and Snyder, C., 2009, A multi-scale comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: storm-scale analysis. *Monthly Weather Review*, **137**, pp. 1805–1824.

Andersson, E., Bauer, P., Beljaars, A., Chevallier, F., Holm, E., Janiskova, M., Kallberg, P., Kelly, G.A., Lopez, P., McNally, A.P., Moreau, E., Simmons, A.J., Thepaut, J.-N. and Tompkins, A.M., 2005, Assimilation and modeling of the atmospheric hydrological cycle in the ECMWF forecasting system. *Bulletin of the American Meteorological Society*, **86**, pp. 387–402.

Asai, T., 1965, A numerical study of the air-mass transformation over the Japan Sea in winter. *Journal of the Meteorological Society of Japan*, **43**, pp. 1–15.

Bauer, P., Lopez, P., Benedetti, A., Salmond, D. and Moreau, E., 2006a, Implementation of 1D+4D-Var assimilation of precipitation-affected microwave radiances in precipitation at ECMWF. I: 1D-Var. *Quarterly Journal of the Royal Meteorological Society*, **132**, pp. 2227–2306.

Bauer, P., Lopez, P., Benedetti, A., Salmond, D., Saarinen, S. and Bonazzola, M., 2006b, Implementation of 1D+4D-Var assimilation of precipitation-affected microwave radiances in precipitation at ECMWF. II: 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, **132**, pp. 2307–2332.

Betts, A.K., 1986, A new convective adjustment scheme. Part I: observational and theoretical basis. *Quarterly Journal of the Royal Meteorological Society*, **112**, pp. 677–692.

Betts, A.K. and Miller, M.J., 1986, A new convective adjustment scheme. Part II: single column tests using GATE wave, BOMEX, and arctic air-mass data sets. *Quarterly Journal of the Royal Meteorological Society*, **112**, pp. 693–709.

Carton, J.A. and Giese, B.S., 2008, A reanalysis of ocean climate using simple ocean data assimilation (SODA). *Monthly Weather Review*, **136**, pp. 2999–3017.

Chevallier, F., Lopez, P., Tompkins, A., Janisková, M. and Moreau, E., 2004, The capability of 4D-Var systems to assimilate cloud-affected satellite infrared radiances. *Quarterly Journal of the Royal Meteorological Society*, **130**, pp. 917–932.

Cotton, W.R., Pielke, Sr., R.A., Walko, R.L., Liston, G.E., Tremback, C.J., Jiang, H., McAnelly, R.L., Harrington, J.Y., Nicholls, M.E., Carrio, G.G. and McFadden, J.P., 2003, RAMS 2001: current status and future direction. *Meteorology and Atmospheric Physics*, **82**, pp. 5–29.

Deeter, M. and Evans, K.F., 1998, A hybrid Eddington-single scatter radiative transfer model for computing radiances from thermally emitting atmospheres. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **60**, pp. 635–648.

Dudhia, J., 1989, Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences*, **46**, pp. 3077–3107.

Errico, R., Mahfouf, J.-F. and Bauer, P., 2007, Issues regarding the assimilation of cloud and precipitation data. *Journal of the Atmospheric Sciences*, **64**, pp. 3785–3798.

Fletcher, S.J. and Zupanski, M., 2006a, A hybrid multivariate normal and lognormal distribution for data assimilation. *Atmospheric Science Letters*, **7**, pp. 43–46.

Fletcher, S.J. and Zupanski, M., 2006b, A data assimilation method for lognormally distributed observational errors. *Quarterly Journal of the Royal Meteorological Society*, **132**, pp. 2505–2519.

Grasso, L.D. and Greenwald, T., 2004, Analysis of 10.7 μm brightness temperatures of a simulated thunderstorm with two-moment microphysics. *Monthly Weather Review*, **132**, pp. 815–825.

Grasso, L.D., Sengupta, M. and DeMaria, M., 2010, Comparison between observed and synthetic 6.5 and 10.7 μm GOES-12 imagery of thunderstorms that occurred on 8 May 2003. *International Journal of Remote Sensing*, **31**, pp. 647–663.

Grasso, L.D., Sengupta, M., Dostalek, J.F., Brummer, R. and DeMaria, M., 2008, Synthetic satellite imagery for current and future environmental satellites. *International Journal of Remote Sensing*, **29**, pp. 4373–4384.

Greenwald, T.J., Hertenstein, R. and Vukicevic, T., 2002, An all-weather observational operator for radiance data assimilation with mesoscale forecast models. *Monthly Weather Review*, **130**, pp. 1882–1897.

Hollingsworth, A., Engelen, R.J., Textor, C., Benedetti, A., Boucher, O., Chevallier, F., Dethof, A., Elbern, H., Eskes, H., Flemming, J., Granier, C., Kaiser, J.W., Morcrette, J.J., Rayner, P., Peuch, V.H., Rouil, L., Schultz, M.G., Simmons, A.J. and The GEMS Consortium, 2008, Toward a monitoring and forecasting system for atmospheric composition: the GEMS project. *Bulletin of the American Meteorological Society*, **89**, pp. 1147–1164.

