RESEARCH ARTICLE

Assimilating visible satellite images for convective-scale numerical weather prediction: A case-study

Leonhard Scheck[1,2] | Martin Weissmann[2,3] | Liselotte Bach[1,4]

[1]Deutscher Wetterdienst, Offenbach, Germany

[2]Hans-Ertel Centre for Weather Research, Ludwigs-Maximilians-Universität, Munich, Germany

[3]Institut für Meteorologie und Geophysik, Universität Wien, Vienna, Austria

[4]Bundeswehr Geoinformation Centre, Euskirchen, Germany

Correspondence
L. Scheck, Meteorologisches Institut, LMU, Theresienstrasse 37, 80333 München, Germany.
Email: leonhard.scheck@lmu.de

Abstract

Satellite images in the visible spectral range contain high-resolution cloud information, but have not been assimilated directly before. This paper presents a case-study on the assimilation of visible Meteosat SEVIRI images in a convective-scale data assimilation system based on a local ensemble transform Kalman filter (LETKF) in a near-operational set-up. For this purpose, a fast look-up table-based forward operator is used to generated synthetic satellite images from the model state. Single-observation experiments show that the assimilation of visible reflectances improves cloud cover under most conditions and often reduces temperature and humidity errors. In cycled experiments for two summer days with convective precipitation, the assimilation strongly reduces the errors of cloud cover and improves the precipitation forecast. While these results are promising, several issues are identified that limit the efficacy of the assimilation process. First, the linearity assumption of the LETKF can lead to errors as reflectance is a nonlinear function of the model state. Second, errors can arise from the fact that visible reflectances alone are ambiguous and only weakly sensitive to the water phase and cloud-top height. And lastly, it is not obvious how to localise vertical covariances as visible reflectances are sensitive to clouds at all heights. For the latter reason, no vertical localisation was used in this study. To investigate the robustness of the results, the horizontal localisation scale, the assigned observation error and the spatial density of observations were varied in sensitivity experiments. The best results were obtained for an observation error close to the Desroziers estimate. High observation density combined with small localisation radii resulted in the smallest 1 hr forecast error. These settings were also beneficial for 3 hr forecasts, but forecasts at that lead time were less sensitive to the observation density and the localisation scale.

KEYWORDS

clouds, convective scale, data assimilation, ensemble Kalman filter, satellite observations, solar spectrum

1 | INTRODUCTION

Cloud-affected satellite observations provide a promising source of information for data assimilation. Clouds cover a large fraction of the Earth and frequently occur in areas that are meteorologically of particular sensitivity (McNally, 2002). Furthermore, clouds are an indicator of primary forecast features, for example fronts and cyclones on synoptic scales, as well as individual thunderstorms on convective scales.

Nevertheless, only a very small fraction of the available cloud information is assimilated in current numerical weather prediction (NWP) models. Several major NWP centres recently started to assimilated cloud-affected observations of microwave channels in global models and achieved significant forecast improvements through the incorporation of this new data source (Zhu *et al.*, 2016; Geer *et al.*, 2017). Extending this direct "all-sky" assimilation approach to infrared satellite channels is an active field of research, but is not yet implemented operationally (Okamoto, 2017; Geer *et al.*, 2018). Furthermore, most NWP centres do not yet include cloud variables in the data assimilation control vector, and thus only achieve improvements through the correction of the underlying dynamics. Satellite channels in the solar spectrum are not yet assimilated operationally in any NWP model despite providing a wealth of information on atmospheric clouds.

Regional convection-permitting models lag even further behind on the assimilation of cloud-affected satellite observations since microwave channels are not available on geostationary satellites and polar-orbiting satellites provide only very limited temporal resolution. Consequently, operational convection-permitting regional models do not yet assimilate cloud-affected satellite observations, although these data potentially provide very relevant information on the position and intensity of convective storms (Gustafsson *et al.*, 2017; Geer *et al.*, 2018).

The high spatio-temporal density of satellite observations, for example measured by the Spinning Enhanced Visible and Infrared Imager (SEVIRI) on Meteosat, is particularly promising to constrain small-scale processes acting on the convective scale (Gustafsson *et al.*, 2017). In convective-scale forecasts, low stratus and convective precipitation are among the less accurately represented processes with major impact for society. Clouds and particularly deficiencies in the model representation of low stratus dominate the uncertainty of solar power production forecasts (Kurzrock *et al.*, 2018). Heavy precipitation preceded by the development of clouds in the process of convective initiation is a major hazard for the security of the public and thus requires timely and accurate warnings. The prediction of both processes would strongly benefit from a better positioning and better constraint of clouds

and convective systems in the analysis through all-sky data assimilation.

For the discussed purpose, satellite observations in the solar part of the electromagnetic spectrum, in particular from visible and near-infrared channels, are a very promising source of information. Measuring the part of the incident sunlight that is reflected by clouds, aerosols and the Earth's surface, solar satellite channels allows us to infer high-resolution information on cloud position and microphysics, the structure and albedo of the Earth's surface and aerosol distributions. Moreover, the information in solar channels is complementary to that obtained from thermal infrared channels. While cloud-affected observations in the thermal infrared range are sensitive to both temperature and clouds, non-absorbing solar channels like the visible 0.6 μm channel are nearly exclusively sensitive to clouds. Also with respect to the cloud-top height, thermal and visible channels are complementary. While the brightness temperature measured in thermal channels provides information on the cloud-top height, visible reflectances only very weakly depend on the vertical position of the cloud. In thermal channels, low clouds are either invisible, because of absorption by water vapour or other trace gases higher up in the atmosphere, or hard to distinguish from the ground or the sea, because of the weak temperature difference between the Earth's surface and the boundary layer. In contrast, the brightness contrast between clouds and the Earth's surface makes low clouds well discernible in visible channels. The latter thus offer possibilities to retrieve information on boundary-layer clouds like low stratus, fog and cumulus clouds, which is integral to the advancement of convective-scale weather prediction for the reasons discussed above. Finally, cirrus clouds that are opaque in thermal channels are often semi-transparent in visible channels, which makes it possible to infer information on medium and low clouds below thin cirrus. All these properties indicate that visible channels can provide new and valuable information for data assimilation, with the only obvious disadvantage that their use is restricted to daytime.

Until recently, the simulation of visible satellite images required for their assimilation has been computationally too expensive for operational purposes. A novel possibility has been opened up by the development of the fast and accurate radiative transfer method MFASIS (Method for FAst Satellite Image Synthesis; Scheck *et al.*, 2016), which has recently also been implemented in the RTTOV (Radiative Transfer for TOVS) package (Saunders *et al.*, 2018). In this work, we demonstrate the feasibility of assimilating solar satellite channels in a convective-scale data assimilation framework using MFASIS. We focus on two summer days with convective precipitation and forecasts with a length of up to three hours. These restrictions limit the

computational effort and allow us to perform a number of sensitivity experiments with different assimilation settings. We assess the impact of the new observations on surface and tropospheric wind, temperature and humidity, cloud cover approximated by reflectance itself as well as precipitation. Furthermore, key challenges for the efficiency of the assimilation process, for example, nonlinearity and ambiguity of the observations, are identified and discussed by means of single-observation experiments.

The paper is organised as follows. In Section 2, we introduce the observations, the forward operator and our modelling and data assimilation framework. In Section 3, the experimental set-up and synoptic situation during the case-studies are explained and an overview of our numerical experiments is given. Sections 4 and 5 are dedicated to the discussion of the results of single-observation and cycled experiments, respectively. Conclusions are presented in Section 6.

2 | DATA AND METHODS

2.1 | Data assimilation and modelling framework

For this study, we employ the Kilometre-scale ENsemble Data Assimilation (KENDA) system (Schraff *et al.*, 2016) developed at Deutscher Wetterdienst (DWD). To compute the first-guess and forecast ensembles, we employ version 5.2 of the limited-area, non-hydrostatic NWP model COSMO (Consortium for Small-scale MOdeling) in the COSMO-DE set-up. This configuration has been operational at DWD until May 2018 (Baldauf *et al.*, 2011), with a grid spacing of 2.8 km and a domain covering Germany and its neighbouring countries (black rectangle in Figure 1). The numerical grid consists of 421×461 columns and uses 50 hybrid layers that are terrain-following in the lower levels and flat in the upper levels. The grid extends up to 22 km. Deep convection is represented explicitly by the model whereas shallow convection is parametrized by means of a simplified version of the Tiedtke scheme (Tiedtke, 1989). A Lin-type (Lin *et al.*, 1983) one-moment bulk cloud microphysics scheme including cloud water, cloud ice, rain, snow and graupel is used that contains a simplified version of the parametrization of Seifert and Beheng (2001) for auto-conversion, accretion and self-collection. The boundary conditions for the COSMO-DE model runs are taken from ICON-EU model runs with parametrized convection and a grid resolution of 7 km.

The Local Ensemble Transform Kalman Filter (LETKF) used in this study is a square-root filter following Hunt *et al.* (2007). Its advantage is the flow-dependent update of the sample background-error covariance matrix as well as its computational efficiency conducting the analysis in the low-dimensional ensemble space (transform) and grid point by grid point (local) on a reduced analysis grid. In a first step, the analysis mean is computed by an optimal linear combination of the first-guess ensemble members. In a second step, the analysis ensemble is spanned such that its spread reflects the estimated analysis uncertainty.

The LETKF cost function is given by

$$
\begin{aligned}
J(\mathbf{w}) = (k-1)(\mathbf{w}^{\mathrm{T}}\mathbf{w}) + (\mathbf{y}^{\mathrm{o}} - \overline{\mathbf{y}}^{\mathrm{b}} - \mathbf{Y}^{\mathrm{b}}\mathbf{w})^{\mathrm{T}} \\
\times \mathbf{R}^{-1}(\mathbf{y}^{\mathrm{o}} - \overline{\mathbf{y}}^{\mathrm{b}} - \mathbf{Y}^{\mathrm{b}}\mathbf{w}),
\end{aligned} \tag{1}
$$

where the weight vector $\mathbf{w}$ is of dimension k of the ensemble space, and $\mathbf{R}$ is the observation-error covariance matrix. In this formulation of the cost function, a linear relationship between the first-guess ensemble in model space and its transformation to observation space is assumed:

$$
H(\overline{\mathbf{x}}^{\mathrm{b}} + \mathbf{X}^{\mathrm{b}}\mathbf{w}) \approx \overline{\mathbf{y}}^{\mathrm{b}} + \mathbf{Y}^{\mathrm{b}}\mathbf{w}. \tag{2}
$$

In Equation (2) (corresponding to equation 18 in Hunt *et al.* 2007), $\overline{\mathbf{x}}^{\mathrm{b}}$ is the first-guess ensemble mean and $\mathbf{X}^{\mathrm{b}}$ the background ensemble perturbation matrix. Column i of $\mathbf{X}^{\mathrm{b}}$ is given by $\mathbf{x}^{\mathrm{b}(i)} - \overline{\mathbf{x}}^{\mathrm{b}}$, the deviation of the state vector of background member i from the mean state. $\mathbf{Y}^{\mathrm{b}}$ is the background ensemble perturbation matrix in observation space with columns $\mathbf{y}^{\mathrm{b}(i)} - \overline{\mathbf{y}}^{\mathrm{b}}$ and the model equivalents $\mathbf{y}^{\mathrm{b}(i)} = H(\mathbf{x}^{\mathrm{b}(i)})$ are obtained by means of the nonlinear forward operator H. The analysis ensemble members in model space are computed as

$$
\overline{\mathbf{x}}^{\mathrm{a}(i)} = \overline{\mathbf{x}}^{\mathrm{b}} + \mathbf{X}^{\mathrm{b}}\overline{\mathbf{w}}^{\mathrm{a}(i)}, \tag{3}
$$

where $\overline{\mathbf{w}}^{\mathrm{a}(i)}$ is the weight vector minimising Equation (1).

The linearity assumption (Equation (2)) may not always be valid, in particular for nonlinear forward operators sensitive to clouds and precipitation processes such as the MFASIS operator applied in this work. We will discuss the effects of this linearisation on the assimilation of solar satellite channels in Section 4.

In the operational KENDA system, conventional observations from SYNOP stations, radiosondes, wind profilers, drifting buoys and aircraft (AMDAR, Aircraft Meteorological DAta Relay, and MODE-S) are assimilated. In this study, all these observations are assimilated, except for MODE-S data and drifting buoys. A second difference to the operational KENDA set-up is that we do not employ latent heat nudging, which pulls the model towards radar-derived precipitation rates during the first-guess integrations.

