Monitoring Regional Atmospheric Stability with Hyperspectral Satellite Observations Enhanced with a Potential Network of Ground-Based Microwave Radiometers

MARIA TOPOROV[a,b] AND ULRICH LÖHNERT[a,b]

[a] *Institute for Geophysics and Meteorology, University of Cologne, Cologne, Germany*
[b] *Hans-Ertel-Centre for Weather Research, Cologne/Bonn, Germany*

(Manuscript received 31 October 2024, in final form 23 June 2025, accepted 21 July 2025)

ABSTRACT: This study attempts to assess the potential benefit of a network of ground-based microwave radiometers (MWRs) in variable configuration in synergy with simultaneous geostationary satellite observations from the Meteosat Third-Generation Infrared Sounder (IRS) for observing regional atmospheric stability. Since both observation types are not yet available, reanalysis fields of the western part of Germany (Rhein–Ruhr area) were used to simulate MWR and IRS observations. From these simulations, neural network retrievals for CAPE and lifted index (LI) were derived for satellite-only (IRS), ground-based-only (MWR), and synergistic (IRS + MWR) observation configurations. The developed retrievals were applied to two years of simulated observations. In general, the IRS + MWR retrievals outperform (up to 0.3 improvement in Heidke skill score) the IRS retrievals under clear sky and especially under cloudy conditions, if both observation types are available simultaneously and at the same location. The impact of adding one to 200 MWRs of the hypothetical observation network to the satellite observations was investigated using a spatial statistical interpolation method. The fields of CAPE/LI from IRS observations were merged with the CAPE/LI values from the MWR network by taking into account the corresponding error covariance matrices of both retrievals, which were derived from reanalysis data. It was shown that the contribution of ground-based observations is more pronounced under cloudy conditions and is most valuable for the first 25 sensors located in the domain.

KEYWORDS: Microwave observations; Remote sensing; Satellite observations; Sensitivity studies; Neural networks

1. Introduction

The accurate prediction of severe weather and issuance of warnings are important for public safety. However, even state-of-the-art high-resolution (kilometer scale, deep convection resolving) numerical weather prediction (NWP) models still have difficulties in predicting the exact position and time of locally influenced high-impact weather events such as convective storms or fog (Burghardt et al. 2014; Novak et al. 2011; Henderson et al. 2021; Martinet et al. 2020). The development of these events depends, among others, on potential instability of the atmosphere (Markowski and Richardson 2010). Atmospheric thermodynamic stability determines the tendency of the air to rise or, in the opposite, to resist vertical motion. It is determined by the vertical distribution of temperature and humidity and can be described in terms of so-called stability indices (Peppler 1988). Two frequently used indices are convective available potential energy (CAPE) and lifted index (LI) (Blanchard 1998).

Besides NWP models, the information needed for the calculation of CAPE and LI (i.e., temperature and humidity fields) is provided by radiosondes and satellite observations. Radiosondes are typically only available 1–2 times a day at a specific location, which is not sufficient for capturing the temporal and spatial variability of the thermodynamic state of the atmosphere.

Satellite observations from geostationary platforms enable continuous monitoring of stability indices with higher temporal and spatial resolution (e.g., Koenig and de Coning 2009). The Meteosat Second Generation Global Instability Index (MSG-GII) product is based on observations of the MSG SEVIRI instrument and covers Africa and Europe. It includes three instability indices, the LI, KOnvektiv-index (KO), and the K index (KI), as well as the total precipitable water (TPW) (EUMETSAT 2013). The GII product is produced with a horizontal resolution of approximately 9 km at the subsatellite point. Although the horizontal resolution of the GII products decreases with increasing latitude, the full disc coverage and repeat cycle of 15 min still represent a significant improvement to the sparsely located radiosonde sounding sites. However, the MSG-GII product is based on infrared observations and therefore restricted to cloud-free and preconvective areas. Moreover, currently operational geostationary instruments have low spectral resolution and provide only limited information on the vertical structure of the atmosphere, especially of the lowest layers (Schmit et al. 2008). Particularly, the clear-sky radiances of SEVIRI are mainly sensitive to the water vapor in the mid- and upper troposphere (Stengel et al. 2009).

Hyperspectral IR observations from geostationary platforms will be available from the recently launched Infrared Sounder on board the Meteosat Third Generation (MTG-IRS; https://www.eumetsat.int). IRS is expected to provide a more detailed picture of four-dimensional water vapor and temperature structures with high temporal resolution and full disc coverage (Holmlund et al. 2021). The first satellite of the MTG series carrying IRS was launched in July 2025 and is designed to perform a full disc scan from the 0° nominal

Denotes content that is immediately available upon publication as open access.

Corresponding author: Maria Toporov, mtoporo1@uni-koeln.de

DOI: 10.1175/JAMC-D-24-0212.1

longitude. The established monitoring of stability indices will be complemented by the IRS observations.

The benefit of simulated geostationary hyperspectral IR observations assimilated into a regional forecast model was shown in a number of experiments (Guedj et al. 2014; McGrath-Spangler et al. 2022; Coopmann 2022). However, besides the higher temporal and horizontal resolution of the measurements, the hyperspectral IR observations from geostationary orbit will be affected by the same limitations, such as cloudiness, uncertain information on surface emissivity, and low accuracy and resolution in the boundary layer, as operational hyperspectral sounders on board of polar orbiting satellites (Pougatchev et al. 2009).

Several convective case studies have shown that the main limitation of the hyperspectral Infrared Atmospheric Sounding Interferometer (IASI) on board EUMETSAT's MetOp polar orbiters is the low accuracy of retrieved humidity profiles in the lower troposphere. The lower-layer moisture is often underestimated and elevated moist layers are not detected, which has a large impact on CAPE calculated from IASI-retrieved profiles (EUMETSAT 2020a). Moreover, theoretical assessment of the IRS-retrieved humidity profiles has shown that over land surfaces, the accuracy of humidity profiles in the layer between the surface and 800 hPa degrades with increasing satellite zenith angle (EUMETSAT 2020b). The accuracy of the temperature profiles obtained from IRS observations in midlatitudes was shown to degrade with increasing instability.

The described limitations of different observing systems can be reduced by constraining the hyperspectral satellite observations with surface observations or with ground-based remote sensing. Particularly, Bloch et al. (2019) have shown that the replacement of the surface-parcel properties in the satellite soundings with the surface or the radiosonde observations leads to significant improvements in the accuracy of surface-based CAPE. There are a growing number of ground-based instruments deployed worldwide for atmospheric boundary layer profiling (Cimini et al. 2020). The measured atmospheric variables include the vertical profiles of temperature, humidity, wind, aerosol, and cloud properties. Previous studies have shown the benefit of combining ground-based remote sensing sensors with satellite observations for the retrieval of atmospheric state (Ebell et al. 2013) and of stability indices (Toporov and Löhnert 2020). Loveless et al. (2022) demonstrated the benefits of combining ground- and satellite-based infrared observations for the retrieval of temperature and humidity profiles. The combination of ground-based Atmospheric Emitted Radiance Interferometer (AERI) with space-based hyperspectral IR sounders resulted in the increase in degrees of freedom for signal (DOFS) by 30%–40% in the layer below 700 hPa compared to the space-based instruments alone. This results in significant improvement in vertical resolution and reduced uncertainties of the retrieved profiles in the lowest 1000 m.

Ground-based microwave radiometers (MWRs) are being considered to fill the observation gap [e.g., through the European National Meteorological Services Profiling Program (EUMETNET E-PROFILE) in Europe] in the atmospheric boundary layer (ABL) because they are suited for 24/7 operations (also in cloudy conditions) and contain valuable information on water vapor and temperature (Löhnert and Maier 2012; Löhnert et al. 2009). In contrast to radiosondes, MWRs provide temporally highly resolved profiles, however, with limited vertical resolution. For temperature profiles, MWRs provide about 3–4 DOFS, mostly within the ABL, whereas for humidity, only about two independent pieces of information are provided on the vertical structure (Löhnert et al. 2009). First attempts to assimilate the ground-based MWR remote sensing observations into high-resolution NWP models showed improvements in the detection of fog (Martinet et al. 2020), in the precipitation forecast (Caumont et al. 2016), and a positive impact on the 6-h forecast of ABL temperature and humidity (Vural et al. 2024).

How many ground-based MWRs are needed in a future observation network? Maintaining an operational observing network is a challenging and expensive task. Therefore, it is essential to carefully evaluate the impact of different components of the observing system. The present study aims to show the impact of two new observation systems (MTG-IRS and ground-based MWR) on the regional retrieval of stability indices (CAPE/LI), which are used as indicators for severe convection.

For this, we build upon the study performed by Toporov and Löhnert (2020), in which observations from a single MWR (at the subsatellite point) were shown to have a large potential in complementing IRS-satellite observations by providing thermodynamic information from the lowest atmospheric layers and below clouds, where the satellite is unable to provide useful information.

This study now goes a step forward and poses the following two research questions: 1) How accurately can the spatially and temporally continuously available MTG-IRS observations retrieve the potential of severe convection (via CAPE/LI) over a regional domain? 2) And how much information can a potential network of ground-based MWR in varying configurations (i.e., spatial density) add to the regional stability information derived by the satellite?

At the time of publication, data from the already launched MTG-IRS was not yet available and extensive ground-based MWR networks were only as a theoretical consideration. Therefore, we base our study on the assessment of simulated observations. For this, reanalysis fields are considered as the true atmospheric state. First, from these atmospheric fields, stability indices are derived and satellite (MTG-IRS) and ground-based MWR observations are simulated to obtain pairs of atmospheric state and measurement vectors. With these data pairs, neural network (NN) retrievals are derived that output time series of CAPE/LI fields from satellite only, ground-based observations only, and a synergy of the two. In the following, we use a statistical interpolation method to estimate the spatial influence of a varying number of ground-based MWR observations in our regional domain: from one instrument to over 1500. The obtained CAPE/LI fields are compared to the original fields calculated directly from reanalysis to assess the accuracy and information content of the retrievals.

The article is structured as follows. Section 2 introduces the reanalysis dataset and the stability indices CAPE and LI. The instruments, their measurement principle, and assumptions

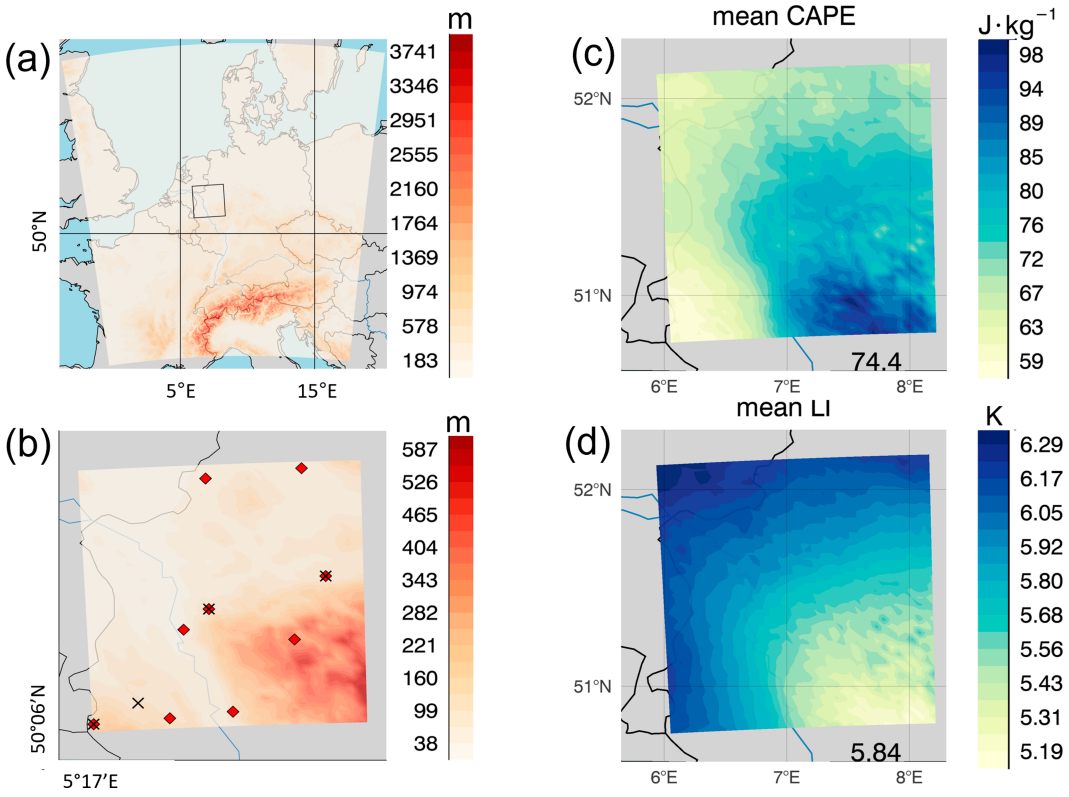

FIG. 1. (a) The target domain used in the study (black square), including the metropolitan Rhein–Ruhr area, within the COSMO-REA2 domain. (b) Detailed topography of the target domain. Red diamonds and black crosses show the locations of DWD SYNOP stations and of E-PROFILE ceilometers, respectively. Note the different range of color bars. (c),(d) Mean values of CAPE and LI for the considered period.

made when simulating the observations are explained in Section 3. Section 4 provides an overview of two methods applied in the study: the neural networks and their configuration (section 4a) and the statistical interpolation (section 4b). Section 5 gives details on the setup of the sensitivity study concerning the spatial number density of the MWRs and the calculation of error covariance matrices required for statistical interpolation. The results are discussed in Section 6, which is divided into two parts: The neural network retrieval performance is discussed in section 6a, whereas the analysis of the spatial statistical interpolation approach is discussed in section 6b. A summary and conclusions are given in Section 7.

