Received: 23 July 2024 | Revised: 11 December 2024 | Accepted: 12 December 2024

DOI: 10.1002/qj.4930

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Improved weather predictions through data assimilation for GFDL SHiELD

Mingjing Tong[1] | **Linjiong Zhou**[2] | **Kun Gao**[2] | **Lucas Harris**[1] | **Alex Kaltenbaugh**[1] | **Xi Chen**[2] | **Baoqiang Xiang**[3]

[1]NOAA/OAR/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA

[2]Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, New Jersey, USA

[3]University Corporation for Atmospheric Research, Boulder, Colorado, USA

Correspondence
Mingjing Tong, NOAA/OAR/Geophysical Fluid Dynamics Laboratory, Princeton, NJ 08540, USA.
Email: mingjing.tong@noaa.gov

Present address
Xi Chen, Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

Funding information
NOAA Research

Abstract

The Geophysical Fluid Dynamics Laboratory (GFDL)'s System for High-resolution prediction on Earth-to-Local Domains (SHiELD) model typically uses the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) analysis to initialize its medium-range forecasts. A data assimilation (DA) system has been implemented for the global SHiELD to demonstrate the prediction skills of the model initialized from its own analysis. The DA system leverages the advanced DA techniques used in GFS and assimilates all the observations assimilated in GFS. Compared to the forecasts initialized from GFS analysis, SHiELD forecast skills are significantly improved by using its own analysis. Remarkable improvement is found in the southern hemisphere with positive impact lasting up to 10 days. The DA system is useful in identifying and understanding model errors. The most noticeable model error detected by the DA system originates from the turbulent kinetic energy (TKE)-based moist eddy-diffusivity mass-flux vertical turbulent mixing (TKE-EDMF) scheme. The model error leads to insufficient ensemble spread. Including two versions of the TKE-EDMF scheme in the ensemble helps increase ensemble spread, further improves forecast skills and alleviate the systematic errors in marine stratocumulus regions. Applying interchannel correlated observation errors for Infrared Atmospheric Sounding Interferometer (IASI) and Cross-track Infrared Sounder (CrIS) also reduces the systematic errors and improves the forecast skill up to day 5. Further investigation of the forecast errors reveals that the ensemble spread is largely affected by the parameterization of eddy diffusivity through its impact on the gradient of the model state. The systematic forecast errors in marine stratocumulus regions are associated with the vertical location of the stratocumulus cloud, which is sensitive to model vertical resolution within the cloud layer. Enhancing eddy diffusion within the cloud or near cloud top elevates cloud top but reduces cloud amount.

KEYWORDS

data assimilation, ensemble spread, interchannel correlated observation error, model error, numerical weather prediction, stratocumulus-topped boundary layer, vertical turbulent mixing

1 | INTRODUCTION

The improvement of numerical weather prediction (NWP) relies on the advances of the forecast models and the improvement in the accuracy of initial conditions facilitated by the increasing computing power. The initial condition (also known as analysis) is created by statistically combining short-range forecasts (also known as first guess or background) with observations via data assimilation (DA) techniques. The quality of the initial condition is therefore determined by the forecast model, observations assimilated and DA methods, which include DA schemes, cycling strategies, pre-processing and quality control of observations, representing background and observation errors, mapping background to observations via observation operators, accounting for model errors, and enhancing balance in analysis.

System for High-resolution prediction on Earth-to-Local Domains (SHiELD) is an experimental global atmospheric model developed by the Geophysical Fluid Dynamics Laboratory (GFDL). It was originally developed as a prototype of the Next-Generation Global Prediction System for the National Weather Service (NWS) by coupling the non-hydrostatic Finite-Volume Cubed-Sphere Dynamical Core (FV3; Lin & Rood, 1996; Lin, 1997; Lin, 2004; Putman & Lin, 2007; Harris, Chen, et al., 2020) with the Global Forecast System (GFS) physics suite (Harris, Zhou, et al., 2020). The SHiELD model is continuously upgraded with advances in dynamics and physics and with new capabilities developed (Gao et al., 2021; Harris, Zhou, et al., 2020; Mouallem et al., 2023; Zhou et al., 2022; Zhou & Harris, 2022). The model can be initialized from a variety of external analyses, including National Centers for Environmental Prediction (NCEP) GFS analysis and European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) analysis, through a well-designed initialization procedure that accurately preserves the initial conditions while ensuring consistency with the FV3 dynamics (Harris et al., 2021). GFS analysis is typically used to initialize SHiELD forecasts, including its real-time forecast run at GFDL (available at https://www.gfdl.noaa.gov/shield/).

SHiELD (formerly known as fvGFS) has participated in model intercomparison studies aiming to understand the role and relative importance of model formulation and initial condition on medium-range forecast errors in mid-latitudes (Magnusson et al., 2019) and tropical cyclone (TC) prediction (Chen et al., 2019). Magnusson et al. (2019) found that the initial condition plays a major role in forecast root-mean-squared error (RMSE) differences for the evaluated configurations, while the model formulation is the main factor impacting forecast biases for the investigated parameters. Chen et al. (2019) show that both forecast model and initial condition play important roles in TC prediction. They found that the large-scale flow and TC track forecasts are significantly improved using the IFS initial condition compared to the forecasts using GFS initial condition. For the 2017 Atlantic hurricane season, SHiELD track forecasts using IFS analyses are even better than IFS, which in most years has lower Atlantic TC track errors than all other operational models.

In a forecast system run with DA, the forecast model, DA and observational data usage are interdependent (Bauer et al., 2015). A more accurate forecast model produces a more accurate first guess that matches observations better. It allows more observations to be assimilated[i] and improves the accuracy of the analysis, which in turn further improves short-range forecasts (first guesses) within DA cycles. In this study, we implemented a global DA cycling system for SHiELD based on the existing DA capabilities available from GFS (Kleist et al., 2023). This allows us to take advantage of our model advances, which potentially help create a better initial condition and further improve the model's prediction skill upon using GFS initial condition. On the other hand, the cumulative effect of DA cycling and the diagnostic statistics from the DA system facilitate our understanding of the model characteristics and the source of forecast errors, which provides us clues to improve the model.

The evaluation of the SHiELD DA cycling system is based on the comparison with the GFS DA cycling system. In Section 2, we first compare the SHiELD and GFS forecast models and provide some details on our implementation of the SHiELD DA system. Section 3 focuses on the evaluation and improvement for the SHiELD DA system. We explored upgrading observation errors of two infrared instruments, the Infrared Atmospheric Sounding Interferometer (IASI) and the Cross-track Infrared Sounder (CrIS), to interchannel correlated observation errors, and the results will be presented in Section 4. The acronyms of the satellite instruments used in this study are listed in Table 1. Section 5 focuses on investigating the root cause of the analysis and forecast errors identified from the DA experiments. Section 6 summaries the results of this study.

2 | MODEL AND DATA ASSIMILATION SYSTEMS

2.1 | SHiELD and GFS configurations

The FV3 dynamic core was transitioned into the operational GFS in its version 15 (GFSv15) in June 2019. Since then, the operational GFS has shared several common model components with SHiELD, while each model has its own upgrade focus and pace. The SHiELD model

TABLE 1 Acronyms of satellite instruments and observations.

Acronym	Expansion
ABI	Advanced Baseline Imager
AHI	Advanced Himawari Imager
AMV	Atmospheric Motion Vector
AMSU-A	Advanced Microwave Sounding Unit-A
ATMS	Advanced Technology Microwave Sounder
AVHRR	Advanced Very High Resolution Radiometer
CrIS	Cross-track Infrared Sounder
GPS RO	GPS Radio Occultation
GOES	Geostationary Operational Environmental Satellite
IASI	Infrared Atmospheric Sounding Interferometer
MHS	Microwave Humidity Sounder
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SSMIS	Special Sensor Microwave Imager/Sounder

has been upgraded almost every year (Harris, Zhou, et al., 2020). The experiments conducted in this study are based on the 2022 version of SHiELD (SHiELD 2022 hereafter). We constructed the SHiELD DA system based on and compared it with the current operational GFS, which is its version 16 (GFSv16). The cumulus convection schemes (Han et al., 2017), the vertical turbulent mixing scheme (Han & Bretherton, 2019), and the Noah land surface model (Ek et al., 2003) are synchronized with GFSv16 in SHiELD 2022 (Zhou et al., 2022). Although the source codes of the physics packages are synchronized with GFSv16, the actual versions of the physics packages used in SHiELD 2022 could be different, and parameter tuning may be different.

Table 2 lists the major model configuration differences between SHiELD 2022 and GFSv16. Some of these differences impact our development and customization of the DA system, for example, the vertical levels and the scheme used to handle air–sea interaction. A mixed-layer ocean (MLO) based on Polland et al. (1973) was implemented in SHiELD to account for the interactions between atmosphere and ocean. The MLO predicts mixed-layer depth and temperature within the mixed layer with tendencies computed from the net surface heat flux and surface wind stress. The sea surface temperature (SST) is nudged toward observational climatology[ii] plus a slowly decaying initial anomaly (Harris, Chen, et al., 2020; Harris, Zhou, et al., 2020). In GFS, a near-sea surface temperature (NSST) scheme is used to predict the near-sea surface temperature profile between the surface and a reference level at about 5 m through a diurnal warming model and a skin-layer cooling parameterization based on Fairall et al. (1996). The NSST scheme is coupled with an NSST analysis (Li & Derber, 2008) to provide better SST for the calculation of air–sea heat and moisture fluxes, and to be used by the radiative transfer model for satellite radiance assimilation. A major upgrade of SHiELD 2022 is the upgrade of the GFDL microphysics scheme to version 3 (GFDL MP v3; Zhou et al., 2022). This GFDL MP upgrade consists of three aspects: code reorganization, optimization and modularization; using more realistic particle size distributions for cloud water and cloud ice; and the use of climatological aerosol data to calculate cloud droplet number concentration. This GFDL MP upgrade significantly improves global medium-range weather prediction skills as documented in Zhou et al. (2022). Beside the listed major configuration differences, many model parameters are also tuned differently for these two models.