FIGURE 1 Location of observation locations between used in the data assimilation experiments between 1000 and 1100 UTC on 5 June 2016. The numbers of actively assimilated observations of the observation types SYNOP, AIREP (aircraft), PILOT (wind profilers), RAD (satellite images) and TEMP (radiosondes) are provided in parentheses. For the corresponding experiment, the satellite image has been thinned such that the distance between two neighbouring observations is about 36 km. The black rectangle indicates the COSMO-DE domain. For the data assimilation experiments, only observations within the red box are considered and thus the Alps and regions close to the model domain boundaries are excluded
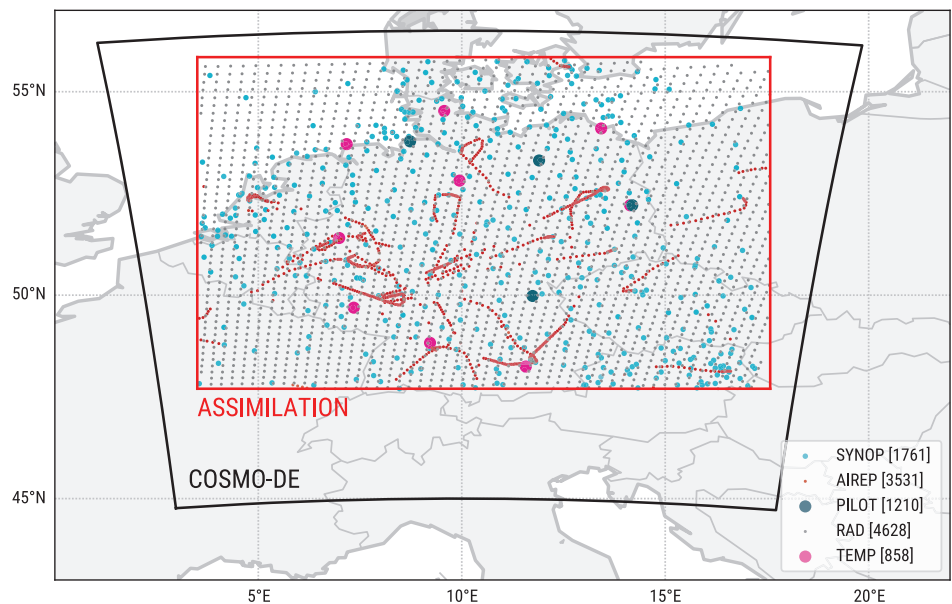

In our near-operational set-up, about 5,000–6,000 conventional observations are available per hour in the assimilation and evaluation region that extends from 47.7° to 56.0°N and from 3.5° to 17.5°E (red box in Figure 1). The assimilation region was chosen to be somewhat smaller than the COSMO-DE model domain and excludes the Alps and areas close to the domain boundary to avoid nesting effects and the misinterpretation of snow as clouds. As discussed in Schraff *et al.* (2016), observations are only assimilated up to a height of 300 hPa to avoid problems in the Raleigh dampening layer on the top of the model column. Since the flight level of aircrafts is typically above 300 hPa, the aircraft report (AIREP) observations displayed in Figure 1 are mainly from ascent and descent flight phases.

2.2 | Satellite observations and forward operator

We assimilate satellite images measured by the SEVIRI instrument onboard the satellite Meteosat-10 of the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) at a wavelength of 0.6 μm in the visible part of the solar spectrum. In contrast to the channels in the near-infrared and infrared spectral range, the 0.6 μm channel is not affected significantly by absorption. Moreover, the contribution of clear-sky Rayleigh scattering to the reflectance is rather small. Thus, the cloud-top height has only a very weak influence on the observed reflectance.

To generate synthetic 0.6 μm SEVIRI images, the model state is transformed to radiative transfer (RT) input parameters as in Scheck *et al.* (2018) and the MFASIS RT method (Scheck *et al.*, 2016) is used to compute reflectances. The RT solver requires profiles of extinction coefficients and effective particle radii for water and ice clouds and cloud cover. The extinction coefficients and effective radii are computed following Kostka *et al.* (2014). Effective droplet radii are calculated adopting the approach of Martin *et al.* (1994) and the parametrization of Wyser (1998) is used for effective ice particle sizes. Water and ice contents are converted to extinction coefficients using the parametrizations of Hu and Stamnes (1993) and Fu (1996), respectively. In contrast to Scheck *et al.* (2016) and Kostka *et al.* (2014), assumptions on the cloud fraction and the overlap of subgrid clouds are taken into account. In the COSMO model, it is assumed that subgrid clouds exist in cells where the relative humidity exceeds a height-dependent critical value (Quaas, 2012 gives an overiew of this approach). We used subgrid cloud water and ice contents and cloud fractions that are consistent with the assumptions used in the model. Consistent with COSMO's internal RT code for computing heating rates, we use the stochastic 3D maximum-random overlap method from Scheck *et al.* (2018). The integrated optical depths and mean effective radii for water and ice clouds obtained by the method are converted to reflectances using MFASIS with the 21MB look-up table described in Scheck *et al.* (2016). Aerosols are not yet taken into account in the operator. However, except for special events, like Sahara dust outbreaks, their impact on visible reflectances
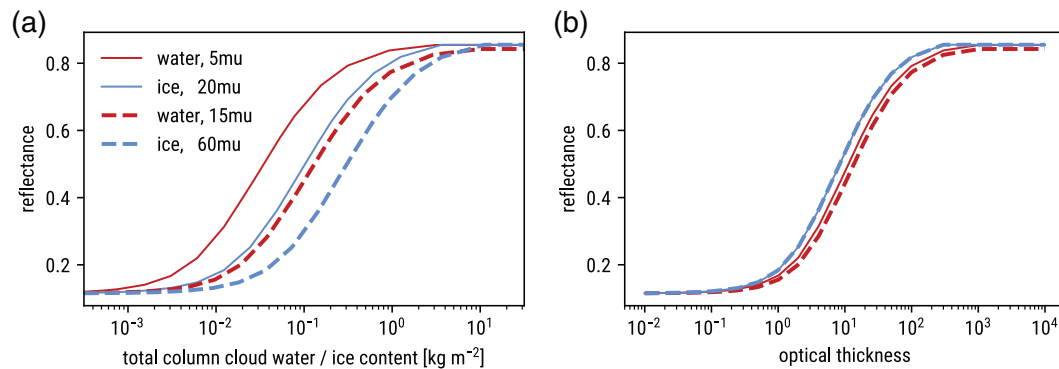
FIGURE 2 Reflectance in the 0.6 μm SEVIRI channel as a function of (a) total column water content and (b) optical depth for a water cloud and an ice cloud for different effective particle radii. The albedo was set to 0.1, the solar zenith angle to 30°, the satellite zenith angle to 60° and the scattering angle to 135°

should be much smaller than the cloud signal. The relative reflectance error of MFASIS with respect to the discrete ordinate method (DISORT; Stamnes *et al.*, 1988) has been found to be less than 2% (Scheck *et al.*, 2016). For a number of high-resolution scenes, the RMS reflectance errors caused by 3D RT effects have been shown to be in the range 0.05–0.2 (Scheck *et al.*, 2018). The agreement of the synthetic reflectance distribution based on operational COSMO-DE forecasts for May and June 2016 with the observed distribution was found to be quite good for not too extreme solar zenith angles in Scheck *et al.* (2018). Therefore, we decided not to perform any bias correction for the visible reflectances.

2.3 | Sensitivity of visible reflectances to the model state

For the interpretation of assimilation experiments, it is important to clarify which model state variables have an influence on the visible reflectances generated by the forward operator. In Figure 2, reflectances for water and ice clouds are shown as functions of the optical depths and the column-integrated water contents. It is obvious that the reflectance, R, is a highly nonlinear function of the optical depth, τ. $R(\tau)$ strongly increases for values between 1 and 100 and saturates for larger optical depths.

Due to differences in the scattering phase function of spherical water droplets and ice particles with complex shapes, the reflectance of water and ice clouds can differ significantly for a given viewing geometry, in particular for special cases like fog bows and halos. However, averaged over all viewing angles, water and ice clouds with the same optical depth have similar reflectances. This is also the case for the viewing geometry of the example in Figure 2. However, ice particles are in general much larger

than cloud droplets. For the same water content, a larger mean particle radius corresponds to a lower particle concentration and, as the increased scattering cross-section cannot fully compensate for this effect, a lower probability of scattering. Consequently, the reflectance of an ice cloud is generally lower than that of a water cloud with the same water content (Figure 2a). However, water clouds cannot be distinguished from ice clouds based on visible reflectance only.

The curves displayed in Figure 2 represent the reflectance for single cloud layers with cloud fraction 1. The sensitivity to cloud water and ice may differ significantly if several cloud layers are present and when the clouds only fill a part of the atmospheric column (i.e., if the column contains grid boxes with cloud fractions between zero and one). The reflectance generally depends on all water and ice cloud layers, but its sensitivity to the water or ice content in specific layers can be much smaller than in the single-layer case shown in Figure 2. For instance, the liquid water content (LWC) of a water cloud under a thick ice cloud has only a very weak impact on the reflectance. The presence of the ice cloud will thus reduce $\partial R/\partial \text{LWC}$. As a second example, the ice water content (IWC) of a thin ice cloud has an impact on the reflectance when a thick water cloud is present.

The variables LWC and IWC used as input variables for the RT do not only contain grid-scale cloud water, but also a contribution from parametrized subgrid clouds. For cells with a relative humidity higher than a height-dependent threshold value (in COSMO typically 70–95%), it is assumed that subgrid clouds are present (Schomburg *et al.*, 2015) and contribute to the LWC. Thus there is also a sensitivity of the reflectance to humidity for grid cells close to saturation. Moreover, the reflectance also depends on the local gradient of the cloud-top height, an effect that is taken into account by means of the cloud-top

inclination correction of Scheck *et al.* (2018). In contrast to infrared radiances, the cloud-top height itself has no significant influence on the 0.6 μm reflectance (discussion in Scheck *et al.*, 2016).

2.4 | Superobbing and thinning

In the COSMO-KENDA system, a diagonal observation-error covariance matrix is employed. To alleviate problems with correlated observation errors, it can be helpful to reduce the spatial density of the observations. One of the standard approaches is to thin the observations, which means to remove satellite pixels in regular spatial intervals such that the distance between neighbouring observation locations is increased. For the satellite images considered here, the thinning process keeps every nth pixel in the zonal direction and only every $2n$th pixel in the meridional direction. Due to the broadening of the radiation beam in the meridional direction, as one satellite pixel of the SEVIRI instrument covers a nearly rectangular surface area with dimensions of about 6×3 km in the COSMO-DE domain, this leads to a regular grid of satellite observations that is nearly equidistant with a thinning length scale of $l_{\text{thin}} = n \times 6$ km.

Superobbing, i.e. averaging the reflectance for every block of $k \times m$ pixels, is a further approach to data reduction. However, since all pixels are included in the averaging process, correlated observation errors cannot be removed as efficiently. Rather, the reason for superobbing is to reduce the representativeness error by drawing the spatial scale represented by the observations towards the effective resolution of the model. To achieve a nearly equidistant grid of observations, we assume $k = 2 \times m$ and define a superobbing length-scale as $l_{\text{so}} = m \times 6$ km.

Superobbing and thinning can also be combined. In this case, the superobbing is performed first and then only every f_{thin}th superobservation is kept in the zonal and meridional directions, which results in a horizontal observation density of

$$\rho^{\text{sat}} = (l_{\text{so}}f_{\text{thin}})^{-2}. \tag{4}$$

To keep this definition meaningful when only thinning is used, we assume $l_{\text{so}} = 6$ km and $f_{\text{thin}} = n$ in this case.

Both superobbing and thinning are suitable methods to reduce the number of observations. Beyond data reduction, reduction of representativeness error and accounting for the assumption of uncorrelated observation errors, this is useful to balance the number of conventional and satellite observations. Schraff *et al.* (2016) define the effective number of observations as the sum of localisation-weighted observations

that influence a grid point in the analysis. The adaptive localisation method for conventional observations in KENDA tries to keep this number constant by locally adjusting the localisation radius to a value between 50 and 100 km.

For observations with a constant spatial density ρ, like the thinned and/or superobbed satellite images considered here, the localisation-weighted number of observations within the cut-off radius $r_{\text{h}} = 2\sqrt{10/3}\, l_{\text{h}}$ (here l_{h} is the localisation scale) can be shown to be

$$\int_0^{r_{\text{h}}} G(r)\, \rho\, 2\pi\, \mathrm{d}r \approx 2\pi\rho\, l_{\text{h}}^2,$$

where G is the Gaspari–Cohn function (Gaspari and Cohn, 1999), which is zero for $r > r_{\text{h}}$. For satellite observations with $\rho = \rho^{\text{sat}}$ the effective number of observations influencing each grid point in the analysis is $N_{\text{gp}}^{\text{sat}} = 2\pi l_{\text{h}}^2 (l_{\text{so}} f_{\text{thin}})^{-2}$.

3 | EXPERIMENTAL SET-UP

3.1 | Synoptic situation

The data assimilation experiments were conducted for two days, namely 29 May and 5 June 2016, from a highly convective period that has been studied extensively (e.g., Necker *et al.* 2018; 2020; Keil *et al.* 2019; Bachmann *et al.* 2020). From 26 May 26 to 9 June 2016, a high number of consecutive severe thunderstorms, heavy rainfall, flash floods, hail and tornados determined the weather situation in Germany and caused losses on the order of billions of Euros. According to Piper *et al.* (2016), the fundamental trigger for these events was a blocking situation due to a large-scale ridge extending from Great Britain over Iceland to central Scandinavia which blocked the exchange of moist and unstable air masses over Europe.

Due to the weak pressure gradient that accompanied the blocking situation, there was basically no advective forcing on 5 June so that the thunderstorms were mostly stationary and caused extremely heavy rainfall and hail with precipitation totals of more than 100 mm within few hours in some locations. In contrast, significantly stronger forcing influenced convection on 29 May so that the convective cells had a much shorter lifetime. For the assimilation of SEVIRI-VIS observations, the two days are particularly interesting as they provide very diverse large- and small-scale convective clouds at the state of convective initiation as well as fully developed cumulonimbus clouds with large anvils.