2. Reanalysis data

The impact of future, not-yet-existing instruments and instrument networks can be assessed by simulation of their observations based on realistic atmospheric profiles. The best, temporally consistent estimates of the atmospheric state of a past time period can be obtained from reanalyses. The high-resolution regional reanalysis COSMO-REA2 is based on the Consortium for Small-Scale Modeling (COSMO[1]) limited-area model (Doms and Baldauf 2018; Doms et al. 2011) and has been developed within the "Hans-Ertel-Centre for Weather Research" (HErZ; Simmer et al. 2016). A detailed description can be found in Bollmeyer et al. (2015) and Wahl et al. (2017).

This study was performed using the COSMO-REA2 reanalysis data for the target domain located in the western part of Germany as shown in Fig. 1. The 150 by 150 km large domain has heterogeneous surface properties and covers the Lower Rhine basin in the central west and the hills of the Rhenish Massif in the southeast. Within the domain, the metropolitan Rhine-Ruhr Area is located—the most densely populated region in Germany. This makes the prediction of severe weather crucial for public safety.

The temporal and horizontal resolutions of the reanalysis fields are 1 h and 4 km (original reanalysis fields were thinned from 2 to 4 km to account for the coarser resolution of IRS observations), respectively, resulting in a field with 39 × 39 grid points and about 7300 profiles per grid point. The vertical grid has 50 layers and extends from the surface to 22 km, with the spacing between levels increasing from 20 to 1000 m. The reanalysis profiles from May to September in the years 2010 and 2011 were used to calculate fields of CAPE and lifted index and to simulate the observations of IRS and MWR.

The detailed description of indices is given in the appendixes A and B. The threshold values for CAPE and LI separating stable

[1] http://www.cosmo-model.org/.

and slightly unstable conditions were taken according to Haklander and Van Delden (2003) and lie by 168 J kg^{-1} and 1.6 K, respectively. The higher/lower the CAPE/LI value, the more unstable the atmosphere and the higher the probability of thunderstorms. The mean values of CAPE and LI for the considered period (Fig. 1) reveal more unstable conditions (larger number of unstable events and higher/lower CAPE/LI values) in the southeast part of the domain.

3. Instruments and simulation of observations

a. IRS

The next generation of Meteosat satellites, MTG, will comprise four imaging (MTG-I) and two sounding satellites (MTG-S) (Holmlund et al. 2021). The latter will replace *Meteosat-11* at 0° longitude and bring an operational hyperspectral instrument into geostationary orbit. The IRS is a sounding Fourier transform spectrometer which will perform highly spectrally resolved measurements of Earth-emitted radiation in 1738 channels. According to the MTG Mission Requirements Document (EUMETSAT 2018), IRS will provide observations in two bands, in the longwave IR (LWIR; 700–1210 cm^{-1}) and in the midwave IR (MWIR; 1600–2175 cm^{-1}) band with a spectral resolution of 0.625 cm^{-1} and a spatial sampling distance of 4 km at nadir. The basic repeat cycle of IRS will take 60 min with an increased frequency of 30 min over Europe. The satellite viewing angle for the considered region is around 60°.

The channels in the LWIR band are mostly sensitive to surface and cloud properties, atmospheric temperature, and ozone, whereas MWIR channels provide information on humidity and temperature. The main objective of the IRS mission is the monitoring of the evolution of vertically resolved water vapor, temperature, and wind structures.

As in the previous study by Toporov and Löhnert (2020), a subset of IRS channels that give information on atmospheric temperature and humidity was selected. The subset of 1113 channels in total consists of the following: 130 channels along the longwave CO_2 absorption band between 700 and 780 cm^{-1}, every second channel between 780 and 1210 cm^{-1} (344 in total), and 639 channels in the water vapor absorption band between 1600 and 2000 cm^{-1}.

The simulation of IRS observations was performed with the fast radiative transfer model RTTOV v12 (Saunders et al. 2018). The detailed description of the radiative transfer and of the simulation of IRS observations from COSMO-REA2 reanalysis profiles can be found in Toporov and Löhnert (2020) and Toporov (2021). Here, we shortly summarize the most important aspects.

For simulation of clear-sky IRS observations, RTTOV requires atmospheric profiles of temperature, humidity, and trace gases [carbon dioxide (CO_2), ozone (O_3), nitrous oxide (N_2O), methane (CH_4), sulfur dioxide (SO_2), carbon monoxide (CO)] along with surface properties. In this study, the profiles of trace gases are assumed to be constant and set to the RTTOV reference profiles. The surface emissivity values for satellite instruments were taken from the RTTOV emissivity atlas University of Wisconsin Global Infrared Land Surface Emissivity (UWIREmis), which provides monthly climatological mean emissivity values at a spatial resolution of 0.1° (Borbas and Ruston 2010).

For the simulation of cloudy observations, additional profiles of cloud liquid/ice water and cloud fraction are required. For water clouds, five different cloud types are considered. For ice clouds, we used the Baran ice scheme (Vidot et al. 2015) that performs direct parameterization of ice optical properties from the ambient temperature and the ice water content.

To produce a realistic observation dataset, the simulated spectra were perturbed with normally distributed noise, which varies between 0.2 and 0.9 K (EUMETSAT 2018). To reduce the dimensionality and to optimally extract atmospheric profile information from simulated IRS observations, principal component analysis (PCA) was applied to the dataset. PCA makes use of redundant information in hyperspectral observations and transforms highly correlated observations into an uncorrelated set of principal components. The first principal components represent the most dominant atmospheric signal contained in the original spectrum, whereas the last principal components consist mostly of random instrument noise and can be discarded. In the following, 50 first principal components, explaining 99.94% of the observations' variability, were used.

b. Ground-based MWR

Microwave radiometer considered in this study is a Humidity and Temperature Profiler (HATPRO; Rose et al. 2005) manufactured by Radiometer Physics GmbH (RPG), Germany. It measures the downwelling radiation emitted by atmospheric components, mainly oxygen, water vapor, and cloud liquid. The first seven channels are located in the K band, at the slope of the pressure-broadened water vapor absorption line (22.24, 23.04, 23.84, 25.44, 26.24, 27.84, and 31.40 GHz). These frequencies provide information for accurate retrievals of integrated water vapor (IWV) and liquid water path (LWP) and for low-resolution humidity profiling. The last seven channels are distributed in the V band along the long-wavelength slope of the 60-GHz oxygen absorption complex (51.26, 52.28, 53.86, 54.94, 56.66, 57.30, and 58 GHz). Increasing opacity toward the band center and homogeneous mixing of oxygen throughout the troposphere make the retrieval of temperature profiles possible. The optically thick channels provide information from atmospheric layers closer to the MWR, whereas more transparent channels receive radiation also from higher atmospheric layers. The resolution and accuracy of the retrieved temperature profiles can be improved by assuming the horizontally homogeneous atmosphere and performing elevation scanning measurements at the most opaque channels in the oxygen absorption band. The accuracy of the retrieved temperature profiles is between 0.5 and 2 K close to the surface and in the lower troposphere, respectively, whereas humidity profile accuracies are in the range of 0.8 g m^{-3} for the midlatitudes (Crewell and Löhnert 2007) and increase up to 1.6 g m^{-3} in more humid environments (Löhnert et al. 2009; Zhang et al. 2018). The simulation of MWR observations was performed with the RTTOV-gb model (De Angelis et al. 2016). RTTOV-gb is based on the original RTTOV that was adapted to handle ground-based

MWR observations and uses 101 pressure levels, which are selected for the ground-based perspective so that they are denser near the ground than those used by RTTOV itself.

The HATPRO measurement vector in this work consists of 30 brightness temperatures: zenith observations at 14 frequencies and additional nonzenith measurements (zenith angles 60°, 70.8°, 75.6°, and 78.6°) at the four opaque frequencies in the V band. The simulated brightness temperature was perturbed with normally distributed random errors in the range of 0.2–0.5 K for zenith and of 0.2 K for scanning observations (Löhnert et al. 2009).

c. *Assumptions made by simulation of satellite and ground-based observations*

Temporal and spatial matching of observations is a crucial and challenging issue in combining and comparing satellite and ground-based sensors. Especially, the differences in the atmospheric column sampled by ground-based and satellite sensors can lead to significant representativeness errors. However, the aim of this work is to show the synergy potential of both sensors in the best possible setting. Therefore, both sensors are assumed to sample the same atmospheric scene. Further, the atmosphere is assumed to be horizontally homogeneous and aerosol free, and the wavelength dependence of diffraction is ignored so that the geostationary sensor samples the same volume of air at all channels. Furthermore, using the emissivity atlas (monthly mean values) and the assumption of constant trace gas profiles leads to the underestimation of the variability of satellite observations. Therefore, for IRS and combined IRS + MWR products, this study shows rather the best possible results, which would most likely not be achieved by real observations.

4. Methods

a. *Neural network retrieval of CAPE and lifted index fields*

In the following, we explain the NN retrievals developed to derive CAPE/LI values from the observations described above. NN approaches have found extensive application in the retrieval of atmospheric properties and cloud characteristics from both satellite (Aires et al. 2002; Boucher et al. 2023) and ground-based microwave observations (Marke et al. 2016; Cadeddu et al. 2009; Jacob et al. 2019). The NNs reproduce a connection between a set of input (simulated observations) and output (CAPE/LI values) vector pairs. Multilayer perceptrons, a basic type of neural network architecture, consist of multiple layers of neurons (input, output, and hidden layers), where each neuron of one layer is connected to neurons in the next layer by weighted links. The NN is characterized by its inputs, outputs, the number of hidden layers, and the number of neurons (nodes) in each hidden layer. Studies have shown that an NN featuring a single hidden layer with a sufficient number of nodes and a nonlinear activation function can effectively reproduce any nonlinear statistical relationship (Hornik et al. 1989).

Each hidden layer node calculates a sum of the weighted input parameters and bias and applies an activation function, which introduces nonlinearity into the NN. In this study, a hyperbolic tangent sigmoid (tansig) activation function was used in the hidden layers and the linear function in the output layer.

Starting from the input vector, the output of the hidden layer is calculated and propagated to the output layer. The calculated output value is compared to the expected output value, and the loss function is calculated (here, the mean-square error). The error is backpropagated by calculating the derivative of the error value with respect to each weight. Subsequently, all weights are updated by the subtraction of the derivative to reduce the error. The training of the NN is performed iteratively until the desired accuracy is reached. Here, the backpropagation of the errors was performed according to the Levenberg–Marquardt algorithm in the MATLAB neural network toolbox (Hagan and Menhaj 1994).

In this study, the simulated observations of IRS, MWR, and combination IRS + MWR (input), along with the corresponding values of CAPE and LI (output), were used for the training of NNs. In addition to observations, the height above sea level (SLH) in the considered grid points was included as an input parameter (Fig. 1), allowing to localize the NN behavior for a specific grid point.

The size of the input layer of the NNs is determined by the number of channels of the instrument. Thereby, the MWR observation vector consists of the zenith measurements from 14 MWR channels, additionally of the four channels scanning at four elevation angles each, and the SLH, resulting in 31 nodes in the input layer. For the IRS, the first 50 principal components were used. The output layer consists of one neuron (CAPE or LI). To find the optimal network configuration for each parameter and sensor, the training needs to be performed for a varying number and size of hidden layers. Based on the experience gained in the first part of the work (Toporov and Löhnert 2020), four network configurations were considered: either one or two hidden layers, each containing 10 or 20 neurons.

For training and validation of the NNs, a subset of 15 observations and corresponding atmospheric profiles per grid point and month were randomly selected from the entire dataset, resulting in a set of about 22.8×10^4 observations (approximately 2% of the entire dataset). With a threshold of 0.02 kg m^{-2} for the liquid water path, approximately 60% of the data are cloudy. This subset of observations was divided into training (75%) and validation (25%) datasets. The remaining 98% of the data were held out as an independent test set.