2.2 | SHiELD data assimilation system

A pragmatic way of developing a DA cycling system for SHiELD to be able to compare with the operational GFS is leveraging all advanced DA techniques from GFS. The building blocks for the system are available from the GitHub repositories of GFS. We started from developing a DA interface for the SHiELD model. A significant amount of effort was made on code integration and validation in order to enable capabilities such as incremental analysis update (IAU; Bloom et al., 1996) and stochastic parameterization schemes (e.g. Leutbecher et al., 2017; Palmer et al., 2009), including the stochastically perturbed parameterization tendency scheme (SPPT; Buizza et al., 1999, Palmer et al., 2009; Shutts et al., 2011), the stochastically perturbed boundary layer humidity scheme (SHUM; Tompkins & Berner, 2008), and the Stochastic Kinetic Energy Backscatter scheme (SKEB; Berner et al., 2009). Our workflow was built based on the GFS workflow with many changes made to customize the DA system for SHiELD. Beside DA cycling and forecast-only workflows, we also developed other options such as replay (Orbe et al., 2017; Takacs et al., 2018), ensemble replay (Zhu et al., 2022), ensemble post-processing, forecast departures evaluation for various DA development and assessment purposes.

The SHiELD DA cycling run for this study closely matches GFSv16.2. It is a hybrid 4D ensemble-variational DA with 80 ensemble members (4DEnVar; Kleist & Ide, 2015) based on the Gridpoint Statistical Interpolation (GSI, Kleist et al., 2009) with 4DIAU (Lei & Whitaker, 2016; Lorenc et al., 2015). The static background covariance was estimated using the National Meteorological Center (NMC) method (Parrish & Derber, 1992) by running pairs of 24 and 48 hours SHiELD forecast initialized from operational GFS analyses at C192 (∼50 km) resolution every

TABLE 2 Major configuration differences between SHiELD 2022 and GFSv16.

Configuration	SHiELD 2022	GFSv16
Vertical levels/model top	91/0.66 hPa (∼ 51 km)	127/0.01 hPa (∼ 80 km)
Tracer advection scheme	Positive-definite	Monotonic
Microphysics	Inline GFDL MP v3	Split GFDL MP v1
Vertical turbulent mixing	TKE-EDMF initial version (EDMF0)	TKE-EDMF newer version (EDMF1)
Ocean surface	MLO	NSST
Gravity Wave Physics	GFS orographic gravity wave drag parameterization (used in GFSv15)	CIRES Unified Gravity Wave Physics Scheme

Note: The description of the split and inline microphysics can be found in Appendix B of Harris, Zhou, Lin et al. (2020) and more details can be found in Zhou et al. (2022). The initial and the newer versions of the TKE-based eddy-diffusivity mass-flux (TKE-EDMF) scheme are implemented as two separate options that can be chosen through a namelist parameter. CIRES stands for University of Colorado Cooperative Institute for Research in the Environmental Sciences. More technical details about the GFS physics suite can be found at https://dtcenter.ucar.edu/GMTB/v4.1.0/sci&uscore;doc/index.html.
Abbreviations: EDMF, eddy-diffusivity mass-flux; GFDL, Geophysical Fluid Dynamics Laboratory; GFS, Global Forecast System; SHiELD, System for High-resolution prediction on Earth-to-Local Domains; TKE, turbulent kinetic energy.

five days for a whole year. The ensemble members are updated with a modulated space form of the local ensemble Kalman filter (LETKF; Hunt et al., 2007, Lei et al., 2018). Multiplicative inflation through relaxation to the prior spread (RTPS; Whitaker & Hamill, 2012) is used to increase ensemble spread. The SPPT, SHUM and SKEB schemes are also used to represent model uncertainties in the ensemble. More DA upgrades adopted from GFSv16 can be found at https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/implementations.php.

All observations assimilated in GFSv16.2 are assimilated in SHiELD except a few channels from microwave (MV) and infrared (IR) instruments due to the lower model top of SHiELD. In particular, AMSU-A channel 14 and ATMS channel 15 are not assimilated in SHiELD. IASI channel 16 (648.75 cm^{-1}), 72 (662.75 cm^{-1}), 303 (720.5 cm^{-1}) and CrIS channel 28 (666.875 cm^{-1}), 32 (669.375 cm^{-1}), 115 (721.25 cm^{-1}) are also excluded from assimilation, because very large observation innovation biases were found in those channels during our initial evaluation of the IR radiance assimilation. The interchannel correlated observation errors for IASI and CrIS, which are used in GFSv16, are not applied in most of our SHiELD DA experiments. The estimation of the interchannel correlated observation errors for IASI and CrIS and their impact will be discussed in Section 4. Te GPS RO bending angles up to 45 km are assimilated in SHiELD instead of up to 55 km in GFSv16.

We adopted the GFS surface cycle to update numerous surface fields such as SST, snow cover, sea ice, soil moisture and temperature etc. from real-time analyses or climatological datasets. When the NSST analysis is not performed, which is the case for SHiELD, the surface cycle updates SST from the NCEP Real-Time Global Sea Surface Temperature (RTG SST) analysis product. The RTG SST analysis product created through a 2DVar technique was discontinued on February 11, 2020. While NCEP still generates the RTG SST analysis data based on the NSST foundation temperature (personal communication with Xu Li at NCEP/EMC) every six hours at the main synoptic times, the RTG SST analysis data are identical in 24 hours and change once at 1800 UTC each day. Our initial use of the RTG SST analysis data in the surface cycle led to distinct negative near-surface temperature and moisture biases over northern-hemispheric oceans compared with surface observations. It largely degraded the prediction skill. We modified the surface cycle to update SST from the GFS surface analysis data, which are also available at the beginning of the IAU window (0300, 0900, 1500, and 2100 UTC). Changes were also made to the MLO to incorporate the IAU capability.

3 | SYSTEM EVALUATION AND IMPROVEMENT

3.1 | Experiments

This study aims to demonstrate the medium-range forecast performance of SHiELD initialized from its native analysis obtained through running DA cycling. The reference is SHiELD free forecast initialized from GFS analysis, which is how the model is utilized in our research and real-time forecast. Before running the medium-range forecast, we need to make sure that SHiELD DA cycling works properly through comparison with GFS DA cycling. The operational GFSv16 and SHiELD 2022 real-time forecast are both run at horizontal resolution of C768 (∼13 km). Due to limited computing resources, SHiELD DA experiments

TABLE 3 Names and descriptions of data assimilation, replay and free forecast experiments.

Experiment	Description
GFS_DA	Experimental GFSv16.2 DA cycling
SHiELD_GFSIC_REPLAY	SHiELD replay to GFS_DA IC with the EDMF0 version of the TKE-EDMF scheme (default 2022 SHiELD configuration)
DA_CNTL	SHiELD control DA cycling with the EDMF0 version of the TKE-EDMF scheme
DA_EDMF1	SHiELD DA cycling with the EDMF1 version of the TKE-EDMF scheme
DA_CNTL_DEDMFENS	Same as DA_CNTL, but use both versions (EDMF0 and EDMF1) of the TKE-EDMF scheme in the ensemble
DA_EDMF1_DEDMFENS	Same as DA_EDMF1, but using both versions (EDMF0 and EDMF1) of the TKE-EDMF scheme in the ensemble
DA_EDMF1_DEDMFENS_RCOV	Same as DA_EDMF1_DEDMFENS, but apply interchannel correlated observation errors for IASI and CrIS
SHiELD_GFSIC_EDMF1	SHiELD free forecast initialized from GFS_DA IC using the EDMF1 version of the TKE-EDMF scheme

Abbreviations: DA, data assimilation; EDMF, eddy-diffusivity mass-flux; GFS, Global Forecast System; SHiELD, System for High-resolution prediction on Earth-to-Local Domains; TKE, turbulent kinetic energy.

(Table 3) were run at a reduced horizontal resolution of C192 (~50 km). The horizontal resolution of the ensemble in the DA experiments is C96 (~100 km). For fair comparison, we also ran a GFS DA experiment at the same reduced horizontal resolution (GFS_DA in Table 3). The evaluation of the system at C768 resolution will be conducted in the near future, as we are porting the system to a newly available High Performance Computing system.

Our evaluation of the DA system is based on short-range forecasts within DA cycles. Both initial condition and model formulation play a role in forecast performance. The relative impact of initial condition and model formulation on short-range forecast can be deduced from comparing SHiELD and GFS DA experiments with a SHiELD replay forecast experiment. The replay technique was first developed at the NASA Global Modeling Assimilation Office (GMAO) based on the IAU capability (Orbe et al., 2017). The IAU is an initialization method applying analysis increments gradually into model integration to reduce the imbalance and model shocks resulting from running the model directly from the analysis (Bloom et al., 1996). The replay forecast is similar to the IAU enabled forecast step in DA cycling, except that the IAU tendencies are computed from a pre-existing analysis (Orbe et al., 2017; Takacs et al., 2018). The SHiELD replay experiment (SHiELD_GFSIC_REPLAY in Table 3) uses the GFS_DA analysis to compute the analysis increments relative to the SHiELD forecast. It is equivalent to re-running the GFS_DA forecast step but replacing the GFS model with the SHiELD model. It is also equivalent to re-running the SHiELD DA forecast step but replacing the SHiELD analysis with the GFS_DA analysis. The intercomparison between the DA and replay experiments helped us identify defects in the SHiELD DA system.

The first cycle of all DA and replay experiments was initialized from the operational GFS analysis at 1800 UTC on 9 June 2022. The DA and replay experiments were run until 0000 UTC on 24 August 2022. The first 10 days' spin-up cycles were discarded from evaluation. The SHiELD_GFSIC_REPLAY and the SHiELD control DA experiment (DA_CNTL) use the default SHiELD 2022 model configuration. A couple of more SHiELD DA experiments (DA_EDMF1, DA_CNTL_DEDMFENS and DA_EDMF1_DEDMFENS in Table 3) were run to understand and mitigate model errors associated with the turbulent kinetic energy (TKE)-based moist eddy-diffusivity mass-flux (TKE-EDMF) vertical turbulent mixing scheme (Han & Bretherton, 2019). Another SHiELD DA experiment (DA_EDMF1_DEDMFENS_RCOV) was conducted to demonstrate the impact of using interchannel correlated observation errors for IASI and CrIS, which will be discussed in Section 4. The medium-range forecasts initialized from the DA_EDMF1, DA_EDMF1_DEDMFENS and DA_EDMF1_DEDMFENS_RCOV analyses are compared with SHiELD free forecast initialized from GFS_DA analysis (SHiELD_GFSIC_EDMF1 in Table 3) to demonstrate the impact of using SHiELD native analysis and the improvement in SHiELD DA system.