3.2 | Set-up and metrics for single-observation experiments

A number of single-observation experiments were conducted to investigate the impact of assimilating VIS observations. The observations were selected such that they are spatio-temporally close to radiosonde observations. In total, 35 radiosondes were launched at 1045 UTC and 1645 UTC on 29 May and 5 June 2016 within the assimilation region shown in Figure 1. The distance between the radiosonde stations is sufficiently large (Figure 1) such that setting the horizontal localisation cut-off radius to 90 km ensures that every analysis grid point is influenced by only one VIS observation. Thereby, multiple single-observation experiments can be conducted simultaneously in one LETKF cycle (as in Schomburg *et al.*, 2015). First-guess forecasts were initialised by a model warm start at 1000 UTC and 1600 UTC on 29 May and 5 June from a reference cycling experiment in which only conventional observations were assimilated. Analyses were computed for 1100 UTC and 1700 UTC using selected VIS observations only. While inflation methods were used in the reference experiment from which the single-observation experiment was started (see next section), they were switched off for the single-observation analysis step. The increments generated in the analysis are thus caused only by the assimilation of the observation. For the sake of simplicity we assimilated individual satellite pixels and not superobservations. An observation error of 0.1 was used.

To assess the error of the model state with respect to radiosonde observations, we define mean profile errors as

$$\epsilon_X^v = \frac{1}{p_{\text{sfc}} - p_{\text{min}}} \int_{p_{\text{sfc}}}^{p_{\text{min}}} |X_O - X_v| \, dp, \qquad (5)$$

where X is one of the variables T, RH, U and V and $v \in$ A,B indicates whether the deviation from the observation X_O is computed for the first-guess (B) or the analysis (A) model state. p_{min} was set to 200 hPa. To quantify the change in the mean profile errors due to data assimilation, we define $\Delta\epsilon_X = \epsilon_X^A - \epsilon_X^B$, for which negative values indicate a reduction of the error.

3.3 | Standard set-up for cycled experiments

For cycled assimilation experiments, we employed near-operational settings. As in Schraff *et al.* (2016), we used 40 ensemble members and an assimilation window of 1 hr. To control spread, both additive and adaptive multiplicative inflation were used together with relaxation to prior perturbations (RTPP; Zhang *et al.*, 2004), for which the α parameter was set to 0.75, as in Harnisch and Keil (2015). The scaling factor for the additive inflation (parameter α_a in Zeng *et al.*, 2018) was set to a value of 0.2, which improved results slightly compared to the operational value of 0.1. For conventional observations, we used the same fixed vertical localisation and adaptive horizontal localisation as in Schraff *et al.* (2016). The adaptive localisation aims at restricting the effective number of observations that influence a grid point to a given value, in this case a number of 100. To achieve this goal, the localisation scale is allowed to vary between 50 and 100 km. The LETKF state vector in our experiments comprises the prognostic variables temperature, specific humidity, surface pressure, horizontal and vertical velocity components, cloud water and cloud ice.

In the standard configuration, we apply superobbing (no thinning) with a length-scale of $l_{so} = 18$ km. The superobbed image resolution is thus somewhat coarser than the effective model resolution, which Bierdel *et al.* (2012) estimated to be about five times the grid spacing of 2.8 km. Within each 1-hr data assimilation window, one visible satellite image is assimilated assuming a constant observation error of 0.2. The horizontal localisation scale for the superobbed reflectance observations is set to a constant value of 50 km such that the effective number of satellite observations that influence a grid cell in the analysis (as defined in Section 2.4) is $N_{gp}^{sat} \approx 50$. The motivation for this choice was to give satellite observations a significant weight relative to the conventional observations without overwhelming the influence of the latter. As visible reflectances are sensitive to clouds at all heights, there is no obvious way to perform a vertical localisation. Therefore, vertical localisation is not used for these observations in our experiments. Furthermore, no bias correction was used in the experiments given the short experimental period.

In the following, the data assimilation experiment using only conventional observations is denoted CONV. The experiment using both conventional and visible satellite observations is denoted VISCONV. Both experiments were cycled throughout the two days described in Section 3.1. The CONV experiment was initialised at 2100 UTC on the previous day and extended until 1800 UTC. The VISCONV experiment was branched off the CONV experiment at 0500 UTC. Thus, we assimilate satellite images in each cycle between 0600 and 1800 UTC. For both 29 May and 5 June, forecasts with a lead time of 3 hr are initialised at 0600, 0900, 1200 and 1500 UTC.

TABLE 1 Overview of the assimilation settings of all experiments described in Section 3.4

Experiment	σ_o	l_h (km)	l_{so} (km)	l_{th} (km)	N_{gp}^{sat}	N_{tot}^{sat} ($10^3 hr^{-1}$)
S6-1	**0.15**	17	6	—	50.4	166.6
S6-2	**0.20**	17	6	—	50.4	166.6
S12-1	**0.15**	35	12	—	53.5	41.7
S12-2	**0.20**	35	12	—	53.5	41.7
S18-1	**0.30**	50	18	—	48.5	18.5
S18-2	**0.20**	50	18	—	48.5	18.5
S18-3	**0.15**	50	18	—	48.5	18.5
S18-4	**0.10**	50	18	—	48.5	18.5
S36-1	**0.20**	100	36	—	48.5	4.6
S36-2	**0.15**	100	36	—	48.5	4.6
S36-3	**0.10**	100	36	—	48.5	4.6
S18T2-1	**0.15**	100	18	36	48.5	4.6
S18T2-2	**0.10**	100	18	36	48.5	4.6
T12-1	0.15	35	—	12	53.5	41.7
T12-2	0.20	35	—	12	53.5	41.7
T18-1	**0.20**	50	—	18	48.5	18.5
T18-2	**0.15**	50	—	18	48.5	18.5
T18-3	**0.10**	50	—	18	48.5	18.5
HL-1	0.20	**70**	18	—	**95.0**	18.5
HL-2	0.20	**50**	18	—	**48.5**	18.5
HL-3	0.20	**35**	18	—	**23.8**	18.5

Note: For each case, the assigned observation error σ_o, the horizontal localisation scale l_h, the superobbing scale l_{so}, the thinning scale l_{th}, the number of satellite observations influencing each grid cell, N_{gp}^{sat}, and the total number of satellite observations assimilated every hour, N_{tot}^{sat} are specified. Section 2.4 gives exact definitions. The experiments S18-2 and HL-2 are identical to the VISCONV standard experiment. Bold type denotes variation in one or two parameters within each group.

## 3.4 | Sensitivity experiments with modified parameters

In addition to the cycled experiments with the standard configuration, a number of experiments with modified assimilation settings was performed to assess the impact of tunable parameters. An overview of the experiments and the assimilation settings is given in Table 1. Eight groups of experiments were conducted and within each group one or two parameters were varied (indicated by bold type in the table).

In the groups S6, S12, S18, S36 the superobbing scale was set to 6, 12, 18 and 36 km, respectively. Within the group, the assigned observation error was varied. Groups T12 and T18 are similar to S12 and S18, but thinning was used instead of superobbing and the number in the identifier refers to the thinning length-scale. In group S18T2, superobbing to 18 km was combined with thinning by a factor of $f_{thin} = 2$ in the zonal and meridional directions (Section 2.4). In the described groups, the total number of satellite observations N_{tot}^{sat} assimilated per 1-hourly cycle varies between 4,600 and 160,000. However, the localisation radius for the satellite observations is adjusted such that $N_{gp}^{sat} \approx 50$ satellite observations influence each grid cell in the analysis in all of the experiments. In contrast, the localisation radius is varied without changing N_{tot}^{sat} in group HL. Therefore, N_{gp}^{sat} varies and more (HL-1) or less (HL-3) weight is given to the satellite observations than in the standard configuration.

Note that the experiments S18-2 and HL-2 are identical to the VISCONV experiment carried out with the standard assimilation settings. As discussed in Section 3.3, no vertical localisation is used for the reflectance observations.
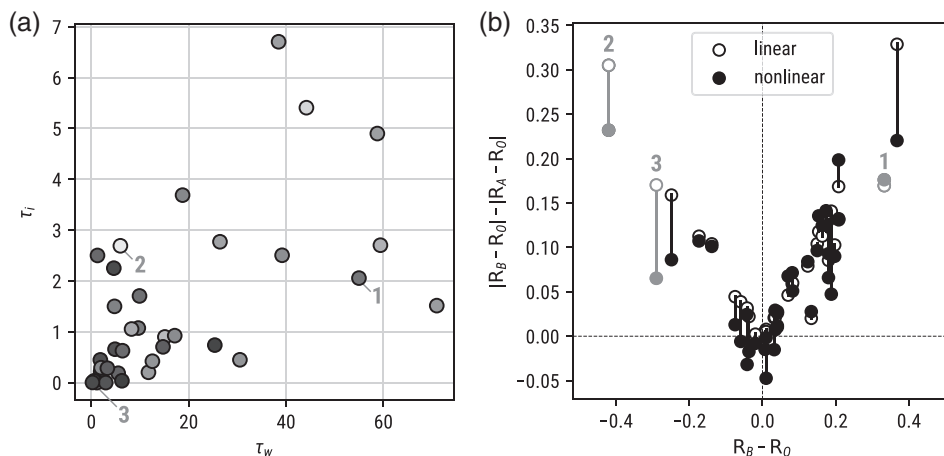
FIGURE 3 (a) First-guess effective ice (τ_i) and water (τ_w) optical depths for the 35 single-observation cases. The brightness of the dots indicates the observed reflectance (the brighter the higher the reflectance). (b) Change in the absolute value of the reflectance departures as a function of the first-guess departure as estimated by the LETKF (open circles) and recomputed using the full nonlinear operator (filled circles). Positive values indicate an error reduction. Cases discussed in more detail in the text are marked with numbers

4 | RESULTS I: SINGLE-OBSERVATION EXPERIMENTS

4.1 | Impact on reflectance

Figure 3a provides an overview of the pairs of effective optical depths of water and ice clouds in the 35 different single-observation experiments. Among these cases are pure water, pure ice and mixed pairs with ice optical depths up to 7 and water optical depths up to 70.

To discuss the impact of the data assimilation on error reduction, we use the following definitions. The mean reflectance of the background members is denoted by

$$R_B = \frac{1}{k}\sum_{i=1}^{k} H(\mathbf{x}^{b(i)}), \qquad (6)$$

the mean reflectance of the analysis members (as estimated by the LETKF) by

$$R_A = \frac{1}{k}\sum_{i=1}^{k} \overline{\mathbf{y}}^b + \mathbf{Y}^b w^{a(i)}, \qquad (7)$$

and the observed reflectances by R_O, where all other variables are defined in Section 2.1, just as in Hunt *et al.* (2007). While Equation (6) involves applying the full nonlinear operator H to the background ensemble members, the latter is not applied in Equation (7). Instead of using H, the reflectances of the analysis members are estimated from the reflectances of the background members using Equation (2). Since, as discussed in Section 2.3, the relation between model state and visible reflectance is nonlinear, we assess the effect of the linearity assumption. For that

purpose, we apply the full nonlinear operator to a 50 s forecast started from the analysis state. It is not possible to apply the operator directly to the analysis state, since the diagnostic variables (e.g., the ones related to the shallow convection, which contributes to the subgrid clouds) are not part of the analysis. However, by applying the operator to a very short forecast, the left side of Equation (2) can be approximated. The obtained reflectance values allow for the computation of the actual mean reflectances of the analysis members,

$$R_A^\star = \frac{1}{k}\sum_{i=1}^{k} H(\mathbf{x}^{a(i)}) = \frac{1}{k}\sum_{i=1}^{k} H(\overline{\mathbf{x}}^b + \mathbf{X}^b w^{a(i)}). \qquad (8)$$

If a single observation is assimilated, the deviation of the mean analysis reflectance from the observation, $|R_A - R_O|$, should be smaller than the one of the background, $|R_B - R_O|$. As shown in Figure 3b, it is $|R_B - R_O| - |R_A - R_O| > 0$ for all experiments (open circles). For larger first-guess departures, a stronger reduction of the error is achieved. However, when the change in the reflectance departure is calculated using analysis reflectances computed with the full nonlinear operator (filled circles in Figure 3b), it becomes evident that in many cases the departure reduction as estimated by the LETKF, $|R_B - R_O| - |R_A - R_O|$, differs from the actual value, $|R_B - R_O| - |R_A^\star - R_O|$. In most cases, the model equivalents have not been drawn as close to the observations as estimated by the LETKF. The nonlinearity of the operator thus limits the effectiveness of the VIS assimilation. However, only for cases with small first-guess departures, the nonlinearity of the operator leads to analysis reflectance errors that are larger than the first-guess errors ($|R_B - R_O| - |R_A^\star - R_O| < 0$). In any other case, the analysis is closer to the observations than the first guess.
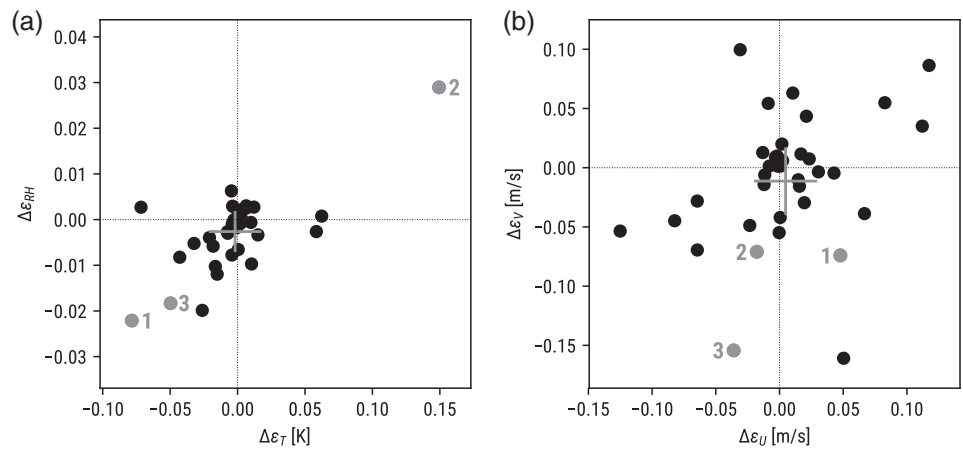
4.2 | Impact on model variables

Figure 4 shows the impact on tropospheric temperature, humidity and the horizontal wind components as verified against radiosondes. For nearly all cases, the impact on relative humidity is either beneficial or neutral (Figure 4a). The only case with increased error (the dot in the upper-right corner) will be discussed below. Results for temperature errors are similar. In most cases the error is reduced and there are only three cases with an error increase of more than 0.05 K. The evaluation of winds overall does not show a systematic effect (Figure 4b).