To prevent overfitting and achieve good generalization, that is, good performance of the NN not only on the training dataset but also on new, "unseen" data, we applied the early stopping technique (Prechelt 1998). For this, during the training, the NN is applied to the training and validation sets. When the validation error increases for several iteration steps, the training is stopped, and the NN properties (weights and biases) at the minimum of the validation error are returned and stored. After training, the NN is applied to the test dataset to ensure the performance of the networks on independent data.

The performance of all networks was evaluated in terms of correlation (CORR) and root-mean-square errors (RMSEs) of the retrieved parameters with respect to the reanalysis

("truth"). The comparisons showed that the NN with one hidden layer of 20 neurons is sufficient for the retrieval of LI. The CAPE retrieval was found to be slightly more accurate when using the NN with two hidden layers containing 20 neurons.

These best NN configurations were applied to 2 years of simulated IRS, MWR, and combined IRS + MWR observations in each grid point of the target domain to obtain time series of CAPE/LI fields. The performance of these retrievals will be shown in section 5a for the entire dataset and in section 6a for clear-sky and cloudy conditions separately.

b. *Spatial statistical interpolation with least squares estimation*

To merge the fields of CAPE/LI retrieved from IRS observations and CAPE/LI values obtained from a varying MWR network configuration, we use the statistical interpolation approach. In this section, we introduce the equation for the least squares estimation, also called the best linear unbiased estimator (BLUE) (Park and Zupanski 2022; Kalnay 2002). This equation builds a basis for common data assimilation algorithms such as optimal interpolation (OI), Kalman filter, and 1D-, 3D-, and 4D-Var, which differ in the approach and simplifications applied (Kalnay 2002; Nichols 2010). A primary goal of data assimilation is to find an optimal "analysis" of the state through the statistical combination of observations and prior knowledge about the state, called background or first guess.

Assume the model state to be a two-dimensional field ordered by the grid point, forming a single vector of the length $n = 39 \times 39$:

$$\mathbf{x}^{\mathrm{T}} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}. \tag{1}$$

Here, we assume that the background $\mathbf{x}^b$ is given by the n CAPE/LI values obtained from IRS observations and these are always available at every grid point. The observation vector consists of m CAPE/LI values corresponding to the number of MWRs distributed in the domain:

$$\mathbf{y}^{\mathrm{T}} = (y_1, y_2, \ldots, y_m)^{\mathrm{T}}. \tag{2}$$

The CAPE/LI values of observation vector are calculated using synergistic retrieval IRS + MWR (see also section 5).

The difference between observation and background values $[y_m - H(x_m^b)]$ is called observational increment or innovation. The term H is an operator that maps the model variables to observed variables and performs a spatial interpolation from the model grid to the observation locations. In general, H can be nonlinear (e.g., radiative transfer calculations that provide satellite-observed radiances from temperature and humidity profiles). In this study, the unit of the background is the same as that of observations (J kg^{-1} for CAPE or K for LI), and the grid points of observations coincide with the grid points of background. Therefore, the operator H performs a simple selection of background values at grid points with available MWR observation (and no forward model error needs to be considered). The least squares estimator or BLUE analysis $\mathbf{x}^a$ is defined by the following linear interpolation equation:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}[\mathbf{y} - H(\mathbf{x}^b)], \tag{3}$$

where $\mathbf{K}$ is a matrix of dimension $(n \times m)$ that determines the relative weight of observations and of background in the analysis. The matrix $\mathbf{K}$ depends on the error covariance matrices of the background $\mathbf{P}^b$ and of the observations $\mathbf{R}$ and can be determined from the minimization of the analysis errors at each grid point.

The optimal weight matrix $\mathbf{K}$ that minimizes the analysis error variance and the error covariance matrix of the analysis fields $\mathbf{P}^a$ are given by

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \tag{4}$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b, \tag{5}$$

where $\mathbf{H} = (\partial H/\partial x)_{x^b}$ is the derivative of the forward model operator H with respect to the model state evaluated at the model background state $\mathbf{x}^b$. The detailed derivation of $\mathbf{K}$ [Eq. (4)] and $\mathbf{P}^b$ [Eq. (5)] can be found in Kalnay (2002).

Thus, the BLUE approach combines two sources of information, a background vector (here CAPE/LI fields from IRS observations only) and an observation vector (here, values of CAPE/LI from a network of MWR), weighted by their uncertainties. The weight given to the observation innovation is optimally determined on the basis of background and observation uncertainty to minimize the analysis error variance. The resulting analysis covers areas where MWR observations are not available and is more accurate than background and observation data in terms of RMSE (Kalnay 2002).

A correct assessment of observation and background errors is crucial to the quality of the analysis, since they determine to what extent the analysis will be "pulled" toward the observations or the background. Thereby, the diagonal components of the IRS retrieval covariance matrix $\mathbf{P}^b$ and the combined IRS + MWR retrieval covariance matrix $\mathbf{R}$ (variances) determine the correction of the background in the grid points in which an MWR observation is available. The background covariances (off-diagonal elements of $\mathbf{P}^b$) between different grid points specify how the observed MWR information will be spread from MWR observation points to the neighbor grid points without MWR observation. While the estimation of the background error covariance could be difficult to obtain within "real" data assimilation, in this study, the error covariance matrices can be computed directly from the differences between the retrieved values of CAPE/LI and the reanalysis (truth). One drawback of statistical methods such as neural networks is that they do not provide uncertainty estimates for each particular retrieval, unlike variational methods such as optimal estimation (Rodgers 2000). Consequently, we use constant background and observation error covariance matrices at all times, regardless of factors such as cloudiness.

The calculation of the $\mathbf{K}$ matrix includes the inversion of an $m \times m$ matrix [Eq. (4)], which can be computationally expensive depending on the number of available observations. Therefore, due to the large number of grid points and available observations, the common data assimilation techniques introduce approximations and simplifications to reduce the

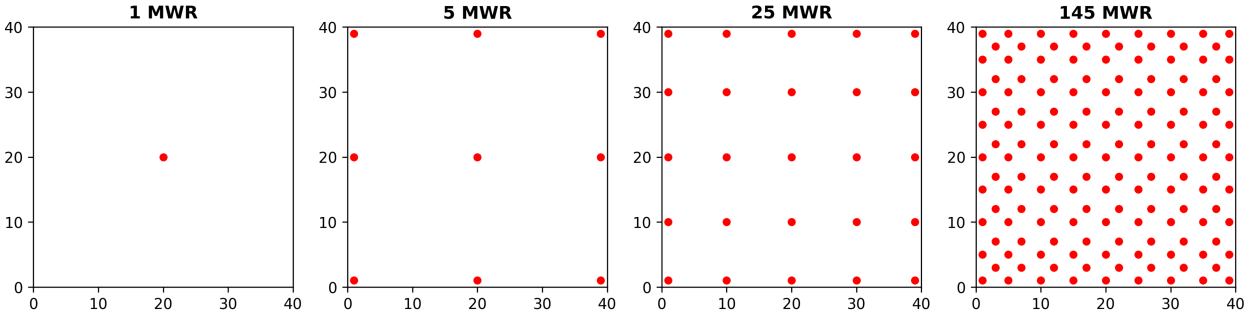

FIG. 2. Possible configurations of the network of ground-based MWR.

computational costs. However, the dimensions of the field used in this study allow the direct use of the least squares method. The assumptions, such as uncorrelated background and observation errors, linearization of the observation operator, unbiased observations, and background, must be fulfilled to ensure the optimal analysis. These aspects will be discussed in the next section.

5. Setup of sensitivity study

In our sensitivity study, CAPE/LI fields retrieved from IRS observations (available at every grid point) serve as the background, and CAPE/LI values derived from the differently configured MWR networks (Fig. 2) are termed observations. The retrieval error variances and covariances were calculated with respect to reanalysis (serving as truth) at each grid point of the domain. The domain covers 39 × 39 grid points (represented as a vector with 1521 values), resulting in $\mathbf{R}$ and $\mathbf{P}^b$ matrices with dimensions 1521 × 1521. In an ideal case, all 1521 grid points would have an MWR observation, which, however, is just considered for theoretical purposes. Thus, the observation covariance matrix $\mathbf{R}$ was adapted depending on the number of MWRs in the network. For example, for a network of three MWR, only the three corresponding diagonal elements (variances) and six nondiagonal elements (covariances) of this matrix are extracted. The covariance matrices of the background $\mathbf{P}^b$ and of observations $\mathbf{R}$ are combined within Eq. (4) to calculate the weight matrix $\mathbf{K}$. The size of the $\mathbf{K}$ matrix depends on the number of MWRs (1521 × number of MWRs).

The statistical interpolation was performed twice: combining IRS-retrieved (available at every grid point) CAPE/LI fields with values of CAPE/LI retrieved using 1) only MWR and 2) synergistic IRS + MWR retrievals. It could be shown that the second approach leads to more accurate analysis than using MWR network values only. In the following, we focus on the interpolation of IRS with IRS + MWR.

The matrix $\mathbf{R}$ corresponds to the error covariance of IRS + MWR retrieval with respect to reanalysis. This could make the assumption of an uncorrelated background and observation error invalid because the IRS observations are included in both background and observation fields. However, the synergistic neural network retrieval utilizes IRS and MWR brightness temperature vectors as one single observation vector and gives each element of this vector a certain weight. It is assumed that elements of IRS measurement vectors are weighted sufficiently different in both IRS and IRS + MWR retrievals, making their errors independent of each other. The posterior error of the interpolated fields was calculated relative to the reanalysis CAPE/LI values as truth.

a. *Error covariance matrices of background (IRS) and observations (IRS + MWR)*

Assuming an MWR placed at all 1521 grid points (i.e., every satellite pixel collocated with an MWR observation), the square root of the diagonal elements of $\mathbf{P}^b$ and of $\mathbf{R}$ gives the uncertainty of the background and of observations in each grid point, respectively (Fig. 3). These values have been derived from the entire dataset including clear-sky and cloudy profiles and demonstrate the performance of the retrievals in each grid point of the domain. Of course, in the case of IRS + MWR, they only show a theoretical state, because it is expected that just a few MWRs will be distributed in the domain. These scenarios are discussed in section 6b. As expected, the IRS + MWR retrieval shows significant improvements compared to the IRS background, reducing the uncertainty of CAPE by on average 30% and of LI by 50%. The spatial variability of the retrieval accuracy is discussed in section 6a.

b. *Impact of a single MWR placed in the domain*

Assuming only one MWR is placed in the domain, the $\mathbf{K}$ matrix (1521 × 1) represents the impact of this single observation on the analysis in all grid points. Varying the location of this single MWR, the average of its impact was calculated in dependency on the distance from the observation point. The average impact of a single MWR is shown in Fig. 3. The average weight given to the observation innovation in the observation point lies around 0.66 and 0.82 for CAPE and LI, respectively. These values describe the influence of an MWR observation on the analysis in the grid point where the observation is performed. Thus, a single MWR observation can be expected to have a stronger impact on the analysis in the observation location for LI than for CAPE. Higher MWR impact for LI results from the higher uncertainty of the IRS compared to the IRS + MWR retrieval (section 5a).