3.2 | Short-range forecast verified against observations

The short-range forecast of SHiELD DA cycling is evaluated from the background departure (also known as observation innovation, observation minus background or OmB) standard deviations of all assimilated observations. The SHiELD forecast fits observations better than GFS forecast in terms of random errors, if the normalized OmB standard deviations are smaller than 100% in Figures 1 and 2. Both initial condition and forecast model contribute to the OmB differences between the SHiELD DA and GFS DA experiments. To separate the impact of initial condition and forecast model, the OmB standard deviations of SHiELD_GFSIC_REPLAY are also included for non-radiance observations in Figure 1. The OmB standard deviations for radiance observations are not shown for SHiELD_GFSIC_REPLAY, because it is simply a re-forecast experiment and does not generate bias correction coefficients necessary for computing bias-corrected radiance OmBs.

First, we would like to compare SHiELD and GFS short-range forecasts with the same GFS_DA initial condition via the SHiELD_GFSIC_REPLAY and GFS_DA experiments. SHiELD forecasts fit better to conventional wind observations in the middle to upper troposphere (Figure 1a), conventional temperature observations in the lower troposphere (Figure 1b), humidity observations near the surface and above 800 hPa (Figure 1c), satellite AMVs above 600 hPa (Figure 1e), and GPS RO bending angles almost through the whole atmosphere, except the upper stratosphere near SHiELD model top (Figure 1d). SHiELD short-range forecasts are not as good as GFS in terms of the fits to wind observations below 900 hPa (Figure 1a), temperature observations between 250 hPa and 50 hPa (Figure 1b), humidity observations between 900 and 800 hPa (Figure 1c), and AMVs below 600 hPa (Figure 1e).

The impact of initial condition on SHiELD short-range forecast can be revealed by comparing SHiELD_GFSIC_REPLAY with SHiELD DA_CNTL experiments. SHiELD forecast shows significant improvement in the fits to conventional wind observations below 150 hPa and AMVs, when initialized from its own analysis (Figure 1a,e). This indicates that the wind field in SHiELD analysis is better than that in GFS analysis in those observed areas. The quality of the initial condition is partially determined by the forecast model. Where the model forecast is better as shown by the SHiELD_GFSIC_REPLAY vs the GFS_DA in Figure 1, the analysis can also be improved, and the short-range forecast initialized from better analysis is further improved through DA cycles. However, there are situations where

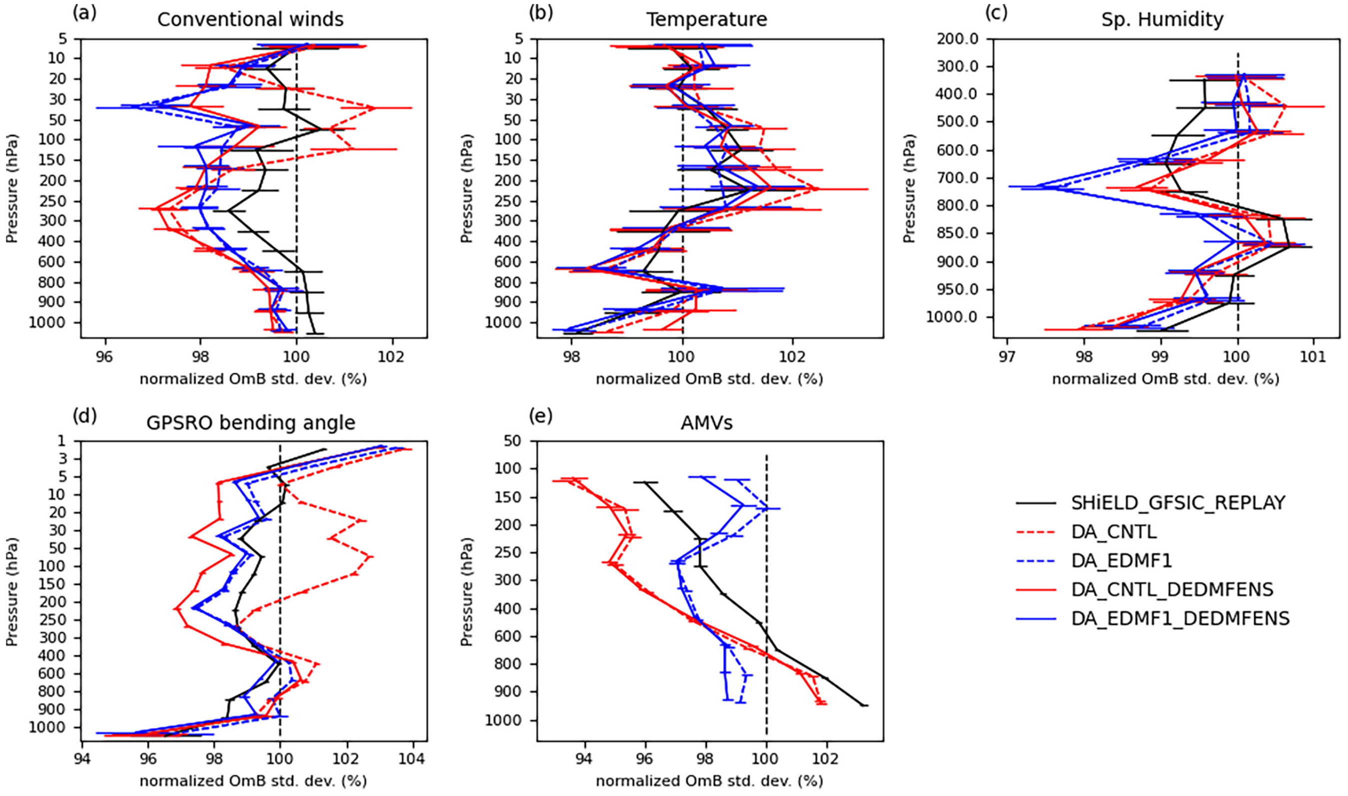

FIGURE 1 Normalized standard deviation of background departures (observation minus background, OmBs) w.r.t the GFS_DA experiment for (a) conventional vector wind; (b) conventional temperature; (c) conventional specific humidity; (d) GPS RO bending angle; and (e) AMV. The standard deviations are computed from 20 June to 24 August 2022, four cycles per day. The error bars indicate the 95% confidence intervals.

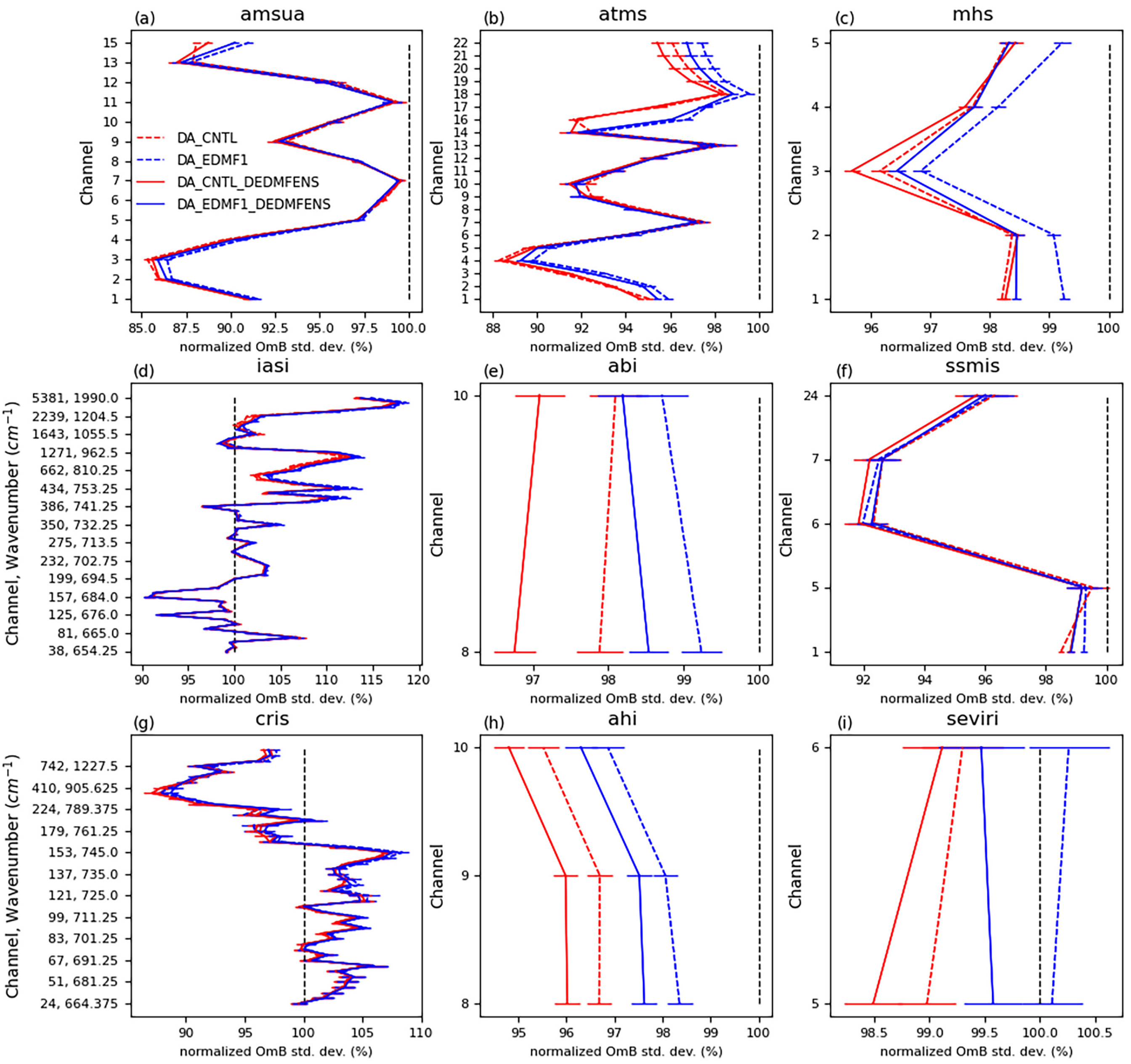

FIGURE 2 As Figure 1, but for satellite radiances after bias correction: (a) AMSU-A; (b) ATMS; (c) MHS; (d) IASI; (e) ABI; (f) SSMIS; (g) CrIS; (h) AHI; and (i) SEVIRI.