Hou, A.Y. and Zhang, S.Q., 2007, Assimilation of precipitation information using column model physics as a weak constraint. *Journal of the Atmospheric Sciences*, **64**, pp. 3865–3878.

Houtekamer, P.L. and Mitchell, H.L., 2001, A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, **129**, pp. 123–137.

Hunt, B.R., Kostelich, E.J. and Szunyogh, I., 2007, Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D*, **230**, pp. 112–126.

Janjic, Z.I., 1994, The step-mountain Eta coordinate model: further developments of the convection closure schemes. *Monthly Weather Review*, **122**, pp. 927–945.

Jung, J.A., Zapotocny, T.H., Le Marshall, J.F. and Treadon, R.E., 2008, A two-season impact study of NOAA polar-orbiting satellites in the NCEP global data assimilation system. *Weather and Forecasting*, **23**, pp. 854–877.

Keppenne, C.L., Rienecker, M.M., Jacob, J.P. and Kovach, R., 2008, Error covariance modeling in the GMAO ocean ensemble Kalman filter. *Monthly Weather Review*, **136**, pp. 2964–2982.

L'Ecuyer, T., Gabriel, P., Leesman, K., Cooper, S. and Stephens, G., 2006, Objective assessment of the information content of visible and infrared radiance measurements for cloud microphysical property retrievals over the global oceans. Part 1: liquid clouds. *Journal of Applied Meteorology and Climatology*, **45**, pp. 20–41.

Lin, X., Zhang, S.Q. and Hou, A.Y., 2007, Variational assimilation of global microwave rainfall retrievals: physical and dynamical impact on GEOS analyses. *Monthly Weather Review*, **135**, pp. 2931–2957.

Liu, H., Anderson, J., Kuo, Y.H., Snyder, C. and Caya, A., 2008, Evaluation of a nonlocal quasi-phase observation operator in assimilation of CHAMP radio occultation refractivity with WRF. *Monthly Weather Review*, **136**, pp. 242–256.

McMillin, L.M., Crone, L.J., Goldberg, M.D. and Kleespies, T.J., 1995, Atmospheric transmittance of an absorbing gas. 4. OPTRAN: a computationally fast and accurate transmittance model for absorbing gases with fixed and variable mixing ratios at variable viewing angles. *Applied Optics*, **34**, pp. 6269–6274.

Meng, Z. and Zhang, F., 2008, Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: comparison with 3DVAR in a month-long experiment. *Monthly Weather Review*, **136**, pp. 3672–3682.

Meyers, M.P., Walko, R.L., Harrington, J.Y. and Cotton, W.R., 1997, New RAMS cloud microphysics parameterization. Part II: the two-moment scheme. *Atmospheric Research*, **45**, pp. 3–39.

Migliorini, S., Piccolo, C. and Rodgers, C.D., 2008, Use of the information content in satellite measurements for an efficient interface to data assimilation. *Monthly Weather Review*, **136**, pp. 2633–2650.

Mitchell, D.L., 2000, Parameterization of the Mie extinction and absorption coefficients for water clouds. *Journal of the Atmospheric Sciences*, **57**, pp. 1311–1326.

Mitchell, D.L., 2002, Effective diameter in radiation transfer: general definitions, applications, and limitations. *Journal of the Atmospheric Sciences*, **59**, pp. 2330–2346.

Miyoshi, T. and Yamane, S., 2007, Local ensemble transform Kalman filtering with an AGCM at T159/L60 resolution. *Monthly Weather Review*, **135**, pp. 3841–3861.

Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J. and Clough, S.A., 1997, Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, **102**, pp. 16663–16682.

Morss, R.E. and Emanuel, K.A., 2002, Influence of added observations on analysis and forecast errors: results from idealized systems. *Quarterly Journal of the Royal Meteorological Society*, **128**, pp. 285–322.

Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V., Kostelich, E.J., Corazza, M., Kalnay, E., Patil, D.J. and Yorke, J.A., 2004, A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, pp. 273–277.

Reichle, R.H., 2007, Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the scanning multichannel microwave radiometer (SMMR). *Journal of Geophysical Research*, **112**, D09108, doi:10.1029/2006JD008033.

Rodgers, C.D., 2000, *Inverse Methods for Atmospheric Sounding: Theory and Practice*, 238 pp. (Singapore: World Scientific).

Saleeby, S.M. and Cotton, W.R., 2004, A large-droplet mode and prognostic number concentration of cloud droplets in the Colorado State University Regional Atmospheric Modeling System (RAMS). Part I: module descriptions and supercell test simulations. *Journal Applied Meteorology*, **43**, pp. 182–195.

Schmit, T.J., Gunshor, M.M., Menzel, W.P., Gurka, J.J., Li, J. and Bachmeier, S., 2005, Introducing the next-generation advanced baseline imager on GOES-R. *Bulletin of the American Meteorological Society*, **86**, pp. 1079–1096.