The impact (weight) of the observation innovation on the analysis in surrounding grid points decreases with the distance. For instance, at a distance of 10 grid points (40 km) from the observation location, the impact of a single MWR observation on the CAPE analysis ($\mathbf{K}$ = 0.3) is half times weaker than on the LI analysis ($\mathbf{K}$ = 0.6). The stronger decrease for

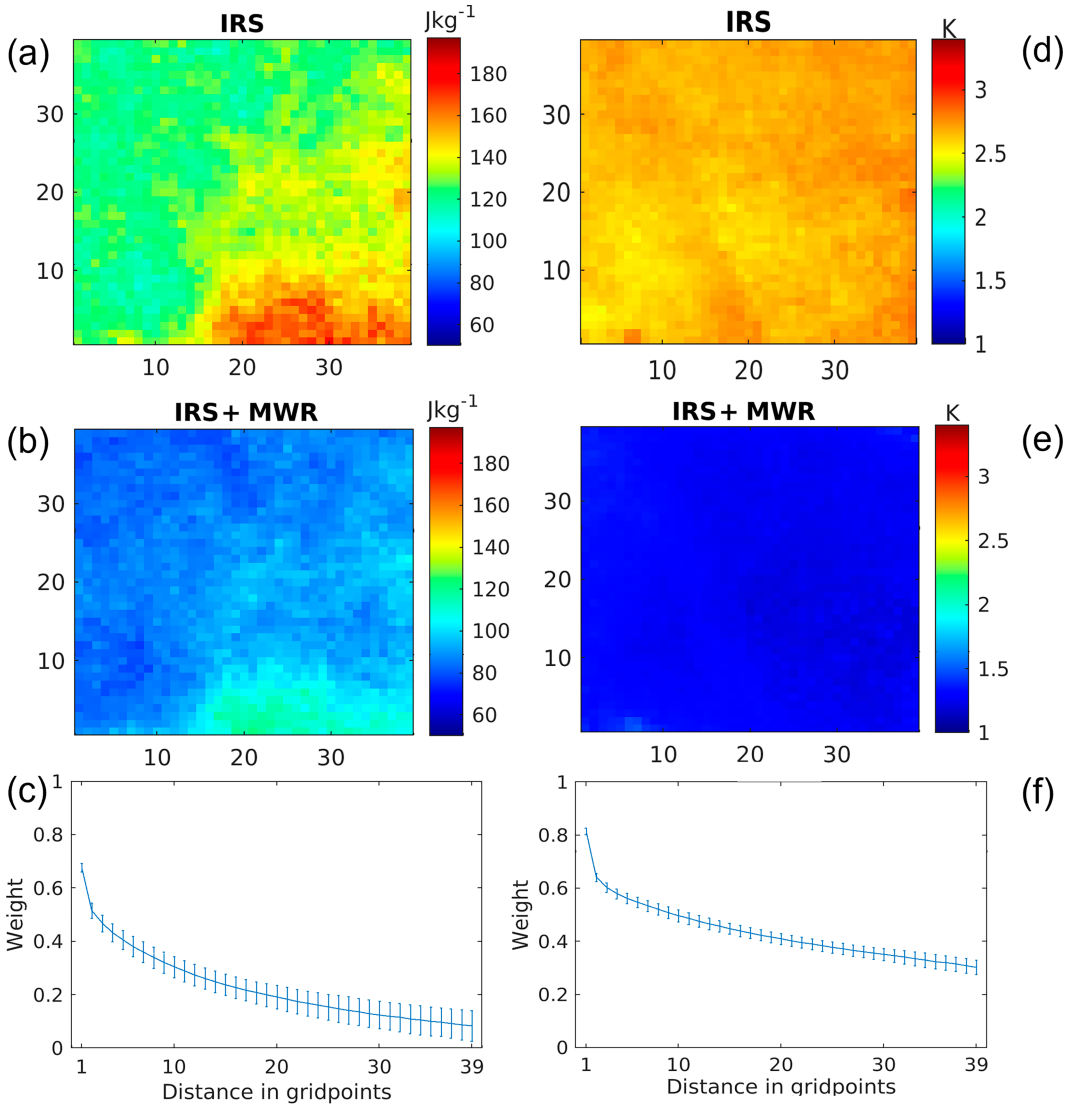

FIG. 3. The standard deviation of IRS and IRS + MWR retrievals in each grid point of the domain for (a),(b) CAPE and (d),(e) LI. These values correspond to the square root of the diagonal elements of covariance matrices $\mathbf{P}^b$ and $\mathbf{R}$. (c),(f) Mean dependence of the weight given to the observation innovation on the distance to the observation point, assuming only one MWR located in the domain. The horizontal grid spacing is 4 km.

CAPE than for LI results from the higher error correlations of the IRS LI retrieval over the domain. Moreover, for LI, the standard deviation of the weights is smaller than for CAPE, indicating that the choice of the location of the MWR is more important for CAPE than for LI.

Generally, a greater effect of MWR can be expected when the sensor is placed in areas where there are large differences between the IRS and the IRS + MWR retrieval uncertainties (e.g., frequently cloudy areas).

6. Results

Section 6a compares the accuracy of the CAPE/LI retrievals from IRS-only, MWR-only, and the combined IRS + MWR observations. It is assumed that both IRS and MWR observations are available simultaneously and at all locations. This allows to assess the pros and cons of the different measurement systems, as well as the benefit of their synergy under clear and cloudy conditions. Then, section 6b presents a sensitivity study for varying number of MWRs located in the target domain and shows the accuracy of the domainwide CAPE/LI obtained using the interpolation method described above.

a. Retrieval performance for domainwide collocated satellite and MWR observations

1) RETRIEVAL PERFORMANCE

A quantitative evaluation of retrieval performance is given by means of Taylor diagrams (Fig. 4) (Taylor 2001). The detailed description of the Taylor diagram can be found in the

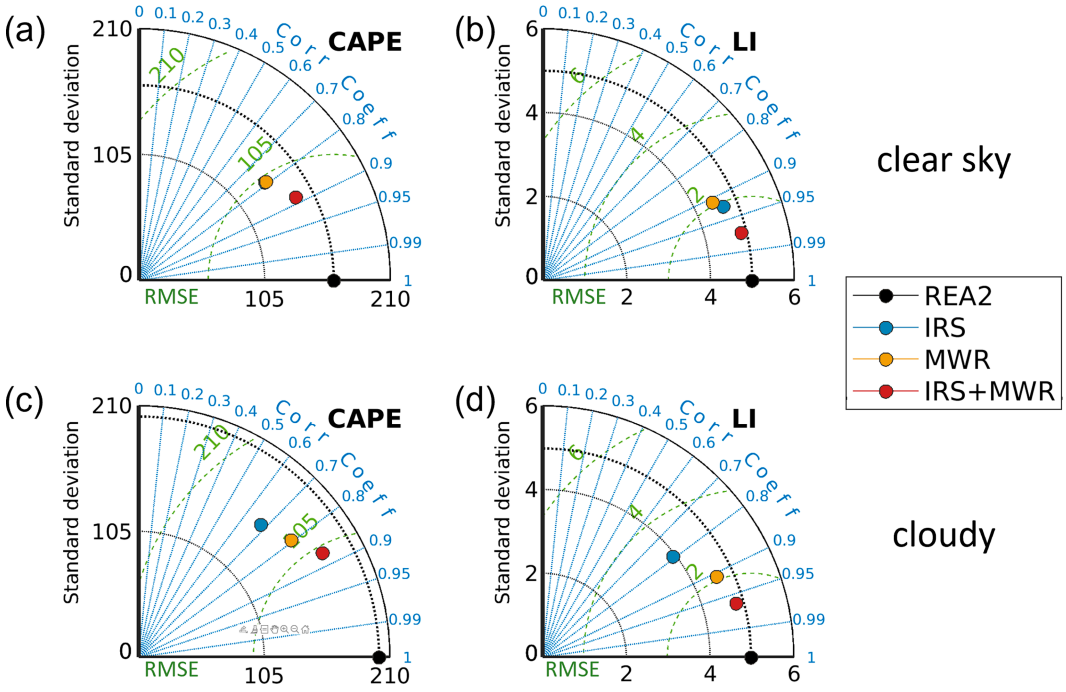

FIG. 4. Taylor diagrams showing the performance of the different CAPE/LI retrievals (IRS, MWR, and IRS + MWR; colored dots) with respect to the reanalysis truth (black dot) for (a),(b) clear-sky and (c),(d) cloudy conditions. Constant values of CAPE/LI standard deviation are given as dotted arcs, whereby the black dot and the corresponding thick dotted arc depict the CAPE/LI standard deviation of the reanalysis datasets. The CORR coefficient between retrieval and reanalysis is shown as blue dotted lines from the origin, whereas green dashed lines indicate the constant centered RMSE. The standard deviation of CAPE and LI is given in joules per kilogram and kelvin, respectively.

appendixes A and B. Here, we only mention that the closer the colored dots (representing the retrievals from different sensors) lie to the black dot (representing reanalysis), the more accurate the retrieval.

Under clear-sky conditions, both single-sensor retrievals MWR and IRS provide similar results for CAPE with a CORR value of 0.79 and RMSE values of 100 J kg^{-1}. For LI, the IRS slightly outperforms the MWR retrieval, with a CORR value of 0.92.

A clear synergy benefit is found when combining IRS and MWR observations: The IRS + MWR retrieval outperforms both single-instrument retrievals with CORR values of 0.89 and 0.97 for CAPE and LI, respectively. The RMSE is reduced by around 23% and 38% compared to IRS retrieval.

Under cloudy conditions, the accuracy of the IRS retrieval degrades for both CAPE (CORR = 0.69) and LI (CORR = 0.89), while MWR accuracy remains largely the same as under clear-sky conditions in terms of CORR. Under cloudy conditions, the synergy benefit is more pronounced than under clear sky. IRS + MWR provides slightly smaller CORR values than under clear-sky conditions but remains the most accurate retrieval with the smallest RMSE values and CORR of 0.87 and 0.96 for CAPE and LI, respectively. This corresponds to an RMSE reduction by 33% and 56% for CAPE and LI, respectively, compared to the IRS retrieval. The accuracy of the IRS retrieval under cloudy conditions depends primarily on the total water and ice content and second on the cloud-top height (CTH). For low clouds with CTH below 2 km and for clouds with liquid/ice water content less than 0.5 kg m^{-2}, the retrieval performs better than for clouds with CTH around 5 km and higher water content.

2) SPATIAL VARIABILITY OF RETRIEVAL PERFORMANCE

To show the spatial variability of retrieval performance, the RMSE and Heidke skill score (HSS) were calculated in each grid point of the domain. HSS is calculated taking into account the exceedance of threshold values for CAPE/LI (see section 2) and includes the number of predicted and not predicted stable and unstable events. It allows the comparison of results based on different datasets such as clear sky and cloudy (Doswell et al. 1990). HSS measures the relative skill, giving the accuracy of the retrieval, relative to that of random chance. The range of HSS is between −1 and 1 with negative values indicating that a retrieval is worse than a randomly generated value. Value 0 means no skill, and a perfect retrieval results in HSS = 1.

The spatial variability of RMSE and HSS of CAPE for clear-sky and cloudy conditions is shown in the left part of Figs. 5 and 6, respectively. Remember that the IRS + MWR retrieval is applied at every grid point of the domain, assuming an MWR is available at every satellite pixel. The RMSE and HSS fields of MWR retrieval are not shown, since we focus on the effect of synergistic (IRS + MWR) observations.

For CAPE, in terms of HSS, the domain-averaged IRS retrieval value degrades from 0.56 under clear skies to 0.42

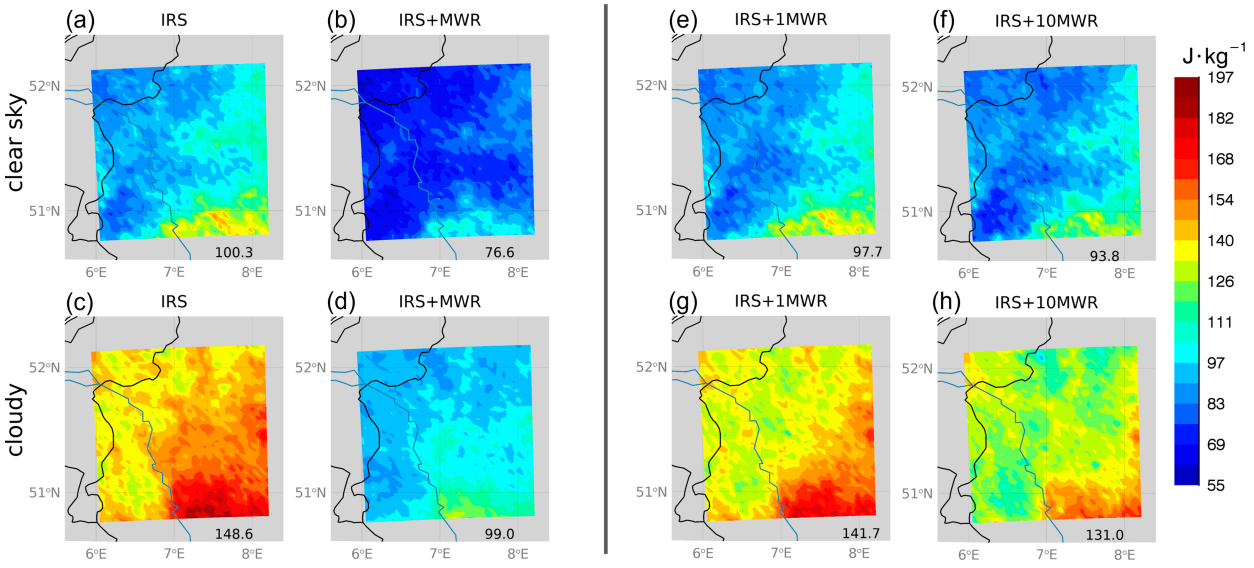

FIG. 5. (a)–(d) RMSE of CAPE derived from IRS and combined IRS + MWR observations for clear-sky and cloudy conditions, assuming an MWR in every satellite pixel. (e)–(h) The RMSE of the analysis based on interpolation between IRS-retrieved CAPE fields and CAPE values from a single MWR (in the center of the domain) and a network of 10 MWRs (as shown in Fig. 1b). The number in the lower-right corner gives the RMSE value for the entire dataset (separated into clear sky or cloudy).

under cloudy conditions, whereby additional MWR observations improve the HSS under both clear-sky (HSS of 0.71) and cloudy (HSS of 0.67) conditions. Under clear-sky conditions, both retrievals, IRS and IRS + MWR, perform slightly better in the southeast part of the domain. Here, the number of unstable events is larger (Fig. 1), probably associated with the topography of the German low mountain range. Higher probability of detection and lower false alarm ratio in this part of the domain result in higher HSS values. The larger RMSE values in this part of the domain (under both clear sky and cloudy conditions) partly reflect higher values and variability of CAPE. Under cloudy conditions, the IRS retrieval shows an almost uniform pattern of low HSS over the domain, which makes it clear that this retrieval has issues in any case if clouds are present. Adding the ground-based MWR, also under cloudy conditions, leads to an improvement of skill especially in the regions with higher CAPE, which can be attributed to the additional temperature and humidity information in the lower troposphere, below the clouds.