SHiELD forecast clearly shows better performance than GFS, but SHiELD forecast initialized from its own analysis is largely degraded. For example, with the same GFS_DA initial condition, the SHiELD_GFSIC_REPLAY forecast shows better fit to GPS RO bending angles than the GFS_DA forecast in almost all vertical levels, but the forecast initialized from the SHiELD DA_CNTL analysis shows much increased OmB standard deviations of GPS RO bending angles (Figure 1d). Similar degradations can also be found in the OmBs of wind and temperature observations in the upper troposphere and lower stratosphere (Figure 1a,b). This indicates that something else in the DA system affected the analysis. The analysis seems not well constrained by those observations. More understanding and mitigation of the degradations will be discussed in the next section.

For the verification against satellite radiance observations, we directly compare SHiELD DA_CNTL with the GFS_DA experiment. Figure 2 shows that SHiELD short-range forecast fits better to all MV radiances (Figure 2a–c, f). The percentages of improvement are quite significant. AMSU-A channels 1–5 and 15, ATMS channels 1–6 and 16–22 are assimilated in all-sky conditions (Zhu et al., 2016; Zhu et al., 2019; Tong et al., 2020). For all the assimilated IR humidity channels of geostationary satellite instruments, the OmB standard deviations of SHiELD

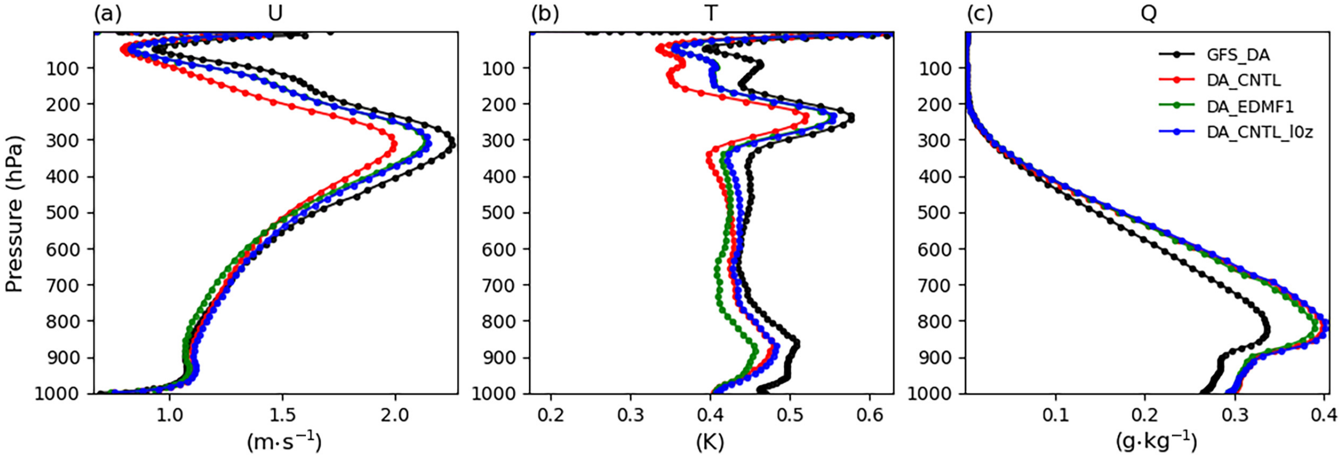

FIGURE 3 Global mean vertical profile of ensemble spread of (a) zonal wind ($\text{m}\cdot\text{s}^{-1}$); (b) temperature (K); and (c) specific humidity ($\text{g}\cdot\text{kg}^{-1}$) for GFS_DA, DA_CNTL, DA_EDMF1 and DA_CNTL_l0z experiments. The spreads are computed from the four DA cycles of 11 June 2022. DA_CNTL_l0z: Same as DA_CNTL, except that the background characteristic mixing length (l0) in the turbulent kinetic energy eddy-diffusivity mass-flux (TKE-EDMF) scheme is not a constant, but exponentially decreases with height as Equation (4) in Table 4.

are also smaller than GFS (Figure 2e, h, i). Smaller OmB standard deviations of SHiELD are also found in all MV humidity channels (Figure 2b, c). All these suggest a better temperature, humidity and cloud forecast of SHiELD in the regions observed by those IR and MV instruments. The OmB standard deviations of IASI and CrIS radiances are very different from those of other radiances. The assimilated observation counts of IASI/CrIS troposphere channels can be up to 50% more/less than those assimilated in GFS_DA. This is largely associated with the cloud detection scheme, which depends on the assigned observation errors (Bathmann & Collard, 2020). In the SHiELD experiments shown in Figure 2, the operational GFSv15 observation errors for IASI and CrIS are used, while new interchannel correlated observation errors estimated for GFSv16 are used in GFS_DA. This motivated us to explore the use of interchannel correlated observation errors for IASI and CrIS in the SHiELD DA system, which will be presented in Section 4.

3.3 | Improve SHiELD analysis through identifying and mitigating model errors

In this section, we mainly focus on improving SHiELD analysis by finding the source of the errors that lead to the degradations in the analysis manifested as the large OmBs of GPS RO bending angles, wind and temperature in the upper troposphere and stratosphere (Figure 1a,b,d). In an ensemble-based DA system, when the analysis is not well constrained by the observations, a possible reason is that the ensemble spread is not sufficient. Figure 3 shows the ensemble spread of zonal wind, temperature and specific humidity of SHiELD and GFS DA experiments. Although the same ensemble inflation methods are used in both models, their ensemble spread shows different characteristics, which reflects the differences in the model configurations. SHiELD DA_CNTL has a smaller zonal wind and temperature ensemble spread and a larger specific humidity ensemble spread than GFS_DA. The relatively larger gaps of zonal wind and temperature spread in the upper troposphere and stratosphere between DA_CNTL and GFS_DA made us suspect the lack of ensemble spread in DA_CNTL in those regions. Whether the ensemble spread is sufficient or not in a DA system is not simply concluded from comparing the spread between different models. To better quantify the error-spread consistency in the assimilation system, we computed ensemble reliability budget with the extended error-spread equation using either observation (Rodwell et al., 2015) or an analysis (Rodwell & Wernli, 2023) as the estimation of the truth. The extended error-spread equation can be simplified as:

$$\begin{aligned}&\text{verified against observation: Depar}^2\\&\quad=\text{Bias}^2+\text{EnsVar}+\text{ObsUnc}^2+\text{Residual},\end{aligned}\tag{1a}$$

$$\begin{aligned}&\text{verified against analysis: Error}^2\\&\quad=\text{Bias}^2+\text{EnsVar}+\text{AnUnc}^2+\text{Residual}.\end{aligned}\tag{1b}$$

The left-hand side of the equation stands for mean-squared ensemble mean forecast errors verified against observation (Depar^2) or analysis (Error^2). The terms on the right-hand side of the equation are squared bias of the ensemble mean forecast relative to the observation or analysis (Bias^2), mean ensemble variance (EnsVar),