4.3 | Selected cases

The cases marked with digits 1–3 in Figures 3 and 4 are now discussed further.

4.3.1 | Case 1

In the first case, the first-guess ensemble is dominated by a thick water cloud below a thin ice cloud and has a mean reflectance of 0.7 (Figure 5a). The observed reflectance is only 0.35 and related to a low-level cloud indicated by the radiosonde humidity profile shown in Figure 6e. In the LETKF analysis, the water content and the cloud fraction of the water cloud are decreased (Figure 6a,c). The reflectance is reduced to about 0.5. The ice content remains unchanged and the cloud fraction is even slightly increased at the ice cloud (Figure 6b). There are two reasons for the ice cloud remaining nearly unchanged. First, in this situation with a thin ice cloud and a much thicker water cloud, the derivative of reflectance with respect to QI should be smaller than the one with respect to QC,

which means that changes in the water cloud are a more efficient way for the LETKF to change the reflectance. Second, the spread in QI is smaller than in QC (not shown). The linearisation used in the LETKF does not lead to a significant error in the ensemble mean reflectance in this case, but causes the reflectance spread to be increased instead of being decreased (compare A and $A^\star$ ensembles in Figure 5a). After the assimilation, the relative humidity and temperature profiles better agree with the ones observed by the radiosonde (Figure 6d,e). The removal of clouds leads to a stronger direct insolation of the surface and higher low-level temperatures.

Even though the observation lies outside the background ensemble (Figure 5a), the LETKF manages to improve reflectance, humidity and temperature in this case. The cloud variables are improved in the sense that a too dense low- to mid-level water cloud is mostly removed. However, the remaining water cloud in the analysis extends further up than the radiosonde profile suggests. This problem results from the fact that cloud-top height is not constrained by the visible reflectance observations (Section 2.3).

4.3.2 | Case 2

In the second case, a thick anvil ice cloud with a reflectance of 0.9 was present at the location of the radiosonde. The observed 0.6 μm reflectance alone only indicates that a optically thick cloud was present, but the significantly lower reflectance in the corresponding 1.6 μm image (not shown) can only be explained by an ice cloud (Wolters *et al.*, 2008). In contrast, the ensemble only contains much thinner ice clouds and mid-level water clouds (Figure 7a–c) with a mean reflectance of 0.5 (Figure 5b). Neither the water cloud nor the ice cloud dominate the
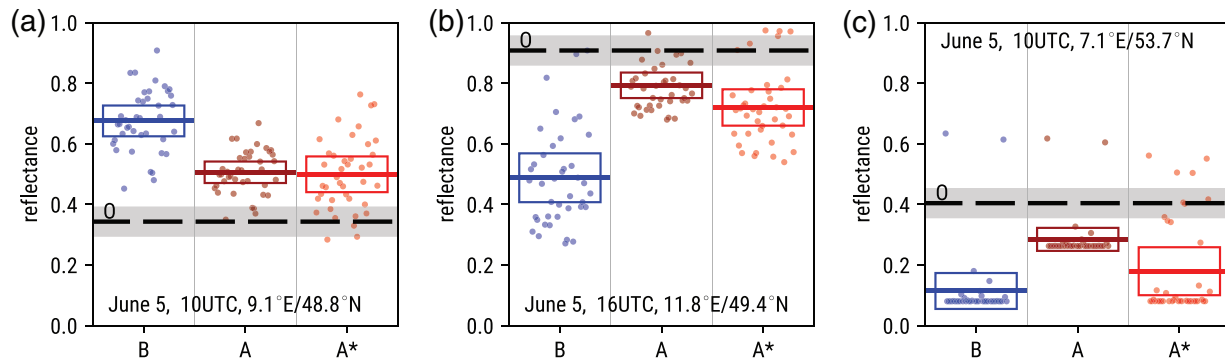
FIGURE 5 Reflectance distributions for the single-observation cases (a) 1, (b) 2 and (c) 3 discussed in the text and marked in Figures 3 and 4. For each case the observed reflectance is given by the dashed black horizontal line and the height of the surrounding grey bar denotes the assigned observation error. The dots indicate the reflectance of the individual ensemble members. The thick solid horizontal lines mark the ensemble mean reflectance and the height of the rectangles denotes the ensemble spread. Three reflectance ensembles are shown for each case, the first-guess ensemble (B), the analysis ensemble computed with the linearisation (A) and with the full nonlinear operator (A⋆). The horizontal position of the dots indicates the index of the ensemble member
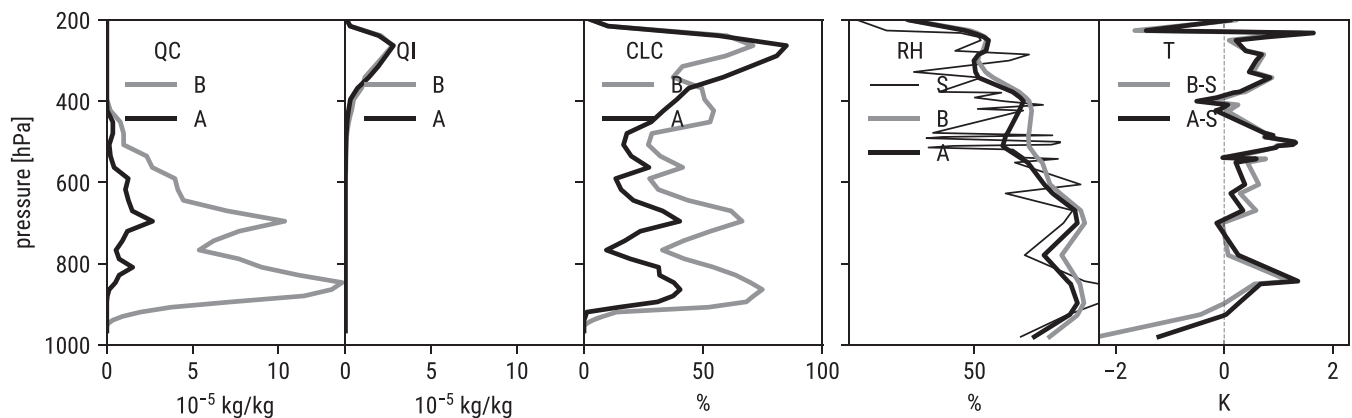
FIGURE 6 Vertical profiles of background (B) and analysis (A) mean variables at the release location of the radiosonde for case 1. The panels show (from left to right) cloud water mixing ratio *QC*, cloud ice mixing ratio *QI*, cloud fraction *CLC*, relative humidity *RH*, and temperature *T*. The thin black line in the *RH* panel is the *RH* observed by the radiosonde. The *T* panel shows the first-guess departure of the ensemble mean for temperature (first-guess minus observation)

mean reflectance of the background ensemble. Consequently, both the ice clouds and the mid-level water clouds are enhanced in the LETKF analysis – the ice cloud mainly by increasing the mixing ratio, the water cloud more by increasing the cloud fraction (Figure 7a–c). These changes cause the reflectance to increase to 0.8. However, the radiosonde profiles show that the humidity had already been too high at mid- and lower levels and this error is increased further in the analysis (Figure 7d). Moreover, the model temperature profile at lower levels is pushed further away from the radiosonde profile (Figure 7e). In this case, the linearisation in the LETKF causes not only an underestimation of the reflectance spread in the analysis, as in case 1, but also leads to a mean analysis reflectance that is not as close to the observed value as indicated by the linear estimate (compare *A* and *A*⋆ ensembles in Figure 5b).

This case is an example of the challenges related to the ambiguity of visible reflectances. Both changes in the water and the ice clouds could reduce the reflectance departure and thus both are modified in the analysis. From the radiosonde profiles it seems unlikely that a dense water cloud was present below the ice cloud. Assimilating additional satellite observations like the 1.6μm channel (in which ice clouds are darker than water clouds) or thermal infrared channels could reduce this ambiguity. Moreover, vertical localisation could have been useful in this case to avoid increments in layers far below the top of the dense ice cloud. Since visible reflectance cannot provide direct information about these layers, the increments probably result from spurious vertical correlations in the absence of vertical localisation.
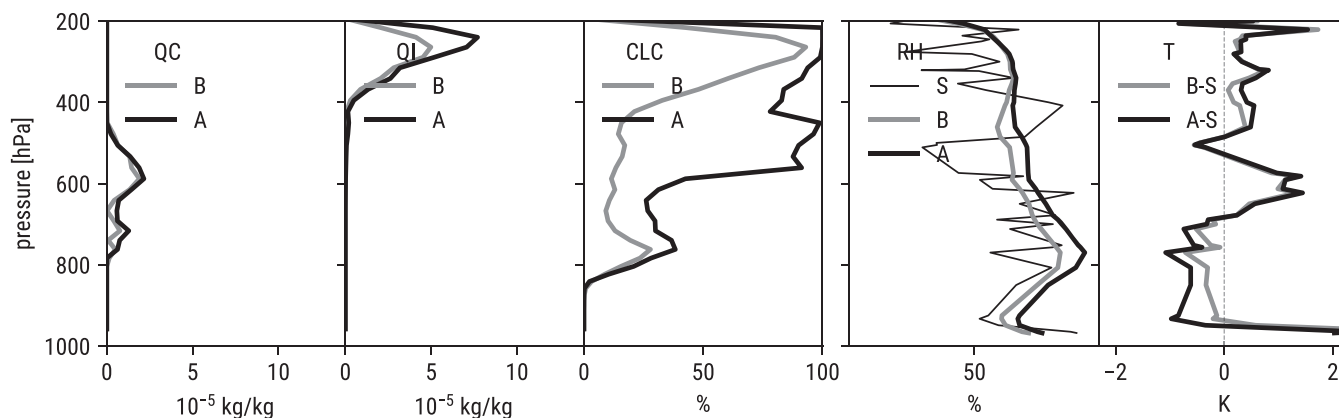
FIGURE 7 As Figure 6, but for case 2

4.3.3 | Case 3

In the third case, an observed low cloud over the ocean is missing in almost all first-guess ensemble members. Only for two of the members are the reflectances considerably higher than the observed reflectance of 0.4, while all other members exhibit low reflectances around 0.1 (Figure 5c). In the analysis, the near-surface cloud is optically thicker and the reflectance is increased. As only the lowermost model levels contain clouds, we do not show vertical profiles for this case.

In this case the LETKF manages to increase reflectance (and even to reduce both humidity and temperature errors with respect to the radiosonde profiles; Figure 4), but this process is hampered by the nonlinear relation between model state and reflectance. While the linear estimates for the analysis reflectances indicate that about two thirds of the error in the mean reflectance have been removed, the actual mean analysis reflectance is only improved slightly (Figure 5c).

This example is thus characterised by a large difference between the linear estimate and the real analysis reflectance. This problem is related to the fact that all ensemble members are so far from the observation that the LETKF has no reliable information from the background ensemble about which change in the cloud water content would be required to correct the reflectance error.

To conclude, the single-observation experiments show that the assimilation of visible reflectances is able to reduce errors in the model state in most of the cases. However, the ambiguity of the observations, spurious correlations and the nonlinearity of the operator can limit the effectiveness of this process.

5 | RESULTS II: CYCLED EXPERIMENTS

This section shows the results of the assimilation of visible reflectances in a near-operational cycling set-up for the two selected days described in Section 3.1. First, we evaluate the impact of assimilating visible reflectances for a standard assimilation set-up (Section 3.3) and then the sensitivity of the results to variations of assimilation parameters (Section 3.4).