For LI, the spatial variability of statistical scores is less pronounced than in the case of CAPE (Fig. 7). Nevertheless, under clear-sky conditions, both retrievals tend to provide slightly higher HSS values over the southeast part of the domain, where the unstable events happen more often. Overall, the IRS retrieval achieves higher HSS values for LI (0.69 and 0.48 for clear sky and cloudy cases, respectively) as in the case of CAPE. Compared to the IRS retrieval, the synergistic IRS + MWR approach leads to 18/64% higher HSS values under clear-sky/cloudy conditions, respectively. The corresponding RMSE is reduced by 38/56% due to the synergistic use of IRS and MWR observations.

b. *Sensitivity study for a varying number of MWR observations*

By applying the spatial statistical interpolation method described in section 4b, we can now analyze the impact of a realistic (section 1) and arbitrary (section 2) MWR network configuration in the target domain on the CAPE/LI analysis performance. For an approaching convective situation, we analyze additionally a time series of CAPE and LI fields, comparing the use of satellite observations only and satellite observations plus a realistic number of MWRs in the domain (section 3).

1) IMPACT OF ADDITIONAL GROUND-BASED OBSERVATIONS ON CAPE AND LI: REALISTIC NETWORK CONFIGURATION

In a first sensitivity study, the impact of adding a microwave radiometer in the center of the domain is analyzed, as well as the deployment of ten MWRs at the E-PROFILE and the DWD Surface Synoptic Observations (SYNOP) stations in the domain (Fig. 1). This states a technically feasible scenario, where the existing measurement infrastructure could be exploited.

The right parts of Figs. 5 and 6 demonstrate the impact of these two measurement configurations on the analysis of CAPE in terms of RMSE and HSS. These figures can be directly compared with the left parts of Figs. 5 and 6 that show the impact of MWRs measuring at every grid point (viz., 1521 MWRs), thus representing a maximum threshold for improvement due to ground-based remote sensing with microwave radiometers.

While in terms of RMSE in the clear-sky case, including up to 10 MWR results in a decrease in the uncertainty of ~6%, the uncertainty reduction is twice as large (12%) in the cloudy

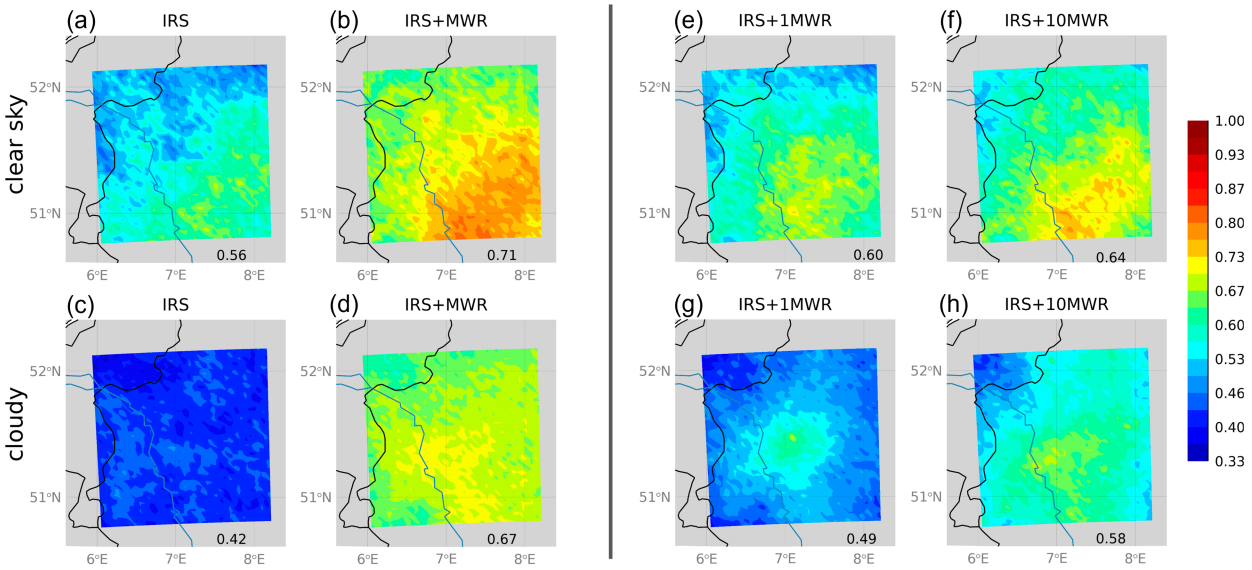

FIG. 6. As in Fig. 5, but illustrating the HSS for CAPE.

case. In the cloudy case, one MWR in the domain can increase the HSS by 0.07, whereas 10 MWRs lead to a skill improvement by 0.16. The influence of adding just one MWR in the cloudy case is especially noticeable in Fig. 6. The relatively low skill across the whole domain in the IRS retrieval is locally significantly enhanced (from 0.4 to 0.7), i.e., at the central location of the MWR.

The corresponding HSS analysis for LI is shown in Fig. 7. As for CAPE, the additional microwave radiometer observations lead to improvements of the RMSE and HSS statistics. While the "perfect" measurement configuration with 1521 MWRs shows a domain-averaged skill of about 0.8 for both clear and cloudy skies, the clear-sky and cloudy-sky analyses with 10 MWRs show lower HSS by only 0.07 and 0.15, respectively. In relation to this, adding 10 MWRs to the satellite-only retrieval leads to an HSS increase of 0.6/0.16 for clear-sky/cloudy conditions, highlighting the significant impact of already 10 MWRs in the domain.

2) IMPACT OF ADDITIONAL GROUND-BASED OBSERVATIONS ON CAPE AND LI: HYPOTHETICAL NETWORK EXTENSION

In a further sensitivity study, the impact of adding up to 200 microwave radiometers in the domain is analyzed following the scheme in Fig. 2. This experiment is only performed to quantify the value of additional instruments, since such a dense

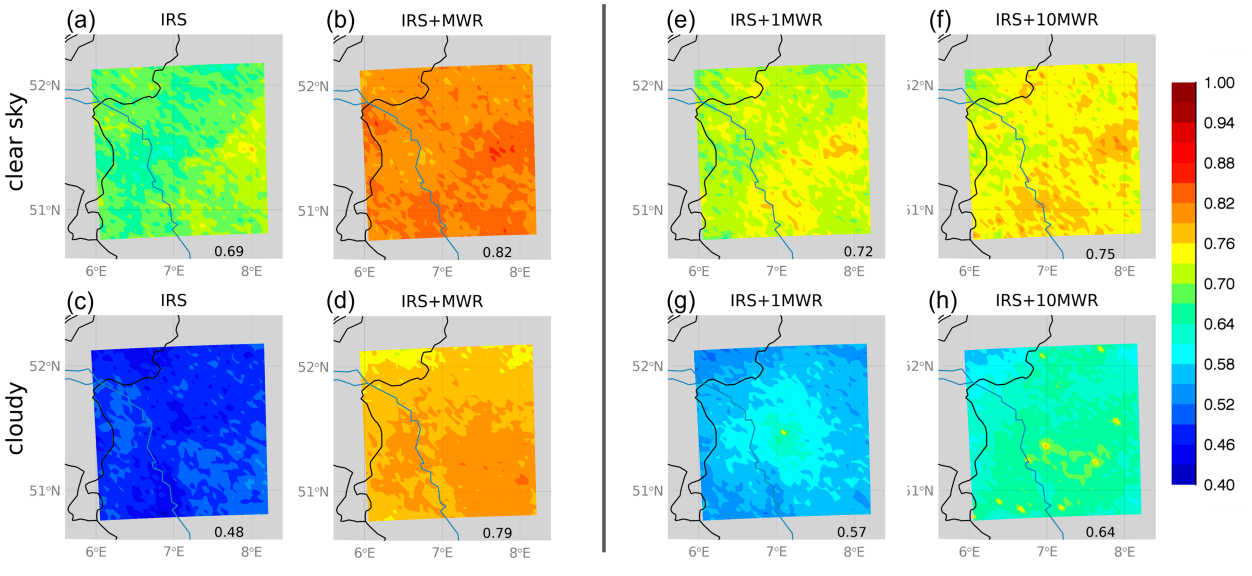

FIG. 7. As in Fig. 5, but illustrating the HSS for LI.

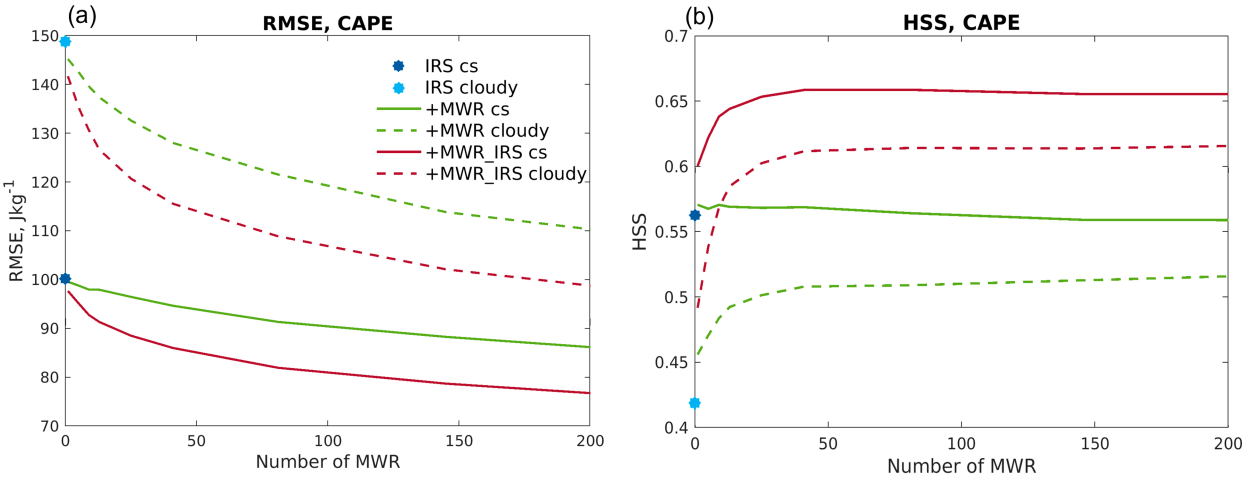

FIG. 8. Dependency of (a) RMSE and (b) HSS of the CAPE analysis (dataset separated into clear sky and cloudy cases) on the number of MWRs distributed in the domain. The blue dots indicate the statistics for the IRS-retrieved CAPE values. Green and red lines correspond to the analysis based on IRS-retrieved fields and MWR network values of CAPE obtained using MWR-only and synergistic IRS + MWR retrievals, respectively.

network is technically hardly feasible. Fig. 8 shows the RMSE and HSS values for CAPE calculated for the entire domain for a network with 0–200 MWR, separated into clear-sky and cloudy conditions. When MWR and IRS sensors are used at the same location in synergy (red lines), the RMSE decreases steadily with a growing number of MWRs, while HSS improves strongly for the first one to 25 MWRs (especially for cloudy conditions) but reaches saturation when more sensors are included. Note that in cloudy conditions, 25 MWRs lead to an increase of almost 0.2 in skill compared to the IRS-only retrieval. In clear-sky cases, this improvement is around 0.1.

The need of using MWR and IRS in synergy can additionally be drawn from Fig. 8. If the MWR-only retrieval is applied to derive CAPE at the MWR locations (green lines) instead of the synergistic IRS + MWR retrieval (red lines), the skill is lower by 0.1 for both clear-sky and cloudy conditions. Specifically, under clear-sky conditions, the added value of the additional ground-based observations is negligible.