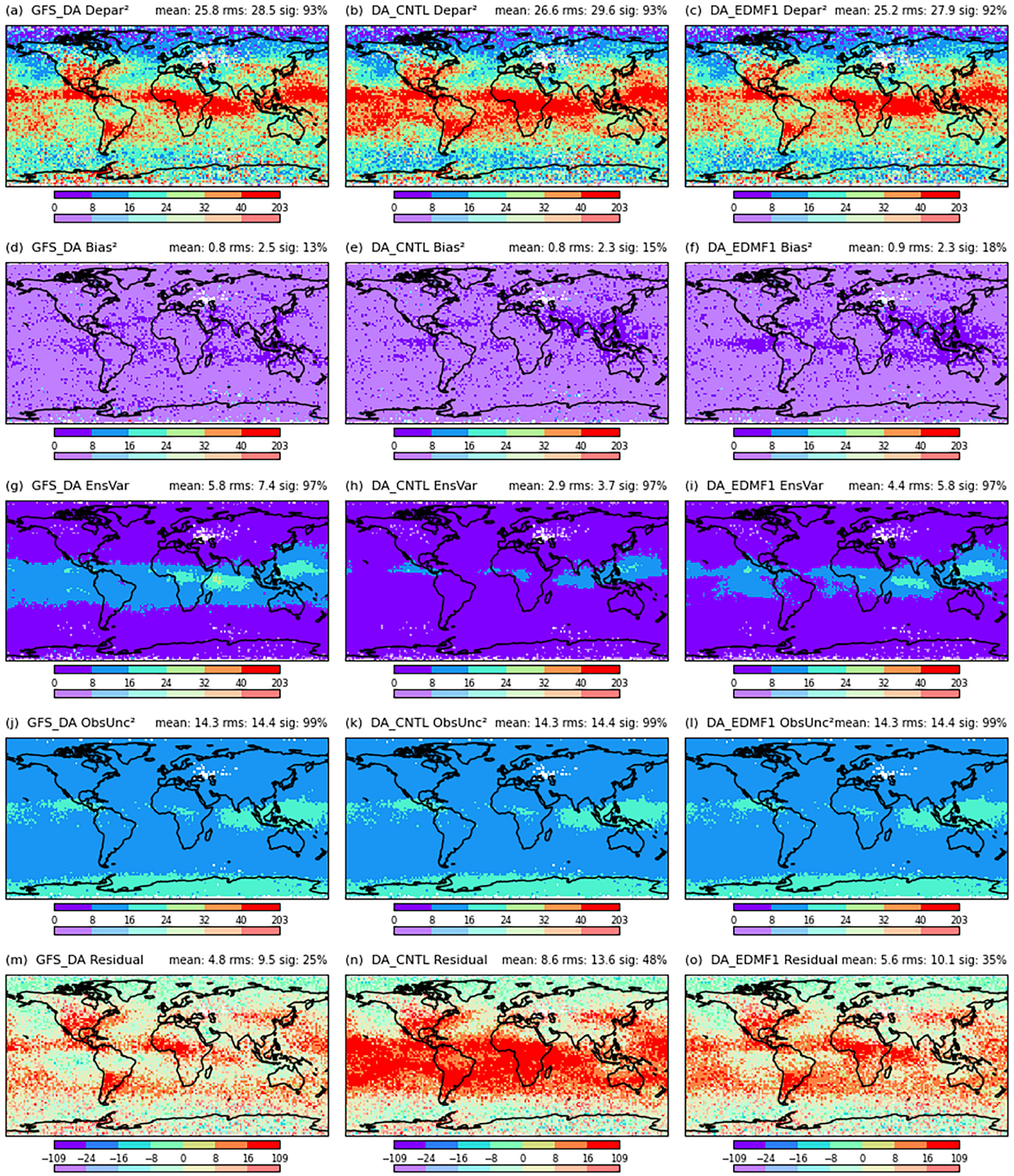

FIGURE 4 Ensemble reliability budget for GPS RO bending angle (10^{-10} radians^2) between 100 and 150 hPa computed from three to nine hours (assimilation window) ensemble forecast of the 0000 UTC and 1200 UTC cycles from 20 June to 24 August 2022 for GFS_DA (left column); DA_CNTL (middle column); and DA_EDMF1 (right column) experiments. (a–c) Mean-squared departure of the ensemble mean relative to the observation (Depar^2). (d–f) Squared bias of the ensemble mean forecast relative to the observation (Bias^2). (g–i) Mean ensemble variance (EnsVar). (j–l) Mean squared observation error (ObsUnc^2). (m–o) The residual term. The data were aggregated to 2° × 2° grid boxes before they are used to compute the budget. The grid boxes with more saturated color indicate that budget terms are different from 0 at 5% significance level. The mean and root-mean-square of the budget terms over the global domain, as well as the percentage of the area exceeding the 5% significance level are shown at the top of each panel.

TABLE 4 Selected differences between EDMF0 and EDMF1 versions of the TKE-EDMF scheme, and the selected difference between the scheme used in SHiELD and the scheme used in GFS.

Parameterization	SHiELD initial version EDMF0	SHiELD newer version EDMF1	GFS newer version EDMF1
Background characteristic mixing length l_0 (m)	$l_0 = 30.0$ (3)	$l_0 = 30.0e^{\left[-2.5\left(1-P/P_s\right)^2\right]}$ (4)	
Minimum TKE e ($\mathrm{m}^2 \cdot \mathrm{s}^{-2}$) threshold used in TKE integration	$e_{\min} = 1.0^{-9}$ (5)	$e_{\min} = \left(\frac{K_0}{c_m l_k}\right)^2$ (6)	
Eddy diffusivity K_ϕ ($\mathrm{m}^2 \cdot \mathrm{s}^{-1}$) at stratocumulus cloud top	$K_\phi = c_\phi l_k \sqrt{e} + K_{\Delta R}$ (7)	$K_\phi = c_\phi l_k \sqrt{e}$ (8)	
Non-local mixing for hydrometeors	Off		On

Note: Equations are based on Han and Bretherton (2019) and Han et al. (2021).

Symbols: K_ϕ, eddy diffusivity ($\mathrm{m}^2 \cdot \mathrm{s}^{-1}$); c_ϕ, a proportionality coefficient; l_k, turbulent mixing length (m); e, TKE ($\mathrm{m}^2 \cdot \mathrm{s}^{-2}$); l_0, background characteristic mixing length (m); P, level pressure (Pa); P_s, surface pressure (Pa); θ_v, virtual potential temperature (K); $K_{\Delta R}$, eddy diffusivity proportional to radiative flux jump at cloud top (ΔR); K_0, background diffusivity ($\mathrm{m}^2 \cdot \mathrm{s}^{-1}$); c_m, coefficient for momentum.

Abbreviations: EDMF, eddy-diffusivity mass-flux; GFS, Global Forecast System; SHiELD, System for High-resolution prediction on Earth-to-Local Domains; TKE, turbulent kinetic energy.

mean squared uncertainty of observation (ObsUnc^2) or analysis (AnUnc^2) and a residual. The exact equations and their derivations can be found in Rodwell et al. (2015) and Rodwell and Wernli (2023). Statistically significant non-zero residual would indicate deficiency in the ensemble spread and in some cases deficiency in the estimation of observation errors (see discussion in Rodwell et al., 2015).

Figure 4 shows the reliability budget for GPS RO bending angles between 100 and 150 hPa. SHiELD DA_CNTL has larger ensemble mean departures than GFS_DA in the tropical region, where about 50% of the GPS RO observations are assimilated. This explains what we found in Figure 1d. The Bias^2 are small and generally not significant. The ObsUnc^2 term is a main contributor to the budget. Since we adopted the GPS RO observation errors from GFS, the ObsUnc^2 are basically identical in all the experiments. The SHiELD DA_CNTL has large and significant positive residuals in the tropics (Figure 4n), which is the result of the ensemble spread (EnsVar) deficit as shown in Figure 4h.

The insufficient ensemble spread of the SHiELD DA_CNTL in the upper troposphere explains the larger background departures shown in Figure 1d. The question is what caused the deficiency in the ensemble spread and how it can be resolved. In order to boost the ensemble spread, we tested increasing the inflation factor of the RTPS scheme and moderately increasing the amplitude of the random perturbations applied to the net physics tendencies in the SPPT scheme from 0.5 to 0.6. This can increase the spread but is still far from closing the gap and producing desired results. We then turned our attention to the model itself. After several sensitivity tests, the problem was identified to be associated with the TKE-EDMF scheme. The initial version (EDMF0) of the TKE-EDMF scheme, as documented in Han and Bretherton (2019), is used in SHiELD 2022, while an updated TKE-EDMF scheme (EDMF1) is used in GFSv16. The differences between the two versions are summarized in Table 4 and will be discussed in Section 5. When we switched to the updated TKE-EDMF scheme and ran the DA cycling experiment (DA_EDMF1), the ensemble spreads above 500 hPa were significantly increased (Figures 3a,b and 4i). The positive GPS RO residual is significantly reduced (Figure 4o). Accompanied with that is the improved background departures of GPS RO, wind and temperature observations in the upper troposphere and stratosphere (DA_EDMF1 vs DA_CNTL in Figure 1a,b,d).

Switching to the upgraded TKE-EDMF scheme improves the ensemble spread in the upper troposphere and stratosphere, but decreases the ensemble spread in the middle to the lower troposphere (DA_EDMF1 in Figure 3). To see if the reduced ensemble spread is a potential source of analysis and forecast errors, we examined the ensemble reliability budget in the lower troposphere. Figure 5 shows the reliability verified against ERA5 reanalysis (Hersbach et al., 2020) for temperature at 850 hPa. We are not showing the reliability budget in observation space, because observations have limited coverage. For example, the GPS RO bending angles at and below the top of the planetary boundary layer (PBL) in stratocumulus regions are largely rejected due to super-refraction. Also, some observation errors are largely inflated to account for representation errors and observation error correlations, such as satellite radiance observation errors. This leads to large negative residuals that obscure the judgment of ensemble reliability. To be able to see the signs of the biases, the square roots of the budget terms in Equation (1b) are computed following the description in Rodwell and Wernli (2023).