5.1 | Impact of reflectance assimilation for a standard set-up

To evaluate the impact of VIS observations on the accuracy of cloud cover and precipitation forecasts, we compute the fraction of ensemble members exceeding a reflectance value of 0.5 for each pixel of the satellite image. This can be interpreted as the probability of the presence of a relatively dense cloud. As an example, the probability of cloudiness for the 1-hr forecast valid at 1000 UTC on 5 June is shown in Figure 8a for the CONV experiment assimilating only conventional observations and in Figure 8b for the VISCONV experiment additionally assimilating satellite observations. Figure 8 shows clear discrepancies between the bright areas, where most of the ensemble members show reflectances larger than 0.5, and the blue contours indicate where the reflectance of the observed SEVIRI image exceeds the threshold value. For instance, a low-cloud field at about 6°E/54°N and a cloud band at 12–15°E/48–49°N are missing while a dense "false alarm" cloud at 6–7°E/51°N is present in most of the ensemble members. These errors are
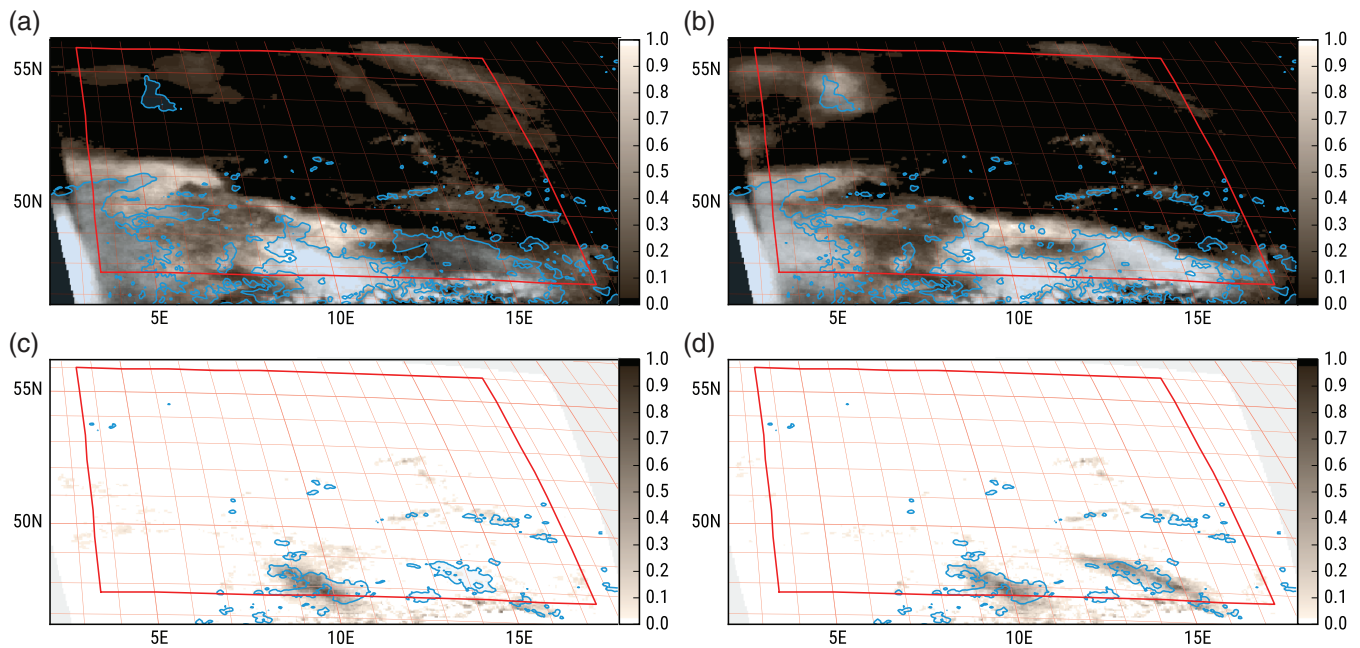
FIGURE 8 Fraction of ensemble members exceeding (a, b) a reflectance of 0.5 and (c, d) a precipitation rate of 1 mm·h^{-1} for the 1 hr forecast valid at 1000 UTC on 5 June 2016. (a, c) are initialised from a cycled experiment with data assimilation of conventional observations only and (b, d) initialised from a cycled experiment with additional visible SEVIRI reflectances. The blue contours indicate the geographical regions in which the threshold values were exceeded by the observation

removed or at least reduced in the VISCONV experiment (Figure 8b). The low-cloud field and the cloud band are present and the cloud at 6–7°E/51°N is significantly weaker.

Analogously to the probability of cloudiness, we consider a probability of precipitation, defined as the fraction of ensemble members in which the precipitation rate exceeds 1 mm·hr^{-1} (shown in Figures 8c,d). A comparison with the regions for which the observed precipitation rate exceeds 1 mm·hr^{-1} (blue contours) shows that the missing cloud band at 12–15°E/48–49°N was supposed to generate precipitation (Figure 8c). In VISCONV, the cloud band not only exists but does indeed produce precipitation (Figure 8d). The observed precipitation rate used here is DWD's radar-based precipitation product RADOLAN/EY.

An overview of the impact of the reflectance assimilation on the cloud cover error during the two days is provided by Figure 9. The figure shows the evolution of the root mean squared reflectance error,

$$ \text{RMSE} = \sqrt{\langle (\overline{\mathcal{R}} - \mathcal{R}_\text{O})^2 \rangle}, \tag{9} $$

and the mean difference of the reflectance fields,

$$ \text{MDIF} = \langle \overline{\mathcal{R}} - \mathcal{R}_\text{O} \rangle, \tag{10} $$

where $\overline{\mathcal{R}}$ is the ensemble mean reflectance field[1], $\mathcal{R}_\text{O}$ is the observed reflectance field and the brackets mean averaging over the field, that is, the satellite image. A positive/negative MDIF indicates in general that the overall cloud cover is too high/low in the model. RMSE and MDIF are displayed for 1-hr (i.e., first guess) and 3-hr forecasts started from the two experiments for each of the two days. It is obvious that at the start of the forecasts the reflectance RMSE is strongly reduced when the satellite observations are assimilated. For all VISCONV forecasts, it remains smaller than in the forecasts from the CONV reference experiment until 3 hr lead time. In almost all cases, the MDIF is closer to zero for VISCONV than for CONV at the start of the forecast, that is, the error in the area covered by clouds is reduced. In some cases, cloud cover changes so strongly during the 3 hr forecast that the MDIF of the CONV experiment ends up closer to zero than that of VISCONV. However, these changes in the mean difference are never strong enough to make the reflectance RMSE of VISCONV larger than the one of CONV. For 5 June, which is the day with weaker synoptic forcing, the impact of reflectance assimilation on cloud cover seems to be slightly more long-lived than for 29 May (compare the RMSEs for the forecasts started at 0600 and 0900 UTC in Figure 9c,d). However, a longer experimental period has to be assessed

---

[1] We use $\mathcal{R}$ for 2D reflectance fields (satellite images) and R for reflectance of individual pixels or super-observations.
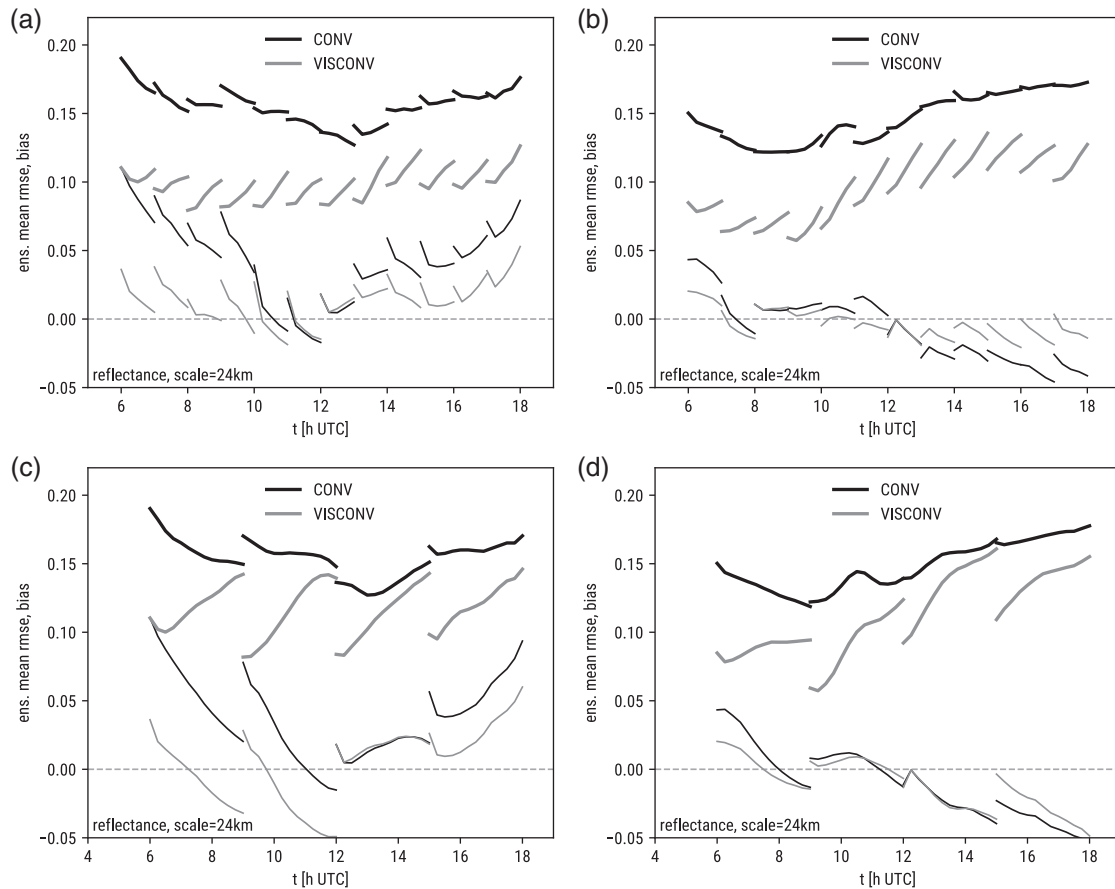
FIGURE 9 Evolution of RMSE (thick lines, Equation (9)) and mean difference MDIF (thin lines, Equation (10)) of the ensemble mean reflectance with respect to the observed reflectance (a, b) during the 1 hr first-guess forecasts of the data assimilation cycles and (c, d) during 3 hr forecasts started at 0600, 0900, 1200 and 1500 UTC from analyses of CONV experiment, with only conventional observation (black) and the VISCONV experiment with additional SEVIRI observation (grey) for (a, c) 29 May and (b, d) 5 June 2016. All synthetic and observed satellite images were coarsened to a resolution of 24 km before computing RMSE and MDIF. To generate these plots, synthetic satellite images were generated every 15 min from the forecast model states

to corroborate such an relation, which would be in agreement with the radar data assimilation experiments of Craig *et al.* (2012).

As an additional metric to measure the impact of the reflectance assimilation, we employ the ensemble mean fraction skill score (FSS) for the reflectance and the precipitation field. The FSS for reflectance with a threshold value of 0.5 and a spatial scale of 24 km (Figure 10a,b) reveals how well the location of dense clouds is predicted on that scale. The results confirm that the assimilation of reflectance has a clearly beneficial impact, which in almost all cases is still present after 3 hr. The FSS for the precipitation rate with a threshold value of 1 mm·h^{-1} and a slightly different window size of 30 km (Figure 10c,d) is also higher for the VISCONV experiment in almost all cases. The only case where the precipitation FSS becomes worse for the VISCONV experiment (5 June after 1100 UTC; Figure 10d) is related to the example from Figure 8. Before 1100 UTC,

the FSS is better for the VISCONV experiment, because it has the precipitating cloud band that is missing in the CONV experiment. In reality, the cloud band stopped raining at 1100 UTC, while in the VISCONV experiment precipitation continued after 1100 UTC in most of the VISCONV ensemble members. Thus, the precipitation in CONV experiment agrees better with the radar-derived precipitation observations from that time on. As visible reflectances do not provide direct information on the precipitation rate, additional observations (e.g., radar) would be beneficial to give greater weight to the ensemble members that also exhibit more accurate precipitation.