Schmit, T.J., Li, J., Gurka, J.J., Goldberg, M.D., Schrab, K.J., Li, J. and Feltz, W.F., 2008, The GOES-R advanced baseline imager and the continuation of current sounder products. *Journal of Applied Meteorology and Climatology*, **47**, pp. 2696–2711.

Schultz, P., 1995, An explicit cloud physics parameterization for operational numerical weather prediction. *Monthly Weather Review*, **123**, pp. 3331–3343.

Shannon, C.E. and Weaver, W., 1949, *The Mathematical Theory of Communication*, 144 pp. (Urbana: University of Illinois Press).

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Wang, W. and Powers, J.G., 2005, *A Description of the Advanced Research WRF Version 2*, NCAR Technical Notes, 5, 22 pp. (Boulder, CO: National Center for Atmospheric Research).

Smith, W.L., Ma, X.L., Ackerman, S.A., Revercomb, H.E. and Knuteson, R.O., 1992, Remote sensing cloud properties from high-spectral resolution infrared observations. *Journal of the Atmospheric Sciences*, **50**, pp. 1708–1720.

Vukicevic, T., Greenwald, T., Zupanski, M., Zupanski, D., Vonder Haar, T. and Jones, A.S., 2004, Mesoscale cloud state estimation from visible and infrared satellite radiances. *Monthly Weather Review*, **132**, pp. 3066–3077.

Vukicevic, T., Sengupta, M., Jones, A. and Vonder Haar, T., 2006, Cloud resolving data assimilation: information content of IR window observations and uncertainties in estimation. *Journal of the Atmospheric Sciences*, **63**, pp. 901–919.

Wang, X., Bishop, C.H. and Julier, S.J., 2004, Which is better, an ensemble of positive/negative pairs or a centered spherical simplex ensemble? *Monthly Weather Review*, **132**, pp. 1590–1605.

Wei, M., Toth, Z., Wobus, R., Zhu, Y., Bishop, C.H. and Wang, X., 2006, Ensemble transform Kalman filter-based ensemble perturbations in an operational global prediction system at NCEP. *Tellus*, **58A**, pp. 28–44.

WENG, F., ZHU, T. and YAN, B., 2007, Satellite data assimilation in numerical weather prediction models. Part II: uses of rain-affected radiances from microwave observations for hurricane vortex analysis. *Journal of the Atmospheric Sciences*, **64**, pp. 3910–3925.

WHITAKER, J.S. and HAMILL, T.M., 2002, Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, **130**, pp. 1913–1924.

WHITAKER, J.S., HAMILL, T.M., WEI, X., SONG, Y. and TOTH, Z., 2008, Ensemble data assimilation with the NCEP global forecast system. *Monthly Weather Review*, **136**, pp. 463–482.

WICKER, L.J. and SKAMAROCK, W.C., 2002, Time-splitting methods for elastic models using forward time schemes. *Monthly Weather Review*, **130**, pp. 2088–2097.

WU, W.-S., PURSER, R.J. and PARRISH, D.F., 2002, Three-dimensional variational analysis with spatially inhomogeneous covariances. *Monthly Weather Review*, **130**, pp. 2905–2916.

YANG, S.-C., KALNAY, E., HUNT, B. and BOWLER, N.E., 2009, Weight interpolation for efficient data assimilation with the local ensemble transform Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **135**, pp. 251–262.

ZOU, X., HUANG, W. and XIAO, Q., 1998, *A User's Guide to the MM5 Adjoint Modeling System*, NCAR Technical Note TN-437_IA, 94 pp. (Boulder, CO: NCAR Publications).

ZOU, X. and XIAO, Q., 2000, Studies on the initialization and simulation of a mature hurricane using a variational bogus data assimilation scheme. *Journal of the Atmospheric Sciences*, **57**, pp. 836–860.

ZUPANSKI, D., 2009a, Information measures in ensemble data assimilation. In *Data Assimilation for Atmospheric, Oceanic, and Hydrologic Applications*, S.-K. Park and L. Xu (Eds.), pp. 85–95 (Berlin: Springer-Verlag).

ZUPANSKI, D., HOU, A.Y., ZHANG, S.Q., ZUPANSKI, M., KUMMEROW, C.D. and CHEUNG, S.H., 2007, Application of information theory in ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **133**, pp. 1533–1545.

ZUPANSKI, D. and ZUPANSKI, M., 2006, Model error estimation employing ensemble data assimilation approach. *Monthly Weather Review*, **134**, pp. 1337–1354.

ZUPANSKI, M., 2005, Maximum likelihood ensemble filter: theoretical aspects. *Monthly Weather Review*, **133**, pp. 1710–1726.

ZUPANSKI, M., 2009b, Theoretical and practical issues of ensemble data assimilation in weather and climate. In *Data Assimilation for Atmospheric, Oceanic, and Hydrologic Applications*, S.-K. Park and L. Xu (Eds.), pp. 67–84 (Berlin: Springer-Verlag).

ZUPANSKI, M., NAVON, I.M. and ZUPANSKI, D., 2008, The maximum likelihood ensemble filter as a non-differentiable minimization algorithm. *Quarterly Journal of the Royal Meteorological Society*, **134**, pp. 1039–1050.