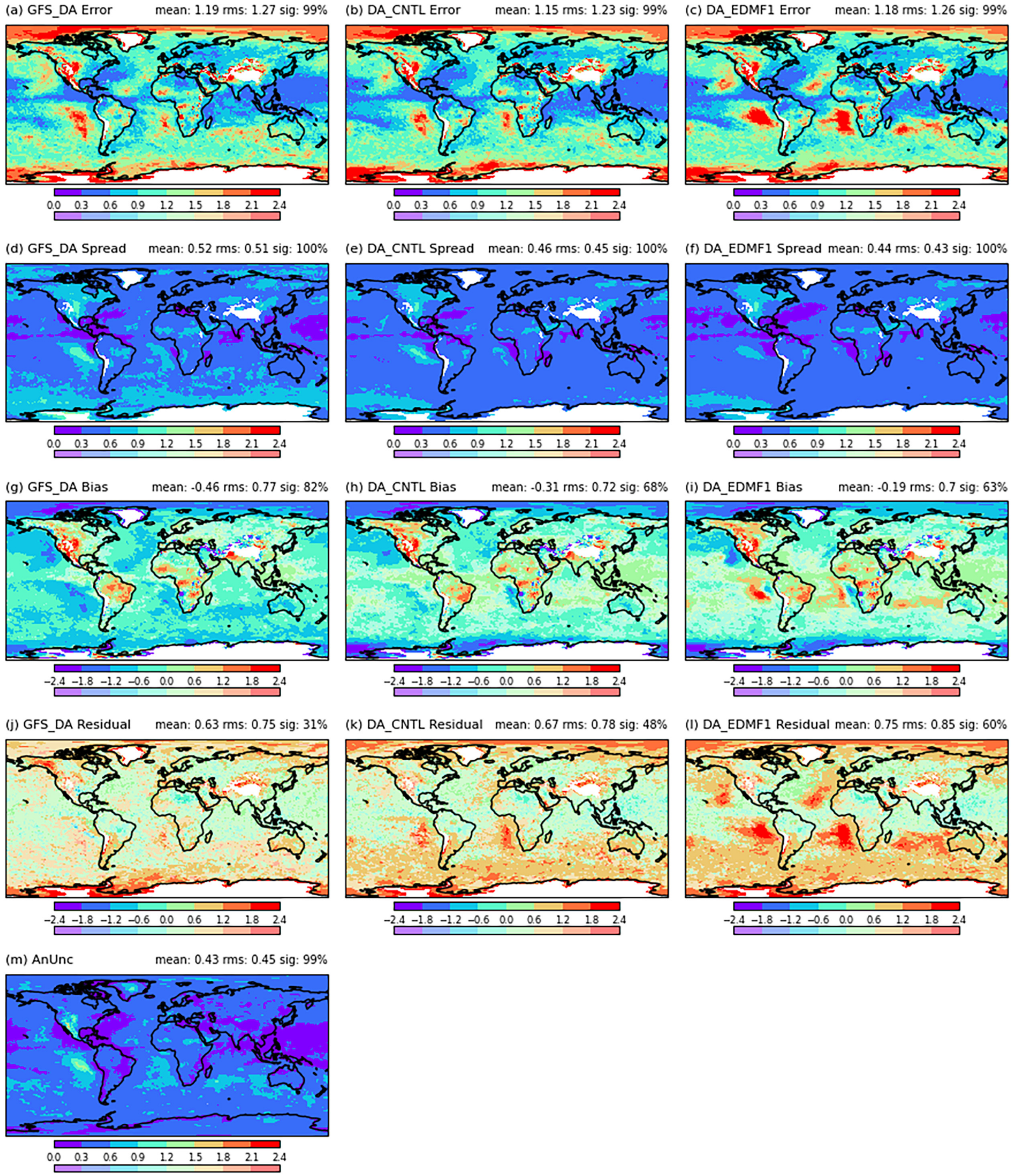

FIGURE 5 Square root of the extended error-spread equation verified against ERA5 reanalysis for temperature (K) at 850 hPa. The budget terms were computed from the six-hour ensemble forecast valid at 0000 UTC from 20 June to 23 August 2022 for GFS_DA (left column); SHiELD DA_CNTL (middle column); and SHiELD DA_EDMF1 (right column). (a–c) Mean of ensemble mean forecast error. (d–f) Mean ensemble spread. (g–i) Ensemble mean forecast bias. (j–l) Square root of the residual term. (m) Uncertainty of ERA5 reanalysis. The residual term is computed as the square root of its absolute value preserving its sign in squared form. Statistical significance is determined from the original squared form. Similar to Figure 4, the saturated colors indicate that the budget terms are different from 0 at 5% significance level. The uncertainty estimation of ERA5 comes from the 10-member ensemble of data assimilation (Isaksen et al., 2010) and it only accounts for random errors not for systematic errors.

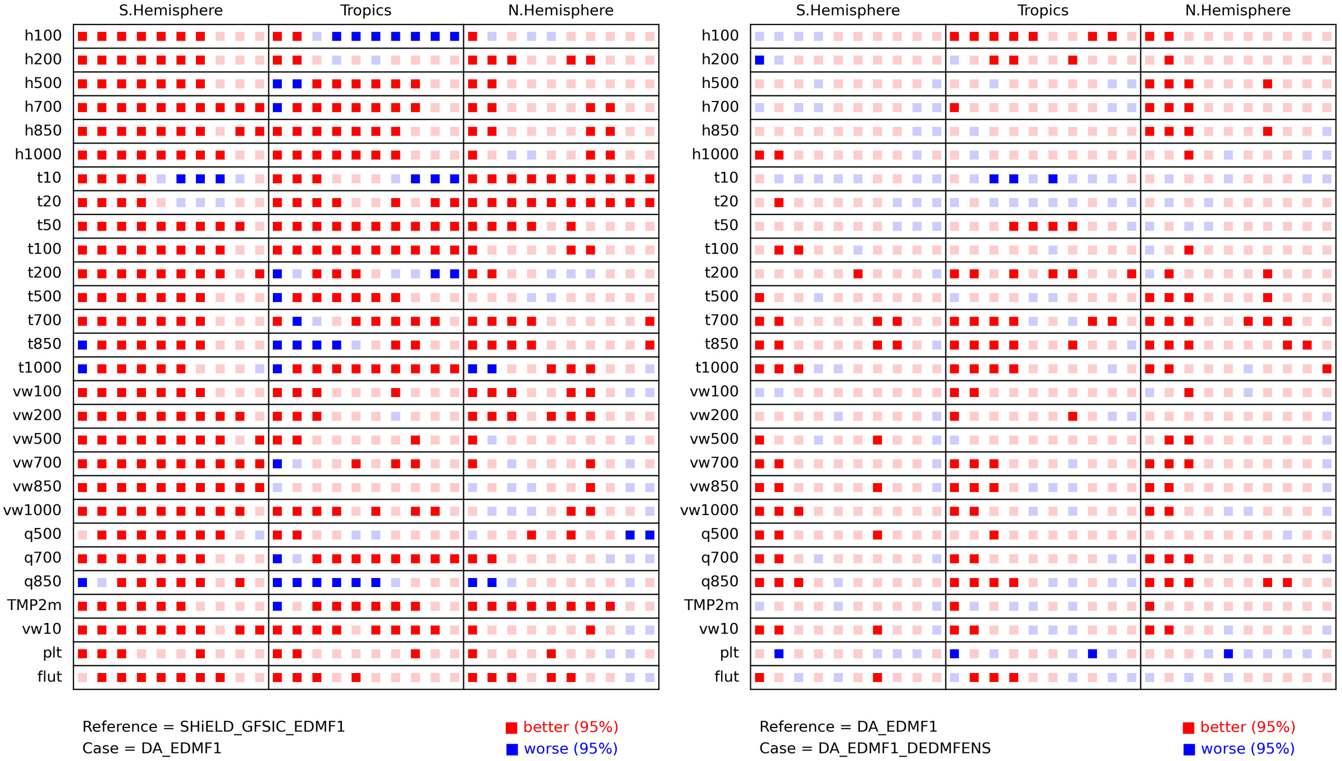

FIGURE 6 System for High-resolution prediction on Earth-to-Local Domains (SHiELD) 10-day forecast scorecards comparing DA_EDMF1 to SHiELD_GFSIC_EDMF1 (left); and comparing DA_EDMF1_DEDMFENS to DA_EDMF1 (right). The scores are the root-mean-squared errors (RMSE) verified against ERA5 reanalysis. The labels on the left side of the scorecards are the verified variables at different pressure levels indicated by the numbers. The variables verified are *h* (geopotential height), *t* (temperature), vw (vector wind), *q* (specific humidity), TMP2m (2-m temperature), vw10 (10-m vector wind), plt (total precipitation), flut (outgoing longwave radiation). Red/blue color means the evaluated experiment (Case) is better/worse than the reference experiment (Reference). Darker color means the difference exceeds 95% confidence interval.

Switching to the EDMF1 version of the TKE-EDMF scheme (DA_EDMF1) increases the ensemble mean forecast error of temperature at 850 hPa (Figure 5c vs b). The degradation is in the marine stratocumulus regions. Figure 5i shows that the positive temperature biases in those regions are a significant contributor to the elevated forecast error. The caveat of the verification in model space is that the forecast error and bias are sensitive to the verifying analysis. When the results were verified against IFS analysis (not shown), we found even larger positive biases and errors in DA_EDMF1 in the marine stratocumulus regions, and the biases in DA_CNTL became positive and only slightly negative in GFS_DA. Nevertheless, this does not change our conclusion for the ensemble reliability. As we can see from Figure 5f,l, the positive residual in DA_EDMF1 largely comes from the lack of ensemble spread. The ensemble spread is more reliable in DA_CNTL at this level, as the positive residual is much smaller and less significant (Figure 5k). In Section 5, we will discuss how the differences in the two versions of the TKE-EDMF scheme affect the ensemble spread.

The ensemble reliability budget clearly shows that both versions of the TKE-EDMF scheme introduce deficiencies in ensemble spread, one in the upper troposphere and one in the lower troposphere. Before we invest in more complicated ways to address the ensemble spread deficiency, we first make use of the complementary errors of the two versions of the TKE-EDMF scheme to enhance the ensemble. Inspired by the multiparameterization scheme of representing model uncertainty, we use both versions of the TKE-EDMF scheme in the ensemble, that is each version of the scheme is used by half of the ensemble members. We tested this idea in two experiments: DA_CNTL_DEDMFENS and DA_EDMF1_DEDMFENS. The difference between the two experiments is the version of the TKE-EDMF scheme used in the deterministic forecast. By doing so, we found significant improvement in short-range forecasts verified against observations (DA_CNTL_DEDMFENS/DA_EDMF1_DEDMFENS vs DA_CNTL/DA_EDMF1 in Figures 1 and 2). For example, the background departures of GPS RO bending angle above 800 hPa are greatly improved when

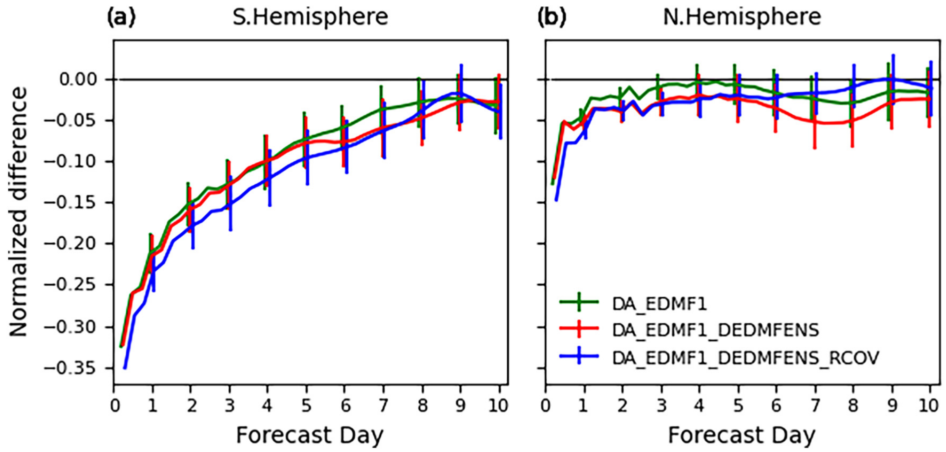

FIGURE 7 SHiELD_GFSIC_EDMF1 normalized root-mean-squared error (RMSE) of 500 hPa geopotential height verified against ERA5 reanalysis in (a) the southern hemisphere and (b) the northern hemisphere for the DA_EDMF1, DA_EDMF1_DEDMFENS and DA_EDMF1_DEDMFENS_RCOV experiments. The error bars indicate the 95% confidence intervals.

the deterministic forecast uses the EDMF0 version of the TKE-EDMF scheme (DA_CNTL_DEDMFENS vs DA_CNTL in Figure 1d). The departures of DA_CNTL_DEDMFENS are even smaller than DA_EDMF1 above 600 hPa, where the DA system suffers much less from the ensemble deficiency there. Similarly, the modified ensemble substantially reduced the background departures of GPS RO bending angles below 400 hPa, where the use of the EDMF1 version TKE-EDMF scheme leads to underdispersed ensemble in the lower troposphere (DA_EDMF1_DEDMFENS vs DA_EDMF1 in Figure 1d).