To assess the impact of assimilating VIS observations on the verification against conventional observations, we computed root mean square departures for the 3 hr forecasts from the CONV and VISCONV experiments for different observation types and variables. The mean relative change from CONV to VISCONV (Figure 11) is in
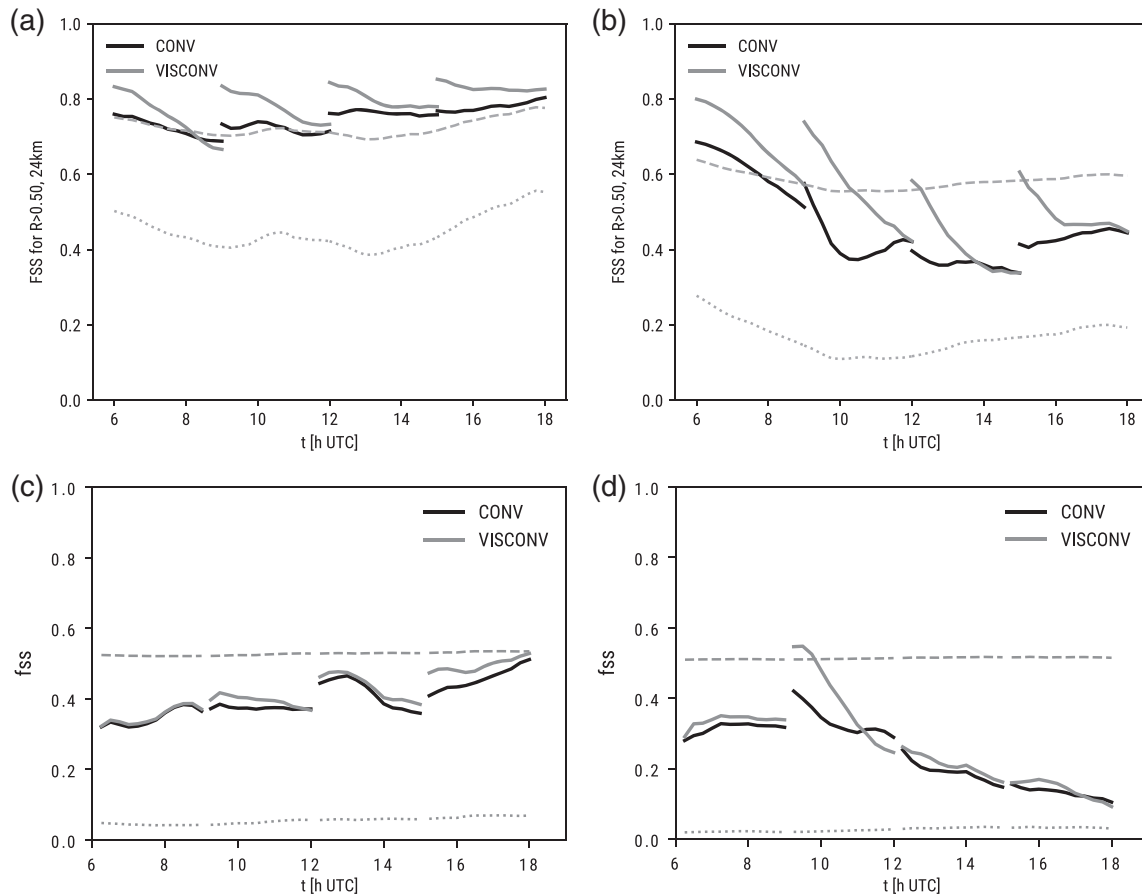
FIGURE 10 (a) Ensemble mean fractions skill score for reflectance with a threshold value of 0.5 and a window of 24 km for 3 hr forecasts for 29 May started from analyses of the CONV (black) and the VISCONV (grey) experiment. (c) is as (a), but for the precipitation rate with a threshold of 1 mm·h^{-1} and a window size of 30 km. (b, d) are as (a, c), but for 5 June. The grey dotted lines indicate the fraction f of the area in which the threshold is exceeded in the observation and which is equal to the skill score of a random forecast. The grey dashed line corresponds to a value of (1+f)/2, which is halfway between random and perfect forecast skill

most cases negative, that is, the reflectance assimilation improves the agreement of the forecast with conventional observations. In some cases, the reflectance assimilation causes an improvement for one day and a degradation for the other day, leading to a neutral total impact. For 5 June, the verification against observations indicates improvements, whereas results are mixed for 29 May. Overall, it is encouraging that we see a neutral to positive impact of the reflectance assimilation, and that for none of the observed variables in Figure 11 does the mean departure become worse by more than a few percent.

5.2 | Sensitivity to assimilation parameters

To investigate the sensitivity of the results presented in the previous section to modified assimilation parameters, we performed a range of further experiments for each of the

two days. The employed settings are listed in Table 1. The experiments are divided into groups corresponding to different settings for superobbing/thinning and localisation. There are groups with superobbing on scales between 6 km (corresponding to averaging over 2×1 pixel blocks) and 36 km (12×6 pixels) as well as groups in which thinning or a combination of superobbing and thinning is used. Within each group, only one parameter is modified (the parameters printed in bold face in Table 1), typically the assigned observation error.

In almost all experiments, the effective number of observations per grid point, N_{gp}^{sat}, is kept at a value close to 50. Only in group HL is the localisation scale varied without changing the superobbing settings, so that in these cases N_{gp}^{sat} also varies. All other assimilation settings not listed in Table 1 are the same as in the standard experiment VISCONV. The VISCONV settings are identical to S18-2 and HL-2, which have only been listed separately to make the variation of parameters within the group clearer.

FIGURE 11 Relative change in the RMSE of model observation equivalents due to assimilation of satellite reflectances. Results are shown for several conventional observation types and variables averaged over all 3 hr forecasts on 29 May and 5 June
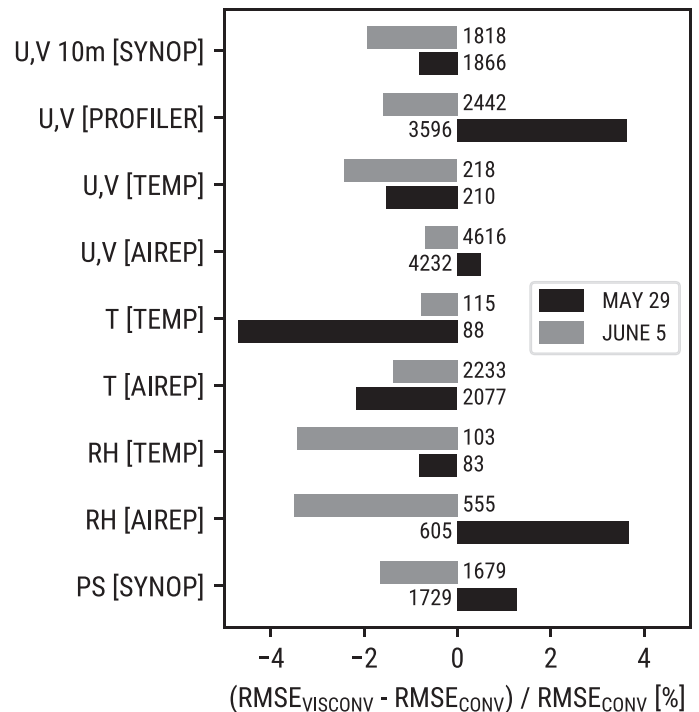
To compare the results for the experiments summarised in Table 1, it is useful to consider the RMSE of the ensemble mean reflectance (Equation 9), divided by the one of a reference experiment. In the following, we discuss this normalised reflectance RMSE for certain forecast lead times, averaged over all forecasts available for the two days. For forecast lead times up to 1 hr, 24 forecasts are available (including the first guesses from LETKF cycling). For longer lead times, only eight forecasts are available (from the 3 hr forecast initialised at 0600, 0900, 1200, 1500 UTC on the two days). The average normalised RMSE of 15 min forecasts serves as a proxy for the analysis error.

Figure 12 shows the average normalised RMSEs for 1-, 2- and 3-hr forecasts as a function of the analysis error proxy for all experiments of Table 1. Results of the CONV experiment were used to normalise the RMSEs. Figure 12 shows that for the best settings the mean RMSE is only 68% of the reference runs after 1 hr, 81% after 2 hr and 89% after 3 hr. It becomes obvious that drawing the analysis closer towards the observations is useful for reducing the error of the 1 hr forecast, but less so for the longer forecasts. The mean normalised RMSEs for the 1 hr forecasts vary much more with the analysis RMSE than the 3 hr forecasts.

To investigate the differences related to modified assimilation settings, it is more useful to normalise the reflectance RMSE with the one obtained for VISCONV, as all experiments for the settings listed in Table 1 are more strongly correlated with each other than with the CONV experiment. The error bars (with a length of one
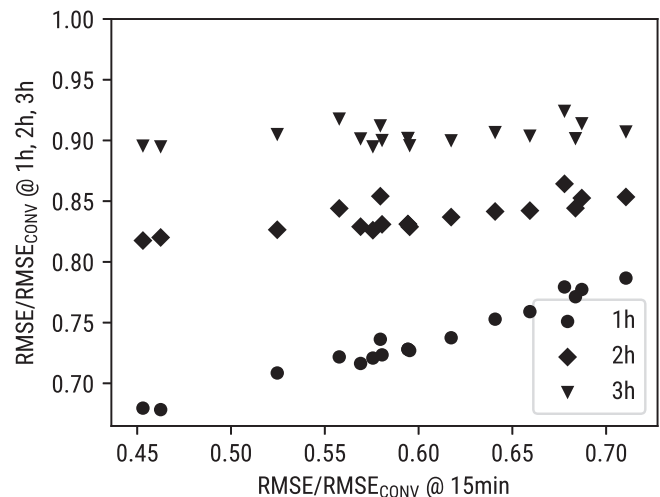
FIGURE 12 Average normalised reflectance RMSE of 1 hr, 2 hr, and 3 hr forecasts versus 15 min forecasts for all assimilation settings shown in Table 1. The CONV experiment with only conventional observations from Section 5.1 was used for the normalisation. All reflectance fields were averaged to 24 km before computing the RMSEs

standard deviation of the $RMSE/RMSE_{VISCONV}$ values for all forecasts) provide information on the significance of differences between the settings. Figure 13a shows a clear, nearly linear relation between the average normalised errors for 1 hr and 15 min forecasts with a variation that is much larger than the error bars. For the experiments with the smallest superobbing or thinning scales, the analysis
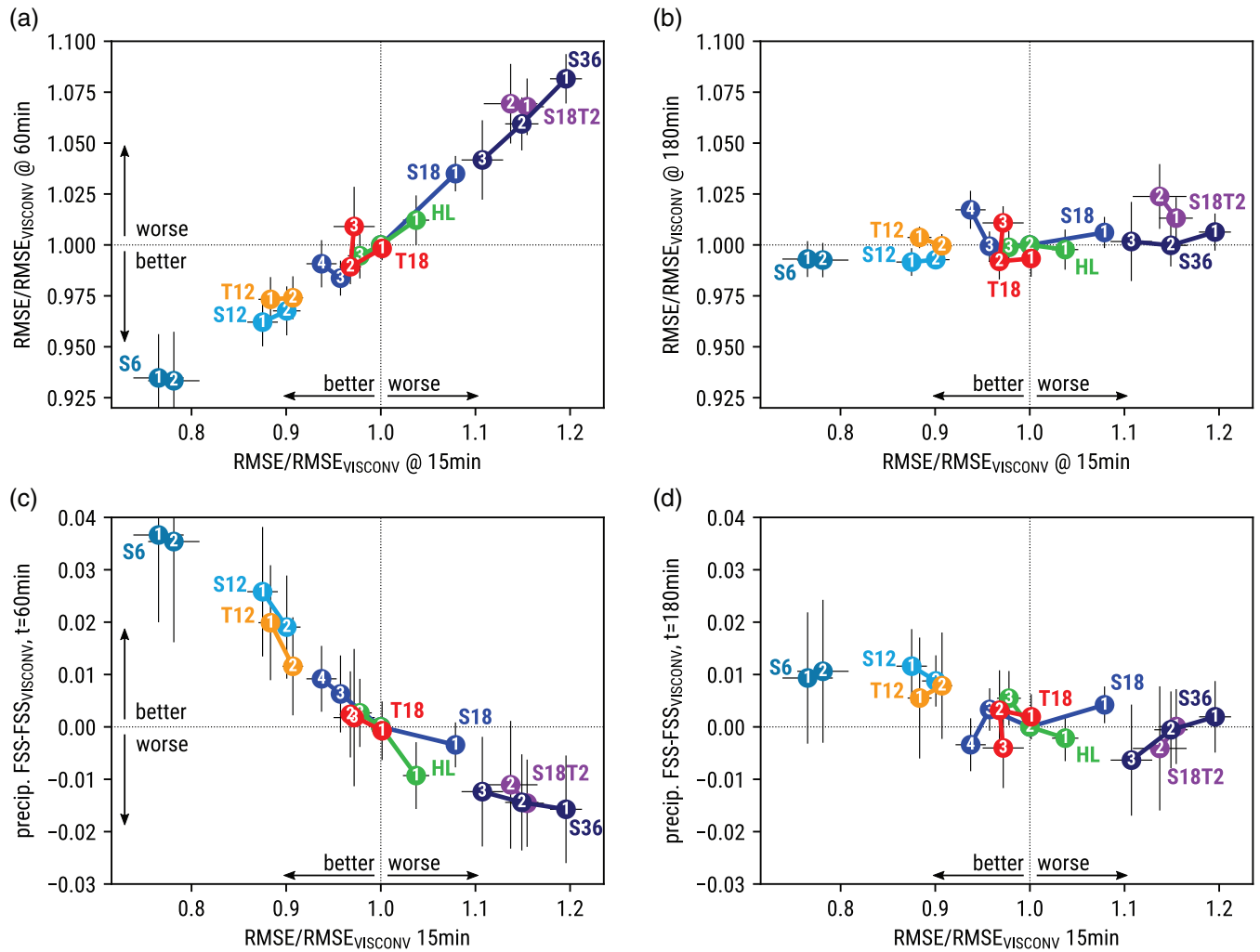
FIGURE 13 (a, b) show the average normalised reflectance RMSE for (a) 1 hr forecasts and (b) 3 hr forecasts versus 15 min forecasts. In contrast to Figure 12, the standard VISCONV experiment was used for the normalisation. All reflectance fields were averaged to 24 km before computing the RMSEs. (c, d) show the average difference of the precipitation FSS (1 mm·hr^{-1}, 30 km) relative to the standard VISCONV settings for (c) 1 hr and (d) 3 hr forecasts. Experiment groups are shown in different colours

and the 1 hr forecast normalised reflectance RMSE are significantly lower than for the larger scales. The smallest errors are found for the S6 group, where superobbing is performed by averaging over 2×1 pixel blocks. This is not very different from assimilating the satellite image at its original resolution

For the 3 hr forecasts (Figure 13b), the situation is less clear, but there is still some indication that smaller analysis errors result in slightly smaller 3 hr forecast errors. To assess the impact of the different assimilation settings on precipitation forecasts, we computed the average difference of the FSS to the one of the VISCONV experiment. The results displayed in Figure 13c,d for the 1 hr and the 3 hr forecasts agree with the reflectance results shown in Figure 13a,b. For the 1 hr forecast, smaller analysis reflectance errors result in clearly better precipitation forecasts while the relation is less obvious for the 3 hr forecast.