The same analysis is carried out for LI (Fig. 9). The retrieval characteristics are very similar to those for CAPE, except that the required number of MWRs in the domain may be a little smaller. The synergistic IRS + MWR retrieval shows a leveling off of HSS already at 10–15 MWRs in the domain, as opposed to 25 for CAPE. This is probably due to the fact that, next to the 2-m temperature observations, only the potential temperature at 500 hPa is used for the calculation of LI. This value is more homogeneous over the domain, compared to the more heterogeneous low-level temperature and humidity profile information required to determine CAPE.

For the network of 25 MWRs, the absolute values of HSS (0.75/0.65) for LI are noticeably higher than those for CAPE (0.65/0.6), which is mostly due to the higher accuracy of the LI retrievals itself (both IRS and IRS + MWR).

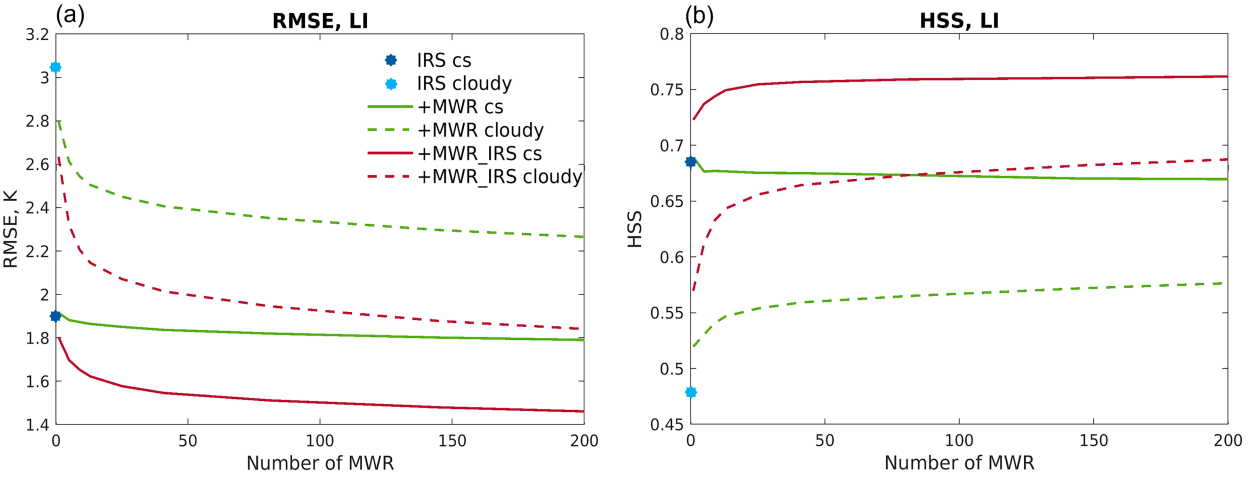

FIG. 9. As in Fig. 8, but illustrating LI.

3) APPLICATION TO A TIME SERIES OF CAPE/LI DEVELOPMENT

(i) CAPE

The time series in Fig. 10 shows the development of CAPE in the period from 1800 UTC 24 August 2011 to 2200 UTC 24 August 2011. During the entire period, the domain is partially covered by liquid clouds (left column) and almost fully covered by thin ice clouds (shadowed area). The time series starts with a strong instability with CAPE values up to 1500 J kg^{-1} in the eastern part of the domain that slightly extends to the west and weakens during the night. The IRS retrieval (third column) can capture only a part of the unstable region but underestimates CAPE even in the absence of clouds (e.g., the southeast part of the domain in Fig. 10). It can be seen that even the presence of thin ice clouds limits the ability of the IRS retrieval to detect high CAPE values.

The interpolation of IRS-retrieved CAPE fields with observations of 10 MWRs (IRS + 10 MWRs analysis) leads to significantly better spatial representation of CAPE at 1800, 1900, and 2000 UTC, especially in the cloudy areas, although the absolute values remain largely underestimated. At 2000–2200 UTC, the performance of the IRS retrieval is very poor and can be only slightly improved by merging with the MWR observations.

Although the MWR addition to the satellite clearly improves the retrieval performance and the synergistic retrieval roughly detects the spatial–temporal trend of CAPE, it clearly misses the finer CAPE structures as given in the reanalysis truth.

(ii) LI

Figure 11 shows the time series of LI between 0800 UTC 21 August 2011 and 1600 UTC 24 August 2011 with a 2-h time step. A different time period as in the case of CAPE has been chosen because after 1600 UTC the LI values over the entire domain lie under the threshold of 1.9 K and the ability of IRS to capture stable conditions can thus not be shown. It can be seen that in the grid points without liquid clouds, the IRS retrieval is mostly able to capture stable (green and yellow) and unstable (blue, below the threshold of 1.9 K) values of LI. In the cloudy grid points, however, the impact of clouds on the IRS retrieval is very strong and leads to an overestimation of LI values (yellow areas in the first two rows and green areas in the last two rows of Fig. 11). Here, the retrieval seems to be very sensitive to cloud-top temperature of the liquid clouds, resulting in too high stability. In the presence of ice clouds, the unstable events can be partly captured by IRS (1400 and 1600 UTC), but some low LI values are overestimated (0800, 1200 UTC).

The impact of 10 MWRs (IRS + 10 MWRs) placed in the domain is different for different times. In the basically liquid-cloud-free situation at 1200 UTC, no impact can be seen. At 1400 and 1600 UTC, liquid clouds strongly influence the IRS retrieval and additional 10 MWR observations lead to a significantly better representation of LI field (considering event/nonevent scores). However, the interpolation of IRS-retrieved field with LI from 10 MWRs (located in a cloudy area) leads to an underestimation of LI in the clear sky southeast of the domain. In this case, a limitation of the radius of influence of single MWR would lead to more accurate results.

The observations of 10 MWRs distributed in the domain lead to improvements in representation of LI in the cloudy grid points. However, the pattern of LI analysis is still strongly influenced by the background (IRS retrieved field). Thus, using the IRS-retrieved LI values in the grid points with liquid clouds without correction is not reasonable, since clouds introduce too high LI values and gradients in the LI field. Assigning higher variances in the background covariance matrix to the IRS observations affected by clouds may mitigate this effect. To improve the IRS-only retrievals, a cloud-clearing method could also be applied to LI field retrieved from IRS observations. For example, the LI values in the grid points with LWP exceeding an empirical threshold, which needs to be determined according to the retrieval accuracy, could be replaced by the LI values in the nearest clear-sky grid point.

7. Summarizing discussion and conclusions

This study evaluates the potential of satellite observations (IRS) and observations of a hypothetical ground-based MWR network for the assessment of atmospheric stability, in terms of CAPE and LI, over a 150 × 150-km domain in the west of Germany.

First, the neural network retrievals of LI and CAPE from IRS and MWR observations, developed in Toporov and Löhnert (2020), were extended to allow for their application to the observations at different locations. In general, the accuracy of retrievals of CAPE and LI for the considered domain is consistent with the results of the previous study. In the presence of clouds, the accuracy of IRS retrievals decreases, while the MWR provides almost the same statistics as under clear-sky conditions. For both indices, the accuracy of the IRS retrieval under cloudy conditions depends on two quantities: the liquid/ice water content and cloud-top height. In general, the IRS retrieval fails to detect convective conditions (high CAPE/low LI values) for cases with water content larger than 0.5 kg m^{-2}.

Under clear-sky conditions, both retrievals, CAPE and LI, benefit from the combination of ground-based and satellite observations. In the presence of clouds, additional ground-based MWR observations are essential and provide information from the atmospheric layers below the cloud, which cannot be captured by the infrared satellite sensor. The synergistic retrieval IRS + MWR results in an HSS around 0.71/0.67 for CAPE (clear sky/cloudy) and around 0.81 for LI (both clear sky and cloudy). The retrieval of CAPE is less accurate compared to LI, even when using synergistic ground-based and satellite observations.

Second, the impact of adding a limited number of ground-based MWR observations to the geostationary IRS measurements was investigated using linear statistical interpolation.

The accuracy of the obtained CAPE/LI fields depends on the error covariances of the background (IRS retrieved fields) and of the additional observations (IRS + MWR retrieval). Combining IRS-retrieved fields with IRS + MWR observed values is beneficial, even if only one ground-based sensor is placed in the center of the domain. Adding 10 MWRs at the

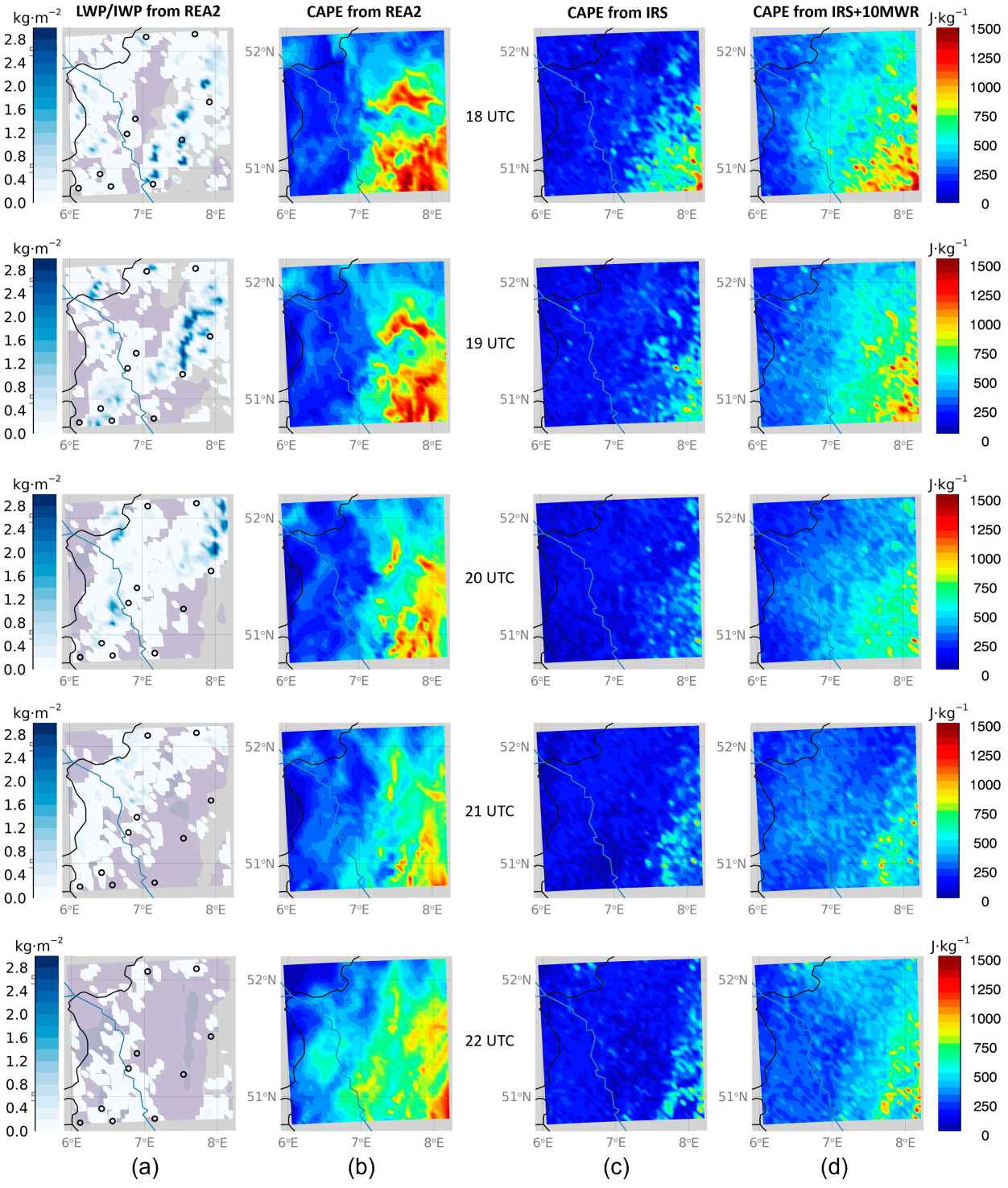

FIG. 10. Time series of CAPE from 1800 UTC 24 Aug 2011 to 2200 UTC 24 Aug 2011. (a) LWP (kg m^{-2}) and ice water path (IWP; shadowed area; lies between 0.001 and 0.7 kg m^{-2}). Black circles indicate the locations of 10 MWRs. (b) CAPE field calculated from reanalysis profiles (truth). (c) CAPE field retrieved from IRS observations. (d) CAPE fields obtained by statistical interpolation of IRS-retrieved CAPE field with CAPE values from 10 MWRs (location of MWRs as shown in Fig. 2).