3.4 | Medium-range forecast impact

The medium-range forecast impact of SHiELD DA is evaluated by comparing SHiELD 10-day forecast initialized from its own analysis to SHiELD forecast initialized from GFS_DA analysis. The 10-day forecasts were run once every day initialized at 0000 UTC from 20 June to 24 August 2022. Here we demonstrate the set of SHiELD DA experiments using the EDMF1 version of the TKE-EDMF scheme in the deterministic forecast. Similar conclusions can be obtained for the set of experiments using the EDMF0 version of the TKE-EDMF scheme.

The scorecard on the left side of Figure 6 shows that SHiELD forecast skills are significantly improved by using its own analysis. There are more improvements in the southern hemisphere than in the tropics and the northern hemisphere. The positive impact can last up to 10 days. Figure 7a shows profound improvement in the forecast of 500 hPa geopotential height in the southern hemisphere. The reduction of the forecast error is more than 15% for the short-range forecast as a result of improved analysis and more than 5% out to day 6. Similar improvement is also found in many other variables evaluated here in the southern hemisphere (not shown). The large degradation in 850 hPa temperature and humidity forecast particularly in the tropics (Figure 8b,e) has been shown in Figure 5c to be largely within the marine stratocumulus regions. The scorecard on the right side of Figure 6 shows the impact of including two versions of the TKE-EDMF scheme in the ensemble. The enhanced ensemble can further improve forecast skills and the positive impact can last more than five days. For example, the positive impact on the 500 hPa geopotential height forecast can extend one more day in the southern hemisphere and extend to day 8 in the northern hemisphere (Figure 7). It is noteworthy that the enhanced ensemble works on the degraded (worse) areas in the DA_EDMF1 experiment as shown on the left scorecard. Particularly, the large short-range forecast errors of 850 hPa temperature and humidity in the tropics are vastly reduced (Figure 8).

4 | IMPACT OF INTERCHANNEL CORRELATED OBSERVATION ERRORS FOR IASI AND CRIS

In this section, we continue enhancing the SHiELD DA system through upgrading the observation errors for IASI and CrIS by accounting for interchannel error correlations. This work focuses on interchannel correlated errors in the context of clear-sky radiance assimilation of IASI and CrIS. Directly using the observation error covariances from GFS did not produce desired results. This indicates that the optimal observation errors can be different for different modeling systems due to representation errors, which motivated us to estimate the observation error covariances for our own system. A notable finding from this effort is that the significant errors in the marine stratocumulus regions, caused by modifications to the vertical mixing scheme, are further diminished with the optimized observations errors (Figures 5, 8 and 15).

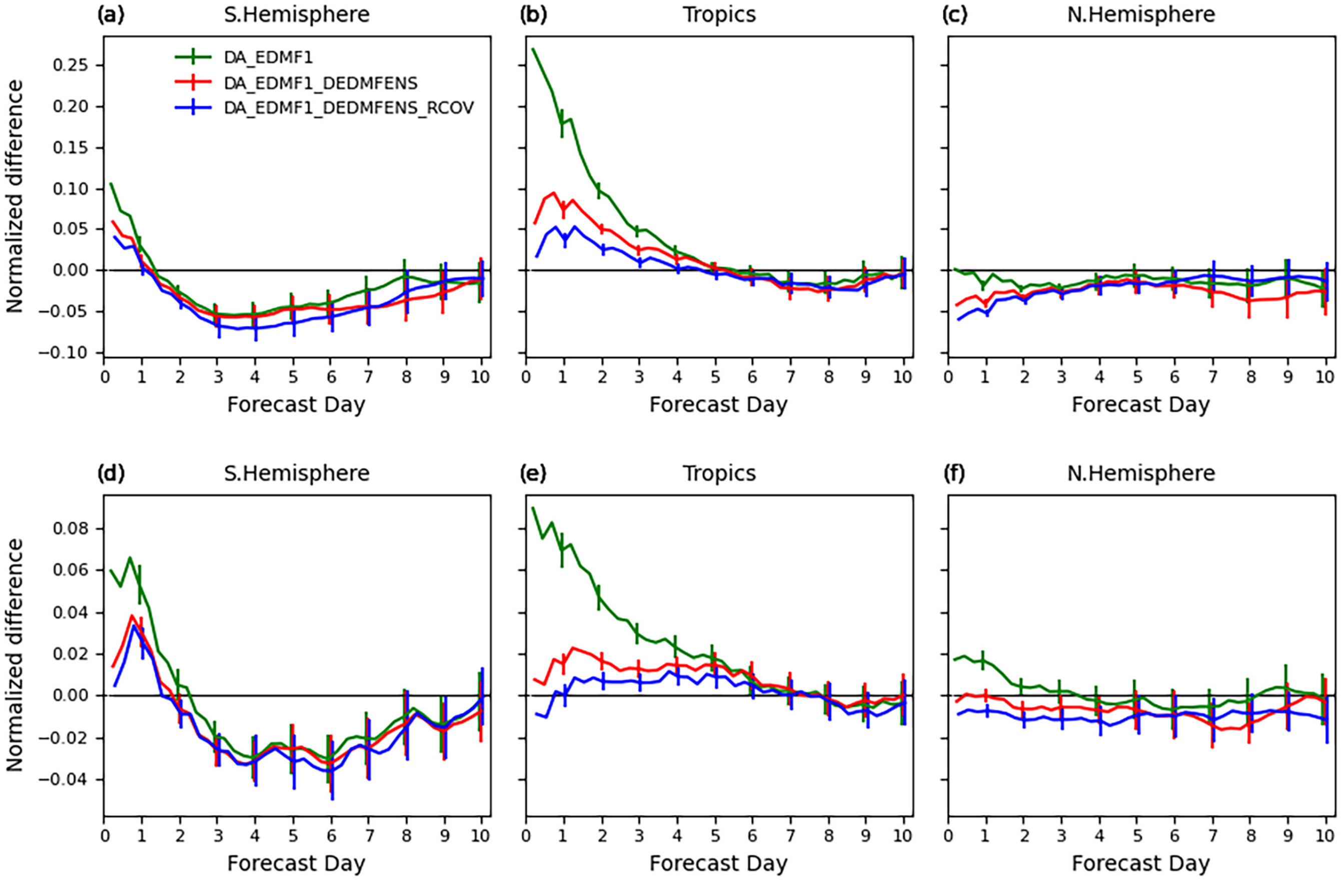

FIGURE 8 As Figure 7, but for the normalized root-mean-squared error (RMSE) of 850 hPa temperature (upper panels) and specific humidity (lower panels) in the southern hemisphere (a,d), the tropics (b,e) and the northern hemisphere (c,f).

4.1 | Method

The use of interchannel correlated observation errors for IASI and CrIS has been investigated and implemented in operational centers (Bathmann & Collard, 2020; Bormann et al., 2016; Eresmaa et al., 2017; Weston et al., 2014). There are some differences in the way to derive and recondition the observation error covariance matrix in those earlier studies. We generally follow the methods used in GFS to compute the error covariance matrices as documented in Bathmann and Collard (2020), that is using the Desroziers et al. (2005) diagnostic to compute the covariance matrices; deriving the covariances over sea and land separately; and using the first method described in Weston et al. (2014) to recondition the covariance matrices. We also tried to use the Hollingsworth and Lönnberg (1986) method for the initial estimation of the covariances (Bormann et al., 2011; Bormann et al., 2016; Bormann & Bauer, 2010). However, with the same error inflation, using the covariances derived from the Desroziers et al. (2005) method leads to smaller OmB standard deviations of other independent observations in our system. So we lean toward using the covariance matrices derived from the Desroziers et al. (2005) method. With the Desroziers et al. (2005) approach, the observation error covariance matrix (**R**) is computed as $R = E\left[d_{\mathrm{a}}^{o}\left(d_{\mathrm{b}}^{o}\right)^{T}\right]$, where E is the statistical expectation operator, d_{a}^{o} is observation minus analysis vector (OmA) and d_{b}^{o} is the observation minus background vector (OmB). The covariance matrix is reconditioned by setting a minimum eigenvalue threshold $\left(\lambda_{\text{thresh}} = \frac{\lambda_{\max}}{K}\right)$, where $\lambda_{\max}$ is the existing maximum eigenvalue and K is a required condition number. Eigenvalues smaller than the threshold are set to the threshold and the matrix is reconstructed using the new eigenvalues and the original eigenvectors (Weston et al., 2014).

The main difference between our estimation and the GFS estimation for the error covariances lies in the way to inflate the diagnosed error covariances. Error inflation is commonly used to compensate for the neglected error correlations. Previous studies show that error inflation is also beneficial to possibly counteract inaccuracies of the diagnosed matrices and suboptimality in the assimilation of the instrument data, when error correlations are taken into account (Bormann et al., 2016). In Bathmann and Collard (2020), only the diagonal values of **R** are multiplied by an inflation factor. While this can further reduce the condition number of **R**, the error correlations are largely diluted as shown in their figures 1 and 3. When we first attempted to inflate the diagonal values of **R**, we barely found an inflation factor that could broadly improve the background fit to other observations. Applying different inflation factors to different spectral regions may help achieve better results as how it is done for GFS, but this requires more careful tuning. From a performance perspective, we chose to apply the inflation factor to the whole covariance matrix.