A comparison of results for the different experiment groups of Table 1 and the members within the groups provides more information on the impact of changes in the following assimilation parameters.
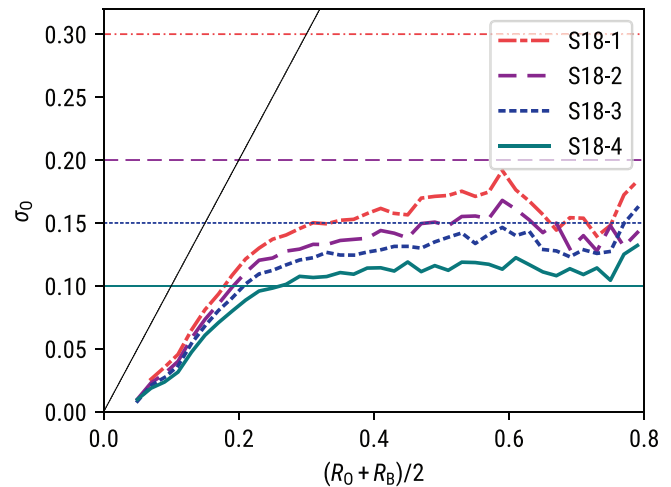
5.2.1 | Observation error

The assigned observation error should not only contain the instrument error, but also the operator error and representativity errors (Janjić *et al.*, 2018). Moreover, larger assumed observation errors can compensate for neglecting correlated errors and other deficiencies of the DA system. The optimal value of the observation error is thus not clear *apriori*. Therefore, we varied the assigned observation error within each of the groups with common superobbing and thinning settings. Taking group S18

as an example, an error of 0.3 as in S18-1 seems to be too high, as for all times the normalised reflectance error (Figure 13a,b) is higher than the one for S18-2, which is identical to VISCONV and uses an error of 0.2. Decreasing the error to 0.15 (S18-3) still leads to slightly reduced errors at 3 hr lead time. However, an error of 0.1 (S18-4) results in significantly increased reflectance errors and less precipitation forecast skill after 3 hr of forecast integration (Figure 13b,d).

As a consistency check, we used the Desroziers method (Desroziers *et al.*, 2005) to estimate observation errors from departure statistics. In this context, it has to be taken into account that reflectance is a bounded variable. There is obviously a lower bound, as reflectance cannot become negative. Mainly due to 3D radiative transfer effects, there is no hard upper limit, but values exceeding 1 are in general rare, so that in practice there is also some upper bound for reflectance. For bounded variables, the observation error should decrease towards the bounds, because otherwise values outside of the bounds would be included in the PDF, which is assumed to be Gaussian in the LETKF framework (Bishop, 2019).

For the sake of simplicity, we ignored this problem and used a constant observation error for the assimilation experiments. However, for the Desroziers estimate we have to take it into account to avoid unrealistic results. For this reason the estimate for σ_o was not calculated as a single value valid for all pixels, but computed as a function of the symmetric reflectance $(R_O + R_B)/2$, i.e., the mean of the observed reflectance and the first-guess ensemble mean reflectance . The estimated observation errors for the S18 group indeed show values much lower than 0.15 and a nearly linear behaviour for reflectances smaller than 0.2 (Figure 14). However, for a wide range of reflectances between 0.3 and 0.7, the estimates indicate that the observation error should be nearly constant, larger than 0.1 and smaller than 0.15. This is in agreement with the results displayed in Figure 13. Both the Desroziers estimates and the normalised reflectance error results also indicate that for the S36, S12, S6 T18, T12 and S18T2 groups the observation error should be around 0.15 (not shown).

Even though Figure 14 is not consistent with our simple approach of using a constant observation error, we consider this simplification a reasonable first step. In the framework of the VISCONV standard set-up, we further verified that the exclusion of observations with $(R_O + R_B)/2 < 0.2$ only results in small changes in the results. It is not clear if assimilating observations with very small values of $R_O + R_B$ in combination with a small observation error will improve the results, particularly because they contain only a weak cloud signal and might be dominated by surface albedo errors.

FIGURE 14 Desroziers estimate (thick lines) for the observation error as a function of the symmetric reflectance $(R_O + R_B)/2$ for the assimilation experiments from group S18 for 5 June. For each experiment, the thin horizontal line with the same line style indicates the corresponding assumed observation error. The diagonal dotted line is given by $(R_O + R_B)/2 = \sigma_o$

5.2.2 | Localisation scale and effective number of observations

In group HL, the localisation scale is varied between 35 km (experiment HL-3) and 70 km (experiment HL-1) while the area density of satellite observations is kept constant. A smaller l_h therefore corresponds to a reduced N_{gp}^{sat}, that is, less weight for the satellite observations compared to the conventional observations. The reduced weight could cause an increased reflectance error of the analysis. However, a smaller localisation radius also allows for a better adaptation of the analysis to the smaller scales reflected in the satellite observations, which could reduce the analysis error.

The reflectance and precipitation forecasts for 15 min and 1 hr are best for the smallest $l_h = 35$ km (Figure 13a,c). This means that the beneficial impact from decreasing the localisation scale from 70 km (HL-1) to 35 km (HL-3) is more important than the detrimental impact of reducing N_{gp}^{sat} from 95 (HL-1) to 24 (HL-3). Therefore, a small localisation radius seems to be more helpful than a high number of observations per grid point. After 3 hr, the reflectance errors are similar for all cases and only a slight advantage for the FSS is left for the $l_h = 35$ km settings (Figure 13b,d).

5.2.3 | Superobbing and thinning

As discussed in Section 2.4, the motivation for thinning is to reduce the potentially negative impact of
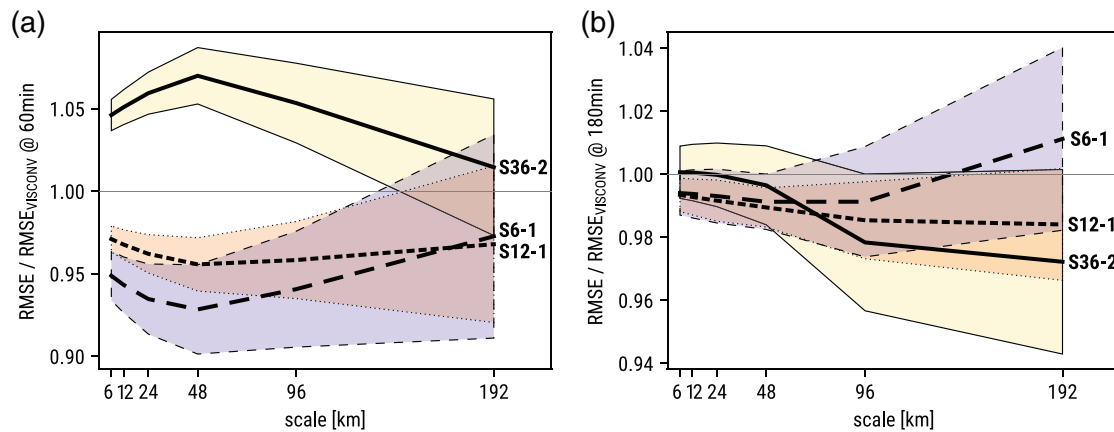
FIGURE 15   Mean (averaged over all forecasts for the two days with the same assimilation settings) reflectance RMSE divided by reflectance RMSE of the VISCONV experiment as a function of the scale on which the reflectance field was averaged (thick lines), for (a) 1 hr forecasts and (b) 3 hr forecasts of the experiments S36-2 (solid), S12-1 (dotted) and S6-1 (dashed). Thin lines with the same line style indicate the standard deviation for each experiment

neglecting correlated observation errors, whereas super-obbing aims at reducing the influence of representativity errors. It is not clear which of these two strategies is more advantageous. Moreover, the optimal spatial scale is unknown. For smaller superobbing or thinning scales, more small-scale information is included in the analysis, but the errors of the assimilated observations are more correlated. N_{gp}^{sat} was kept constant for the superobbing and thinning groups S36, S18, S12, S6, T18, T12 and S18T2. A smaller superobbing or thinning scale therefore also means a smaller horizontal localisation scale in these cases.

Figure 13a,c shows that the smallest errors in the 15 min and 1 hr forecasts are obtained for smallest super-obbing and localisation scales (e.g., compare groups S6, S12, S18, S36). The differences between superobbing and thinning (compare group S12 to T12 and group S18 to T18) are much smaller than the differences related to different spatial scales. One might be tempted to conclude that it is only the localisation scale that is important. However, changing only the localisation scale (as from setting HL-2 to HL-3) has a much weaker impact than the differences between T18 and T12 or S18 and S12. It is thus a smaller localisation scale in combination with a higher spatial density of observations that leads to reduced short-term forecast errors.

In group S6 the superobbing scale is well below the effective model resolution, but we still observe smaller forecast errors than in the groups with larger superob-bing scales. Compared to S12, the additional information available in S6 concerns scales that are not well resolved by the model and is thus probably not useful to reduce forecast errors. However, the improved ability to adapt to the observations accompanied by a smaller localisation

radius seems to be sufficient to give S6 an advantage over S12.

For Figures 13a,b the reflectance fields were averaged to 24 km before computing the RMSEs. An example for the impact of different superobbing scales on the scale-dependent normalised reflectance error after 1 hr and 3 hr is displayed in Figure 15, for which the reflectance fields were averaged on different scales between 6 km and 192 km. Here, the $l_{so} = 36$ km, $l_h = 100$ km settings lead to the largest errors on all scales for the 1 hr forecast, but to the lowest 3 hr forecast error at scales ≥ 96 km. In contrast, the errors for the $l_{so} = 6$ km, $l_h = 17$ km and $l_{so} = 12$ km, $l_h = 35$ km settings are lower than for the reference settings $l_{so} = 18$ km, $l_h = 50$ km for both forecast times and all scales.

These results show that, by assimilating observations with higher spatial density and simultaneously reducing the localisation scale, the analysis can be drawn significantly closer to the observations, in particular on smaller scales. When the analysis error is reduced in this way, the error growth in the forecast is enhanced. However, the error growth slows down after 3 hr for all cases and we still see slightly smaller errors for the high-spatial-density cases. If the goal is to strongly reduce the cloud displacement error in the first forecast hour, it thus seems advantageous to assimilate satellite images at high spatial resolution and there is no indication that this would make 3 hr and longer forecasts worse. However, in this approach large amounts of satellite data must be processed – for our 6 km superobbing experiments about 1.6×10^5 observations per hour (N_{tot}^{sat} in Table 1). If the primary goal is to improve forecasts with lead times of 3 hr or longer, it may be more advantageous to use superobbing or thinning on larger scales, which reduces the number of assimilated

satellite observations significantly, for example, to 4,600 per hour in the 36 km superobbing experiments. Thereby, the computational effort could be reduced or other additional types of observation could be assimilated.

6 | CONCLUSIONS

Visible satellite observations from instruments on geostationary satellites provide high-resolution cloud information that could be valuable for convective-scale data assimilation. However, these observations are not yet directly assimilated in operational NWP models as sufficiently fast and accurate forward operators have become available only recently. In this work, we present the first assimilation results for visible images from the SEVIRI instrument on Meteosat-10 using the KENDA/COSMO ensemble data assimilation system of DWD. These results are based on a case-study for two days of a period with strong convective precipitation in summer 2016.

Visible reflectance depends on the atmospheric state in a nonlinear and ambiguous way, which potentially poses challenges for current data assimilation algorithms. Single-observation experiments showed that nonlinearity and ambiguity of the observations can indeed limit the effectiveness of the assimilation process. However, the assimilation of these observations had a clear beneficial impact on cloud cover in most cases. Humidity and temperature fields were often improved as well. Cycled experiments with a near-operational set-up including conventional observations furthermore demonstrated that the assimilation of visible satellite images can strongly improve cloud cover forecasts and has a beneficial impact on precipitation forecasts.

Superobbing and thinning are standard techniques to overcome some of the limitations of current data assimilation systems like the assumption of diagonal observation covariance matrices with uncorrelated observation errors and their inability to provide reliable information on scales smaller than the effective model resolution. In a sensitivity study, we compared thinning and superobbing on different scales. The spatial scale turned out to be more important than the choice between superobbing and thinning. If the superobbing/thinning scale is reduced together with the horizontal localisation scale, short-term forecast errors on smaller scales can be reduced without a detrimental impact on 3 hr forecasts. This result is valid for both visible reflectance (providing information on cloud cover) and precipitation forecasts. The best results were obtained when the satellite images were assimilated nearly at their original resolution, with a superobbing scale of only 6 km. These results are particularly interesting for "seamless prediction" efforts combining nowcasting and very-short-range forecasting, as for example, in DWD's SINFONY (Seamless INtegrated FOrecastiNg sYstem) project. In such approaches, nowcasting information is to be used in NWP and *viceversa*. To alleviate this transfer of information between model and nowcasting clouds, it is essential to predict the position of cloud as accurate as possible in the first hours of the forecast (U. Blahak, personal communication, 2020).