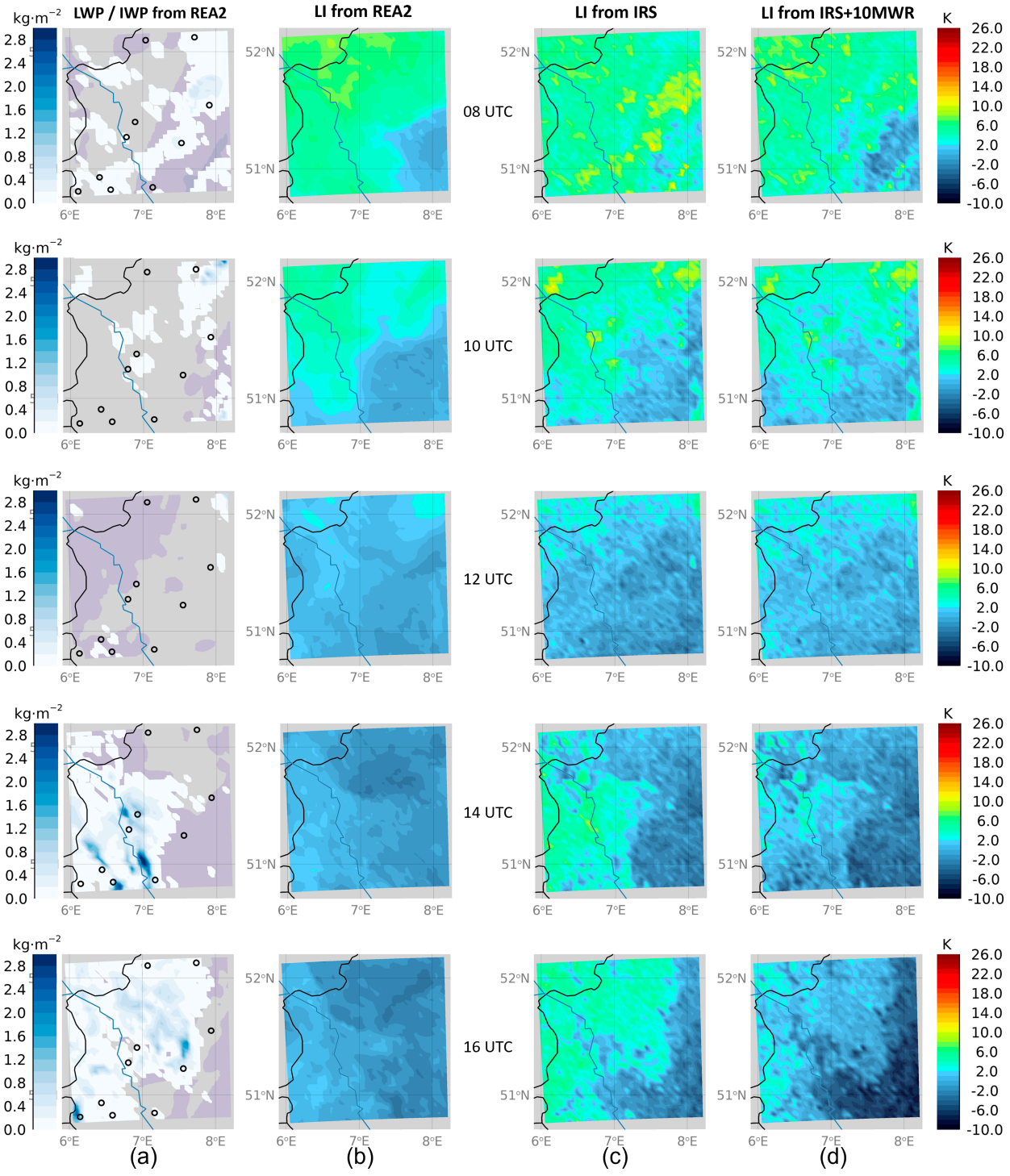

FIG. 11. Time series of LI from 0800 UTC 24 Aug 2011 to 1600 UTC 24 Aug 2011, with 2-h temporal resolution. (a) LWP (kg m^{-2}) and IWP (shadowed area; lies between 0.001 and 0.7 kg m^{-2}). (b) LI field calculated from reanalysis profiles (truth). (c) LI field retrieved from IRS observations. (d) LI fields obtained by statistical interpolation of IRS-retrieved LI field with LI values from 10 MWRs (location of MWR as shown in Fig. 2).

E-PROFILE and DWD SYNOP stations in the domain results in further improvement of the accuracy of CAPE/LI fields. The HSS value of CAPE for the entire domain increases to 0.64/0.58 and that of LI grows to 0.75/0.64 under clear/cloudy conditions, respectively. With a growing number of MWRs in the network, an increase in HSS and a decrease in RMSE are significant only for the first 25 MWRs for CAPE and 10–15 MWRs for LI. If we upscale these findings to the entire area of Germany, this would imply a ground-based remote sensing network of more than 100 instruments. For this, current microwave profilers would need to be more affordable and reliable to operate within a network.

The analysis of the time series of the CAPE and LI analysis fields reveals the weak points of the method.

For CAPE, the main problem lies in the accuracy of the retrievals itself. Further improvements to CAPE retrieval could be achieved by more precise selection of IRS channels sensitive to surface temperature and humidity, by inclusion of scanning MWR observations at lower elevation angles, or by inclusion of in situ measurements from surface stations. We hypothesize that the retrieval would perform better for the mixed-layer CAPE (MLCAPE) instead of the most unstable CAPE used in this study. For the calculation of MLCAPE, the mean conditions in the lowest 100-hPa layer are required. As it was shown by Blumberg et al. (2017), the ML parcels are better captured by passive sensors as the most unstable (MU) parcel.

In the case of LI, the inclusion of surface measurements would also be beneficial. However, more attention should be given to the observation error covariance matrix used in statistical interpolation. Assigning larger error variances to cloud-affected IRS observations could give more weight to more accurate MWR observations.

It is important to note that the results are sensitive to assumptions made concerning observation errors (in space of brightness temperature). For IRS observations, diffraction is not taken into account, and the horizontal inhomogeneity of the atmosphere was neglected. The training, validation, and testing of the neural networks were performed with a 10-month dataset over a certain region. This limits the applicability of the networks to different geographic locations and atmospheric conditions. Further assumptions that would not be fulfilled in the real atmosphere are that the sensors first measure the same air column and second have no representation errors because of the mismatch between observation locations and model grid points. Therefore, the results of this study represent the theoretical evaluation of the synergy of satellite and ground-based sensors in the best possible setting.

Finally, we conclude that significant improvements in the assessment of stability could be achieved if IRS satellite observations were complemented by ground-based MWR observations at existing observational sites. To close the observational gap in the atmospheric boundary layer, we are currently extending our approach to obtain temperature and humidity profiles, which are particularly important in nowcasting and short-term forecast applications.

Acknowledgments. This study was carried out within the Extramural Research Programme of the German Weather Service (Grant 2015EMF-09) as well as within the Hans-Ertel-Centre for Weather Research (Grant 4818DWD5B) funded by the German Federal Ministry for Transport. The ideas of this research have been inspired by the WWRP Working Group "Data Assimilation and Observing Systems" (DAOS) and Cloud and Precipitation Exploration Laboratory (CPEX-LAB; within the Geoverbund ABC/J; http://www.cpex-lab.de). Collaborative efforts for this research have been supported through the EU COST Action CA18235 "PROBE" (European Cooperation in Science and Technology), a funding agency for research and innovation networks (http://www.cost.eu). Responsibility for the content of the publication lies with the authors.

Data availability statement. The whole COSMO-REA2 dataset is archived in the ECMWF file storage system (ECFS). Data for the selected domain are available at the Institute for Geophysics and Meteorology of the University of Cologne.

APPENDIX A

Stability Indices

The CAPE describes the maximum energy available to an ascending air parcel after it has been lifted by some process, such as diabatic heating or forced flow over topography, up to the level of free convection (LFC). Above the LFC, the air parcel becomes warmer than the environment and rises up to the equilibrium level (EL). In this study, the most unstable CAPE is calculated by taking the most unstable parcel within the lowest 300 hPa (MU) and integrating the buoyancy of this parcel from the LFC to the EL (expressed using the virtual temperature of the parcel Tv_p and that of the environment Tv_e):

$$\mathrm{CAPE} = -R_d \int_{p_{\mathrm{MU}}}^{p_{\mathrm{EL}}} (\mathrm{Tv}_p - \mathrm{Tv}_e) d(\ln p), \tag{A1}$$

where $R_d = 287.05\ \mathrm{J\ kg^{-1}}$ is the gas constant for dry air. In general, CAPE values between 0 and 1000 J $\mathrm{kg^{-1}}$ indicate weak instability, and values between 1000 and 2500 J $\mathrm{kg^{-1}}$ and values above 2500 J $\mathrm{kg^{-1}}$ indicate moderate and strong instability, respectively.

The LI describes the degree of instability of the atmosphere below 500 hPa and is calculated as the difference between the temperature of the environment at 500 hPa and the temperature of an air parcel lifted adiabatically from the surface to 500 hPa:

$$\mathrm{LI} = T_{500} - T_{\mathrm{SFC}\rightarrow 500}. \tag{A2}$$

The threshold values for CAPE and LI separating stable and slightly unstable conditions were taken according to Haklander and Van Delden (2003) and lie by 168 J $\mathrm{kg^{-1}}$ and 1.6 K, respectively. The higher/lower the CAPE/LI value, the more unstable the atmosphere and the higher the probability of thunderstorms. The mean values of CAPE and LI for the considered period (Fig. 1) reveal more unstable conditions

(larger number of unstable events and higher/lower CAPE/LI values) in the southeast part of the domain.

APPENDIX B

Taylor Diagram

A quantitative evaluation of the retrieval performance can be given by means of Taylor diagrams (Fig. 4) (Taylor 2001). The Taylor diagram allows the visual comparison of three metrics: the standard deviation of retrieved and true (reanalysis) values, the correlation coefficient between retrieved and true values, and the centered RMSE of retrieved values relative to the truth. The black dot on the horizontal axis shows the CAPE/LI standard deviation of the reanalysis dataset (with correlation of 1 and centered RMSE of 0). The colored dots represent the performance of the CAPE/LI retrievals based on different observations (IRS, MWR, and IRS + MWR). Points lying close to or on the thick black dotted arc indicate a retrieval with the correctly reproduced variability of CAPE/LI. The azimuthal positions of colored points give the correlation coefficients between retrieval and reanalysis. The distance between the dashed green circles and the reanalysis point is equivalent to the centered RMSE between retrieved values and reanalysis. Note that the diagram does not provide information on biases. However, in this study, the retrieval biases were found to be small, leading to negligible differences between RMSE and centered RMSE. Thus, the points lying closest to the reanalysis point indicate retrievals with the highest correlation and lowest RMSE that agree best with the reanalysis.

REFERENCES

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

Blanchard, D. O., 1998: Assessing the vertical distribution of convective available potential energy. *Wea. Forecasting*, **13**, 870–877, https://doi.org/10.1175/1520-0434(1998)013<0870:ATVDOC>2.0.CO;2.

Bloch, C., R. O. Knuteson, A. Gambacorta, N. R. Nalli, J. Gartzke, and L. Zhou, 2019: Near-real-time surface-based CAPE from merged hyperspectral IR satellite sounder and surface meteorological station data. *J. Appl. Meteor. Climatol.*, **58**, 1613–1632, https://doi.org/10.1175/JAMC-D-18-0155.1.

Blumberg, W. G., T. J. Wagner, D. D. Turner, and J. Correia Jr., 2017: Quantifying the accuracy and uncertainty of diurnal thermodynamic profiles and convection indices derived from the atmospheric emitted radiance interferometer. *J. Appl. Meteor. Climatol.*, **56**, 2747–2766, https://doi.org/10.1175/JAMC-D-17-0036.1.

Bollmeyer, C., and Coauthors, 2015: Towards a high-resolution regional reanalysis for the European CORDEX domain. *Quart. J. Roy. Meteor. Soc.*, **141**, 1–15, https://doi.org/10.1002/qj.2486.

Borbas, E. E., and B. C. Ruston, 2010: The RTTOV UWiremis IR land surface emissivity module. Associate Scientist Mission Rep. AS09-04, Doc. NWPSAF-MO-VS-042, 25 pp.

Boucher, E., F. Aires, and V. Pellet, 2023: Towards a new generation of artificial-intelligence-based infrared atmospheric sounding interferometer retrievals of surface temperature: Part I—Methodology. *Quart. J. Roy. Meteor. Soc.*, **149**, 1180–1196, https://doi.org/10.1002/qj.4447.

Burghardt, B. J., C. Evans, and P. J. Roebber, 2014: Assessing the predictability of convection initiation in the high plains using an object-based approach. *Wea. Forecasting*, **29**, 403–418, https://doi.org/10.1175/WAF-D-13-00089.1.