We see a slightly more long-lasting impact of the VIS assimilation for the day on which the synoptic forcing was weaker and convection was triggered locally, which would be in agreement with the results of Craig *et al.* (2012). For both days, the impact is still present after 3 hr lead time, which is an encouraging results for such a synoptic situation with locally triggered strong summertime convective precipitation. However, it should be kept in mind that the experimental period of only two days is too short to draw robust general conclusions on the impact of this new observation type. In future studies we plan to investigate longer experimental periods with different synoptic situations and forecasts with considerably longer forecast horizons. In particular, a longer-lasting forecast improvement may be achievable in stable atmospheric conditions with low clouds such as fog and low stratus.

Issues related to spurious correlations and the ambiguity of visible reflectances should also be investigated further in future studies. Unlike the brightness temperature observations from thermal infrared channels, visible reflectances do not contain information about the vertical position of clouds which is needed for vertical localisation in observation space that the current implementation of the KENDA-LETKF depends on. Furthermore, combining the VIS assimilation with other satellite channels such as the 1.6 μm channel which allows for distinguishing between water and ice clouds (Wolters *et al.*, 2008), or thermal infrared channels that are sensitive to the cloud-top height, may reduce errors resulting from the ambiguity of the observations for individual channels.

For the sake of simplicity, we did not vary inflation parameters and used a constant assigned observation error in each experiment in this study. An adaptive error model like the ones developed for thermal infrared channels (e.g., Harnisch *et al.*, 2016; Minamide and Zhang, 2017; Okamoto *et al.*, 2019; Geer, 2019) could further improve the assimilation of visible channels. Moreover, a better understanding of the circumstances under which nonlinearity causes large errors in the LETKF analysis should be involved in the development of an error model or a quality control method. The choice of covariance inflation methods and parameters should also be assessed in the context of assimilating visible satellite images. And finally, a concurrent project aims at an improved model representation of clouds and hydrometeors that is key for the successful

assimilation of related observations and improved forecasts of clouds and precipitation.

ACKNOWLEDGEMENTS

ORCID

Leonhard Scheck https://orcid.org/0000-0003-1539-7197

REFERENCES

Bachmann, K., Keil, C., Craig, G.C., Weissmann, M. and Welzbacher, C.A. (2020) Predictability of deep convection in idealized and operational forecasts: effects of radar data assimilation, orography, and synoptic weather regime. *Monthly Weather Review*, 148, 63–81.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. and Reinhardt, T. (2011) Operational convective-scale numerical weather prediction with the cosmo model: description and sensitivities. *Monthly Weather Review*, 139, 3887–3905.

Bierdel, L., Friederichs, P. and Bentzien, S. (2012) Spatial kinetic energy spectra in the convection-permitting limited-area NWP model COSMO-DE. *Meteorologische Zeitschrift*, 21, 245–258. https://doi.org/10.1127/0941-2948/2012/0319

Bishop, C.H. (2019) Data assimilation strategies for state-dependent observation error variances. *Quarterly Journal of the Royal Meteorological Society*, 145, 217–227.

Craig, G.C., Keil, C. and Leuenberger, D. (2012) Constraints on the impact of radar rainfall data assimilation on forecasts of cumulus convection. *Quarterly Journal of the Royal Meteorological Society*, 138, 340–352.

Desroziers, G., Berre, L., Chapnik, B. and Poli, P. (2005) Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131, 3385–3396. https://doi.org/10.1256/qj.05.108

Fu, Q. (1996) An accurate parameterization of the solar radiative properties of cirrus clouds for climate models. *Journal of Climate*, 9, 2058–2082.

Gaspari, G. and Cohn, S.E. (1999) Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757. https://doi.org/10.1002/qj.49712555417

Geer, A.J. (2019) Correlated observation error models for assimilating all-sky infrared radiances. *Atmospheric Measurement Techniques*, 12, 3629–3657. https://doi.org/10.5194/amt-12-3629-2019

Geer, A.J., Baordo, F., Bormann, N., Chambon, P., English, S.J., Kazumori, M., Lawrence, H., Lean, P., Lonitz, K. and Lupu, C. (2017) The growing impact of satellite observations sensitive to humidity, cloud and precipitation. *Quarterly Journal of the Royal Meteorological Society*, 143, 3189–3206. https://doi.org/10.1002/qj.3172

Geer, A.J., Lonitz, K., Weston, P., Kazumori, M., Okamoto, K., Zhu, Y., Liu, E.H., Collard, A., Bell, W., Migliorini, S., Chambon, P., Fourrié, N., Kim, M.-J., Köpken-Watts, C. and Schraff, C. (2018) All-sky satellite data assimilation at operational weather forecasting centres. *Quarterly Journal of the Royal Meteorological Society*, 144, 1191–1217. https://doi.org/10.1002/qj.3202

Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissman, M., Reich, H., Brousseau, P., Montmerle, T., Wattrelot, E., Bučánek, A., Mile, M., Hamdi, R., Lindskog, M., Barkmeijer, J., Dahlbom, M., Macpherson, B., Ballard, S., Inverarity, G., Carley, J., Alexander, C., Dowell, D., Liu, S., Ikuta, Y. and Fujita, T. (2017) Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quarterly Journal of the Royal Meteorological Society*, 144, 1218–1256. https://doi.org/10.1002/qj.3179

Harnisch, F. and Keil, C. (2015) Initial conditions for convective-scale ensemble forecasting provided by ensemble data assimilation. *Monthly Weather Review*, 143, 1583–1600.

Harnisch, F., Weissmann, M. and Periáñez, Á. (2016) Error model for the assimilation of cloud-affected infrared satellite observations in an ensemble data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 142, 1797–1808.

Hu, Y.X. and Stamnes, K. (1993) An accurate parameterization of the radiative properties of water clouds suitable for use in climate models. *Journal of Climate*, 6, 728–742.

Hunt, B.R., Kostelich, E.J. and Szunyogh, I. (2007) Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230, 112–126. https://doi.org/10.1016/j.physd.2006.11.008

Janjić, T., Bormann, N., Bocquet, M., Carton, J.A., Cohn, S.E., Dance, S.L., Losa, S.N., Nichols, N.K., Potthast, R., Waller, J.A. and Weston, P. (2018) On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144, 1257–1278. https://doi.org/10.1002/qj.3130

Keil, C., Baur, F., Bachmann, K., Rasp, S., Schneider, L. and Barthlott, C. (2019) Relative contribution of soil moisture, boundary-layer and microphysical perturbations on convective predictability in different weather regimes. *Quarterly Journal of the Royal Meteorological Society*, 145, 3102–3115.

Kostka, P.M., Weissmann, M., Buras, R., Mayer, B. and Stiller, O. (2014) Observation operator for visible and near-infrared satellite reflectances. *Journal of Atmospheric and Oceanic Technology*, 31, 1216–1233.

Kurzrock, F., Cros, S., Ming, F.C., Otkin, J.A., Hutt, A., Linguet, L., Lajoie, G. and Potthast, R. (2018) A review of the use of geostationary satellite observations in regional-scale models for short-term cloud forecasting. *Meteorologische Zeitschrift*, 27, 277–298. https://doi.org/10.1127/metz/2018/0904

Lin, Y.-L., Farley, R.D. and Orville, H.D. (1983) Bulk parameterization of the snow field in a cloud model. *Journal of Climate and Applied Meteorology*, 22, 1065–1092.

Martin, G.M., Johnson, D.W. and Spice, A. (1994) The measurement and parameterization of effective radius of droplets in warm stratocumulus clouds. *Journal of the Atmospheric Sciences*, 51, 1823–1842.

McNally, A.P. (2002) A note on the occurrence of cloud in meteorologically sensitive areas and the implications for advanced infrared sounders. *Quarterly Journal of the Royal Meteorological Society*, 128, 2551–2556. https://doi.org/10.1256/qj.01.206

Minamide, M. and Zhang, F. (2017) Adaptive observation error inflation for assimilating all-sky satellite radiance. *Monthly Weather Review*, 145, 1063–1081.

Necker, T., Geiss, S., Weissmann, M., Ruiz, J., Miyoshi, T. and Lien, G.-Y. (2020) A convective-scale 1,000-member ensemble simulation and potential applications. *Quarterly Journal of the Royal Meteorological Society*, 146, 1423–1442. https://doi.org/10.1002/qj.3744

Necker, T., Weissmann, M. and Sommer, M. (2018) The importance of appropriate verification metrics for the assessment of observation impact in a convection-permitting modelling system. *Quarterly Journal of the Royal Meteorological Society*, 144, 1667–1680. https://doi.org/10.1002/qj.3390

Okamoto, K. (2017) Evaluation of IR radiance simulation for all-sky assimilation of Himawari-8/AHI in a mesoscale NWP system. *Quarterly Journal of the Royal Meteorological Society*, 143, 1517–1527. https://doi.org/10.1002/qj.3022

Okamoto, K., Sawada, Y. and Kunii, M. (2019) Comparison of assimilating all-sky and clear-sky infrared radiances from Himawari-8 in a mesoscale system. *Quarterly Journal of the Royal Meteorological Society*, 145, 745–766. https://doi.org/10.1002/qj.3463

Piper, D., Kunz, M., Ehmele, F., Mohr, S., Mühr, B., Kron, A. and Daniell, J. (2016) Exceptional sequence of severe thunderstorms and related flash floods in May and June 2016 in Germany – Part 1: Meteorological background. *Natural Hazards and Earth System Sciences*, 16, 2835–2850. https://doi.org/10.5194/nhess-16-2835-2016

Quaas, J. (2012) Evaluating the 'critical relative humidity' as a measure of subgrid-scale variability of humidity in general circulation model cloud cover parameterizations using satellite data. *Journal of Geophysical Research; Atmospheres*, 117(D9). https://doi.org/10.1029/2012JD017495

Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., Vidot, J., Roquet, P., Matricardi, M., Geer, A., Bormann, N. and Lupu, C. (2018) An update on the RTTOV fast radiative transfer model (currently at version 12). *Geoscientific Model Development*, 11, 2717–2737. https://doi.org/10.5194/gmd-11-2717-2018

Scheck, L., Frèrebeau, P., Buras-Schnell, R. and Mayer, B. (2016) A fast radiative transfer method for the simulation of visible satellite imagery. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 175, 54–67.

Scheck, L., Weissmann, M. and Mayer, B. (2018) Efficient methods to account for cloud-top inclination and cloud overlap in synthetic visible satellite images. *Journal of Atmospheric and Oceanic Technology*, 35, 665–685. https://doi.org/10.1175/JTECH-D-17-0057.1

Schomburg, A., Schraff, C. and Potthast, R. (2015) A concept for the assimilation of satellite cloud information in an ensemble Kalman filter: single-observation experiments. *Quarterly Journal of the Royal Meteorological Society*, 141, 893–908. https://doi.org/10.1002/qj.2407

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periáñez, A. and Potthast, R. (2016) Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Quarterly Journal of the Royal Meteorological Society*, 142, 1453–1472. https://doi.org/10.1002/qj.2748

Seifert, A. and Beheng, K.D. (2001) A double-moment parameterization for simulating autoconversion, accretion and selfcollection. *Atmospheric Research*, 59-60, 265–281. https://doi.org/10.1016/S0169-8095(01)00126-0

Simmer, C., Adrian, G., Jones, S., Wirth, V., Göber, M., Hohenegger, C., Janjić, T., Keller, J., Ohlwein, C., Seifert, A., Trömel, S., Ulbrich, T., Wapler, K., Weissmann, M., Keller, J., Masbou, M., Meilinger, S., Riß, N., Schomburg, A., Vormann, A. and Weingärtner, C. (2016) HErZ: the German Hans-Ertel centre for weather research. *Bulletin of the American Meteorological Society*, 97, 1057–1068. https://doi.org/10.1175/BAMS-D-13-00227.1

Stamnes, K., Tsay, S.-C., Jayaweera, K. and Wiscombe, W. (1988) Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. *Journal of Applied Optics*, 27, 2502–2509.

Tiedtke, M. (1989) A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, 117, 1779–1800.

Weissmann, M., Göber, M., Hohenegger, C., Janjić Pfander, T., Keller, J., Ohlwein, C., Seifert, A., Trömel, S., Ulbrich, T., Wapler, K., Bollmeyer, C. and Deneke, H. (2014) Initial phase of the Hans Ertel centre for weather research. A virtual centre at the interface of basic and applied weather and climate research. *Meteorologische Zeitschrift*, 23, 193–208. https://doi.org/10.1127/0941-2948/2014/0558

Wolters, E.L.A., Roebeling, R.A. and Feijt, A.J. (2008) Evaluation of cloud-phase retrieval methods for SEVIRI on Meteosat-8 using ground-based lidar and cloud radar data. *Journal of Applied Meteorology and Climatology*, 47, 1723–1738.

Wyser, K. (1998) The effective radius in ice clouds. *Journal of Climate*, 11, 1793–1802.

Zeng, Y., Janjić, T., de Lozar, A., Blahak, U., Reich, H., Keil, C. and Seifert, A. (2018) Representation of model error in convective-scale data assimilation: additive noise, relaxation methods, and combinations. *Journal of Advances in Modeling Earth Systems*, 10, 2889–2911. https://doi.org/10.1029/2018MS001375

Zhang, F., Snyder, C. and Sun, J. (2004) Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Monthly Weather Review*, 132, 1238–1253.

Zhu, Y., Liu, E., Mahajan, R., Thomas, C., Groff, D., Van Delst, P., Collard, A., Kleist, D., Treadon, R. and Derber, J.C. (2016) All-sky microwave radiance assimilation in NCEP's GSI analysis system. *Monthly Weather Review*, 144, 4709–4735.