Cadeddu, M. P., D. D. Turner, and J. C. Liljegren, 2009: A neural network for real-time retrievals of PWV and LWP from Arctic millimeter-wave ground-based observations. *IEEE Trans. Geosci. Remote Sens.*, **47**, 1887–1900, https://doi.org/10.1109/TGRS.2009.2013205.

Caumont, O., and Coauthors, 2016: Assimilation of humidity and temperature observations retrieved from ground-based microwave radiometers into a convective-scale NWP model. *Quart. J. Roy. Meteor. Soc.*, **142**, 2692–2704, https://doi.org/10.1002/qj.2860.

Cimini, D., and Coauthors, 2020: Towards the profiling of the atmospheric boundary layer at European scale—Introducing the COST action PROBE. *Bull. Atmos. Sci. Tech.*, **1**, 23–42, https://doi.org/10.1007/s42865-020-00003-8.

Coopmann, O., 2022: IRS OSSE experiments in AROME—13th meeting IRS MAG—EUMETSAT. EUMETSAT NWP SAF, 26 pp., https://www.eumetsat.int/media/50323.

Crewell, S., and U. Löhnert, 2007: Accuracy of boundary layer temperature profiles retrieved with multifrequency multiangle microwave radiometry. *IEEE Trans. Geosci. Remote Sens.*, **45**, 2195–2201, https://doi.org/10.1109/TGRS.2006.888434.

De Angelis, F., D. Cimini, J. Hocking, P. Martinet, and S. Kneifel, 2016: RTTOV-gb—Adapting the fast radiative transfer model RTTOV for the assimilation of ground-based microwave radiometer observations. *Geosci. Model Dev.*, **9**, 2721–2739, https://doi.org/10.5194/gmd-9-2721-2016.

Doms, G., and M. Baldauf, 2018: A description of the nonhydrostatic regional COSMO-model, Part 1: Dynamics and numerics. COSMO Doc., 167 pp.

——, and Coauthors, 2011: A description of the nonhydrostatic regional COSMO model, Part 2: Physical parameterization. COSMO Doc., 161 pp.

Doswell, C. A., III, R. Davies-Jones, and D. L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **5**, 576–585, https://doi.org/10.1175/1520-0434(1990)005<0576:OSMOSI>2.0.CO;2.

Ebell, K., E. Orlandi, A. Hünerbein, U. Löhnert, and S. Crewell, 2013: Combining ground-based with satellite-based measurements in the atmospheric state retrieval: Assessment of the information content. *J. Geophys. Res. Atmos.*, **118**, 6940–6956, https://doi.org/10.1002/jgrd.50548.

EUMETSAT, 2013: ATBD for the MSG GII/TOZ Product. EUMETSAT Doc. EUM/MET/DOC/11/0247, 32 pp., https://user.eumetsat.int/s3/eup-strapi-media/pdf_met_atbd_msg_gii_toz_5988cacb60.pdf.

——, 2018: MTG End-User Requirements Document [EURD]. EUMETSAT DOC. EUM/MTG/SPE/07/0036, 173 pp., https://user.eumetsat.int/s3/eup-strapi-media/pdf_mtg_eurd_d7b503ecc4.pdf.

——, 2020a: Hyperspectral sounding for severe storm forecasting. ESSL Rep. 2020–01, 101 pp., https://www.eumetsat.int/media/47966.

——, 2020b: Study for IRS retrievals and applications at high Satellite Zenith Angle. Final Study Rep., 195 pp., https://www.eumetsat.int/media/47689.

Guedj, S., V. Guidard, B. Ménétrier, J.-F. Mahfouf, and F. Rabier, 2014: Future benefits of high-density radiance data from MTG-IRS in the AROME fine-scale forecast model Final Report. Météo-France & CNRS/CNRM-GAME Res. Rep., 33 pp., https://hal-meteofrance.archives-ouvertes.fr/meteo-01133380.

Hagan, M., and M. Menhaj, 1994: Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Network*, **5**, 989–993, https://doi.org/10.1109/72.329697.

Haklander, A. J., and A. Van Delden, 2003: Thunderstorm predictors and their forecast skill for the Netherlands. *Atmos. Res.*, **67–68**, 273–299, https://doi.org/10.1016/S0169-8095(03)00056-5.

Henderson, D. S., J. A. Otkin, and J. R. Mecikalski, 2021: Evaluating convective initiation in high-resolution numerical weather prediction models using GOES-16 infrared brightness temperatures. *Mon. Wea. Rev.*, **149**, 1153–1172, https://doi.org/10.1175/MWR-D-20-0272.1.

Holmlund, K., and Coauthors, 2021: Meteosat Third Generation (MTG): Continuation and innovation of observations from geostationary orbit. *Bull. Amer. Meteor. Soc.*, **102**, E990–E1015, https://doi.org/10.1175/BAMS-D-19-0304.1.

Hornik, K., M. Stinchcombe, and H. White, 1989: Multilayer feedforward networks are universal approximators. *Neural Networks*, **2**, 359–366, https://doi.org/10.1016/0893-6080(89)90020-8.

Jacob, M., F. Ament, M. Gutleben, H. Konow, M. Mech, M. Wirth, and S. Crewell, 2019: Investigating the liquid water path over the tropical Atlantic with synergistic airborne measurements. *Atmos. Meas. Tech.*, **12**, 3237–3254, https://doi.org/10.5194/amt-12-3237-2019.

Kalnay, E., 2002: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp., https://doi.org/10.1017/CBO9780511802270.

Koenig, M., and E. de Coning, 2009: The MSG global instability indices product and its use as a nowcasting tool. *Wea. Forecasting*, **24**, 272–285, https://doi.org/10.1175/2008WAF2222141.1.

Löhnert, U., and O. Maier, 2012: Operational profiling of temperature using ground-based microwave radiometry at Payerne: Prospects and challenges. *Atmos. Meas. Tech.*, **5**, 1121–1134, https://doi.org/10.5194/amt-5-1121-2012.

——, D. D. Turner, and S. Crewell, 2009: Ground-based temperature and humidity profiling using spectral infrared and microwave observations. Part I: Simulated retrieval performance in clear-sky conditions. *J. Appl. Meteor. Climatol.*, **48**, 1017–1032, https://doi.org/10.1175/2008JAMC2060.1.

Loveless, D. M., T. J. Wagner, R. O. Knuteson, D. D. Turner, and S. A. Ackerman, 2022: Information content of a synergy of ground-based and space-based infrared sounders. Part I: Clear-sky environments. *J. Atmos. Oceanic Technol.*, **39**, 771–787, https://doi.org/10.1175/JTECH-D-21-0119.1.

Marke, T., K. Ebell, U. Löhnert, and D. D. Turner, 2016: Statistical retrieval of thin liquid cloud microphysical properties using ground-based infrared and microwave observations. *J. Geophys. Res. Atmos.*, **121**, 14 558–14 573, https://doi.org/10.1002/2016JD025667.

Markowski, P., and Y. Richardson, 2010: Convection initiation. *Mesoscale Meteorology in Midlatitudes*, Series on Advancing Weather and Climate Science, Vol. 3, Wiley-Blackwell, 185–244.

Martinet, P., D. Cimini, F. Burnet, B. Ménétrier, Y. Michel, and V. Unger, 2020: Improvement of numerical weather prediction model analysis during fog conditions through the assimilation of ground-based microwave radiometer observations: A 1D-Var study. *Atmos. Meas. Tech.*, **13**, 6593–6611, https://doi.org/10.5194/amt-13-6593-2020.

McGrath-Spangler, E. L., W. McCarty, N. C. Privé, I. Moradi, B. M. Karpowicz, and J. McCorkel, 2022: Using OSSEs to evaluate the impacts of geostationary infrared sounders. *J. Atmos. Oceanic Technol.*, **39**, 1903–1918, https://doi.org/10.1175/JTECH-D-22-0033.1.

Nichols, N. K., 2010: Mathematical concepts of data assimilation. *Data Assimilation*, W. Lahoz, B. Khattatov, and R. Menard, Eds., Springer, 13–39, https://doi.org/10.1007/978-3-540-74703-1.

Novak, D. R., F. E. Barthold, M. J. Bodner, K. F. Brill, and M. Eckert, 2011: Quantifying extreme rainfall threats at the hydrometeorological prediction center. *24th Conf. on Weather and Forecasting/20th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 14A.4, https://ams.confex.com/ams/91Annual/webprogram/Paper179714.html.

Park, S. K., and M. Zupanski, 2022: *Principles of Data Assimilation*. Cambridge University Press, 398 pp., https://doi.org/10.1017/9781108924238.

Peppler, R. A., 1988: A review of static stability indices and related thermodynamic parameters. Illinois State Water Survey, SWS Miscellaneous Publication MP-104, 94 pp., https://www.ideals.illinois.edu/items/49020.

Pougatchev, N., and Coauthors, 2009: IASI temperature and water vapor retrievals—Error assessment and validation. *Atmos. Chem. Phys.*, **9**, 6453–6458, https://doi.org/10.5194/acp-9-6453-2009.

Prechelt, L., 1998: Automatic early stopping using cross validation: Quantifying the criteria. *Neural Networks*, **11**, 761–767, https://doi.org/10.1016/s0893-6080(98)00010-0.

Rodgers, C. D., 2000: *Inverse Methods for Atmospheric Sounding: Theory and Practice*. Series on Atmospheric, Oceanic and Planetary Physics, Vol. 2, World Scientific, 256 pp., https://doi.org/10.1142/3171.

Rose, T., S. Crewell, U. Löhnert, and C. Simmer, 2005: A network suitable microwave radiometer for operational monitoring of the cloudy atmosphere. *Atmos. Res.*, **75**, 183–200, https://doi.org/10.1016/j.atmosres.2004.12.005.

Saunders, R., and Coauthors, 2018: An update on the RTTOV fast radiative transfer model (currently at version 12). *Geosci. Model Dev.*, **11**, 2717–2737, https://doi.org/10.5194/gmd-11-2717-2018.

Schmit, T. J., J. Li, J. Li, W. F. Feltz, J. J. Gurka, M. D. Goldberg, and K. J. Schrab, 2008: The GOES-R advanced baseline imager and the continuation of current sounder products. *J. Appl. Meteor. Climatol.*, **47**, 2696–2711, https://doi.org/10.1175/2008JAMC1858.1.

Simmer, C., and Coauthors, 2016: HErZ: The German Hans-Ertel Centre for Weather Research. *Bull. Amer. Meteor. Soc.*, **97**, 1057–1068, https://doi.org/10.1175/BAMS-D-13-00227.1.

Stengel, M., P. Undén, M. Lindskog, P. Dahlgren, N. Gustafsson, and R. Bennartz, 2009: Assimilation of SEVIRI infrared radiances with HIRLAM 4D-var. *Quart. J. Roy. Meteor. Soc.*, **135**, 2100–2109, https://doi.org/10.1002/qj.501.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, https://doi.org/10.1029/2000JD900719.

Toporov, M., 2021: Combining satellite observations with a virtual ground-based remote sensing network for monitoring atmospheric stability. Ph.D. thesis, University of Cologne, 132 pp., https://kups.ub.uni-koeln.de/53399/.

——, and U. Löhnert, 2020: Synergy of satellite- and ground-based observations for continuous monitoring of atmospheric stability, liquid water path, and integrated water vapor: Theoretical evaluations using reanalysis and neural networks. *J. Appl. Meteor. Climatol.*, **59**, 1153–1170, https://doi.org/10.1175/JAMC-D-19-0169.1.

Vidot, J., A. J. Baran, and P. Brunel, 2015: A new ice cloud parameterization for infrared radiative transfer simulation of cloudy radiances: Evaluation and optimization with IIR observations and ice cloud profile retrieval products. *J. Geophys. Res. Atmos.*, **120**, 6937–6951, https://doi.org/10.1002/2015JD023462.

Vural, J., and Coauthors, 2024: Improving the representation of the atmospheric boundary layer by direct assimilation of ground-based microwave radiometer observations. *Quart. J. Roy. Meteor. Soc.*, **150**, 1012–1028, https://doi.org/10.1002/qj.4634.

Wahl, S., C. Bollmeyer, S. Crewell, C. Figura, P. Friederichs, A. Hense, J. D. Keller, and C. Ohlwein, 2017: A novel convective-scale regional reanalysis COSMO-REA2: Improving the representation of precipitation. *Meteor. Z.*, **26**, 345–361, https://doi.org/10.1127/metz/2017/0824.

Zhang, J., P. Zuidema, D. D. Turner, and M. P. Cadeddu, 2018: Surface-based microwave humidity retrievals over the equatorial Indian Ocean: Applications and challenges. *J. Appl. Meteor. Climatol.*, **57**, 1765–1782, https://doi.org/10.1175/JAMC-D-17-0301.1.