One-month (July 2022) departure data from the DA_EDMF1_DEDMFENS experiment (Table 3) were used to compute the observation error covariances. Following GFS, the IASI matrices were reconditioned to $K = 200$ and the CrIS matrices were reconditioned to $K = 125$. Several ensemble replay experiments, where the background error is fixed by reusing the ensemble from the DA_EDMF1_DEDMFENS experiment and the DA cycling is only run for the deterministic component of 4DEnVar, were conducted for IASI and CrIS separately to determine the optimal inflation factors based on the OmBs of other independent observations. The short-range forecast compared to IFS analysis was also used as a reference. In GSI, the observation error is used in the cloud detection scheme to remove IR radiances affected by clouds (equation 1 in Bathmann & Collard, 2020). We also ran experiments to evaluate the impact of using the old and new diagnosed observation errors in the cloud detection scheme. The optimal inflation factors determined from the experiments are 1.5 for IASI and 2.0 for CrIS. We found that using the old observation errors in the cloud detection scheme can produce better short-range forecasts, especially in the tropics. Finally, we conducted a full 4DEnVar experiment (DA_EDMF1_DEDMFENS_RCOV) using the new observation error covariances for IASI and CrIS altogether with the optimal inflation factors applied and the old observation errors used in the cloud detection scheme.

4.2 | Forecast impact

The impact of using the new observation error covariances for IASI and CrIS are evaluated by comparing the DA_EDMF1_DEDMFENS_RCOV experiment to the DA_EDMF1_DEDMFENS experiment. The short-range forecast impact evaluated by the OmB standard deviations of all other assimilated observations are presented in Figures 9 and 10. The short-range forecast is generally improved as indicated by the smaller OmB standard deviations of the majority of the observations. The background fit to GPS RO observations is improved in all vertical levels (Figure 9d). The improvement in the wind field is

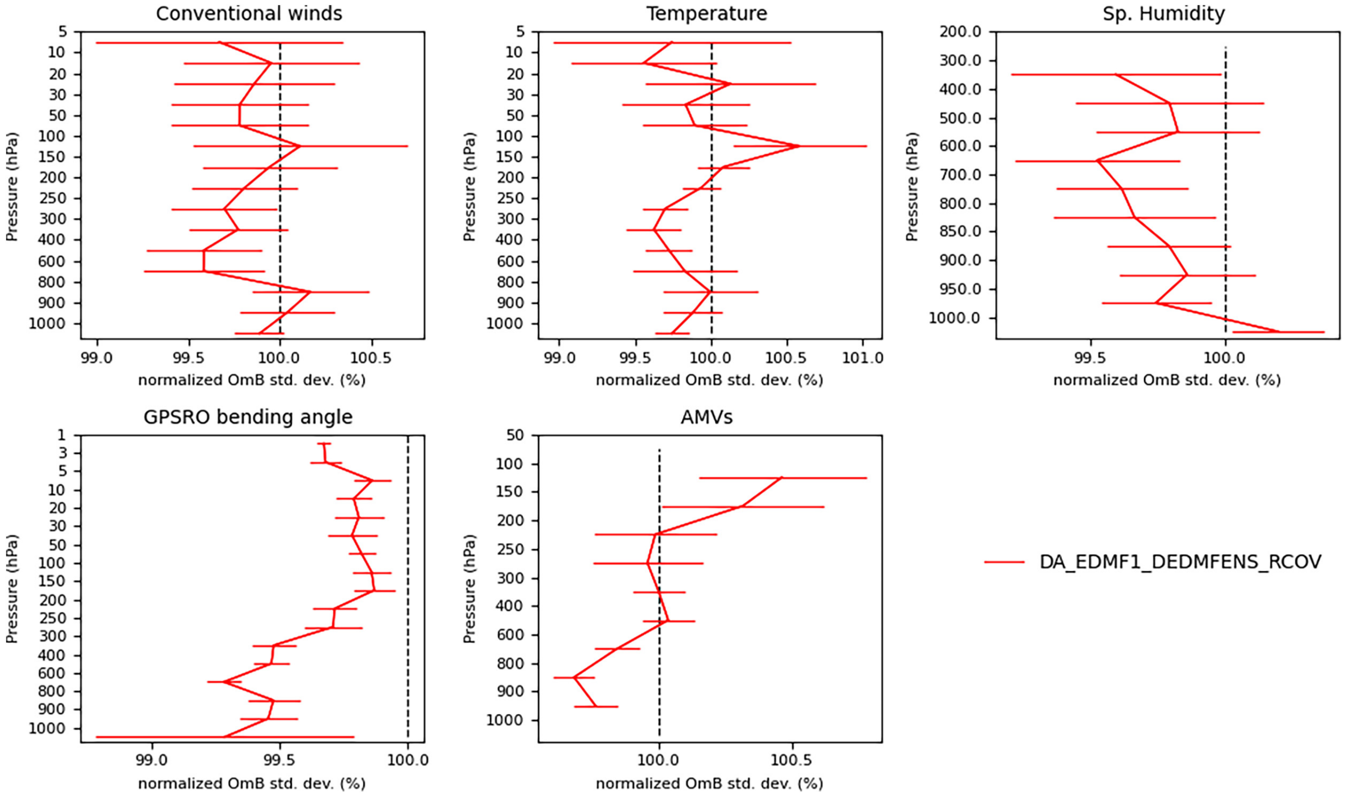

FIGURE 9 As Figure 1, but for DA_EDMF1_DEDMFENS_RCOV normalized by the observation minus background (OmB) standard deviations of DA_EDMF1_DEDMFENS.

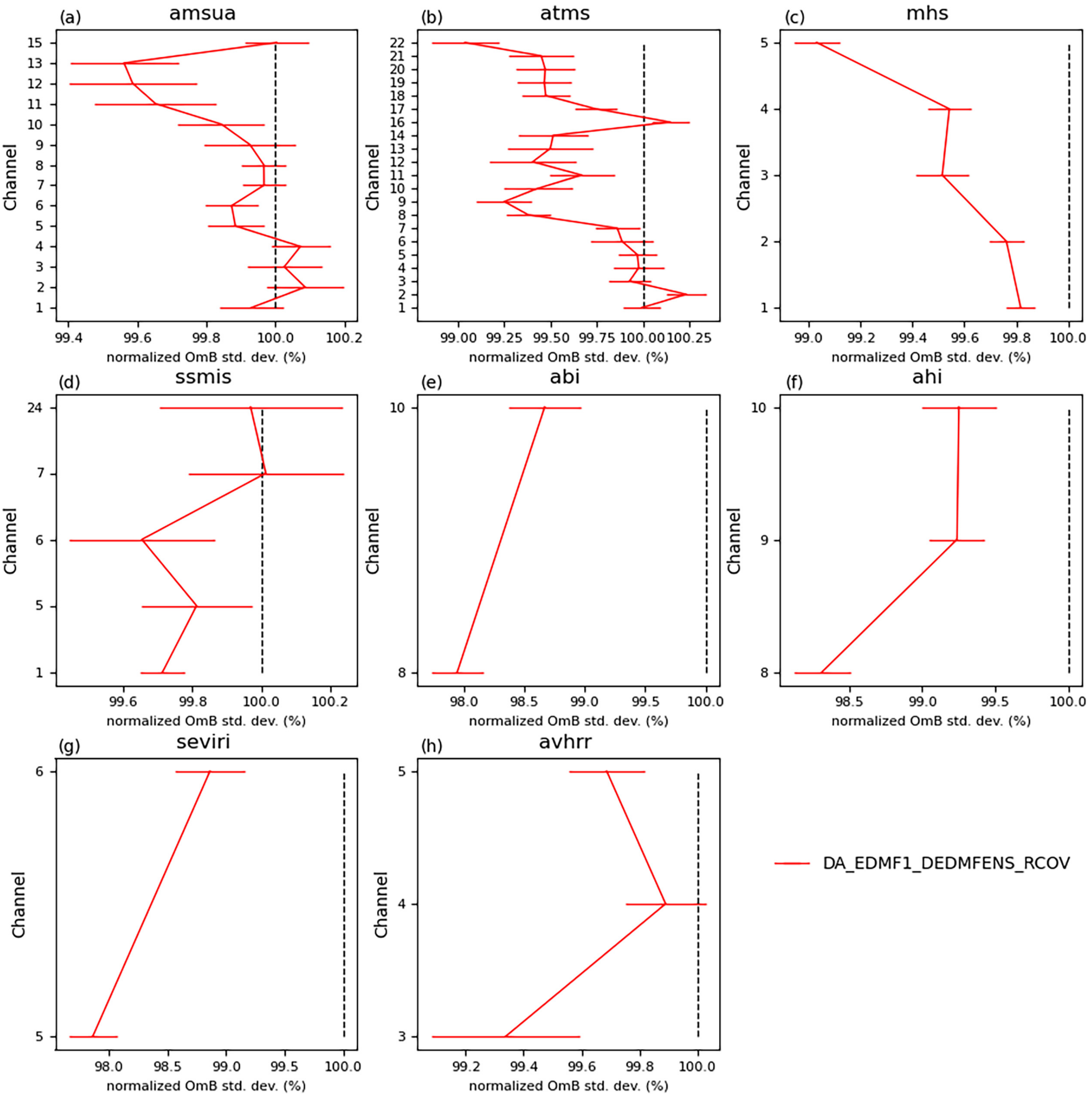

FIGURE 10 As Figure 9, but for satellite radiances: (a) AMSU-A; (b) ATMS; (c) MHS; (d) SSMIS; (e) ABI; (f) AHI; (g) SEVIRI; and (h) AVHRR.

mainly in the lower and middle troposphere (Figure 9a,e), while there is some degradation in the fit to AMVs in the upper troposphere. There is clear improvement in humidity, as the standard deviations of background departures of all MV and IR water vapor channels (Figure 10b,c,e–g) are smaller. There is also significant improvement in the background fit to the AMSU-A and ATMS stratosphere temperature-sounding channels. The improvements for the AMSU-A and ATMS are even better in the tropics, especially for the temperature-sounding channels assimilated in all-sky conditions (not shown).

The forecast of DA_EDMF1_DEDMFENS verified against ERA5 reanalysis shows significant improvement upon the forecast using GFS analysis for almost all variables in all three geographical regions (Figure 11 left panel). The addition of the new interchannel correlated observation errors for IASI and CrIS shows robust improvement up to day 5. More improvement is found in the southern hemisphere, and there is noticeable reduction of the 500 hPa geopotential height forecast error (Figure 7a) that is significant up to day 4 (Figure 11). Most interestingly, the new observation error covariances

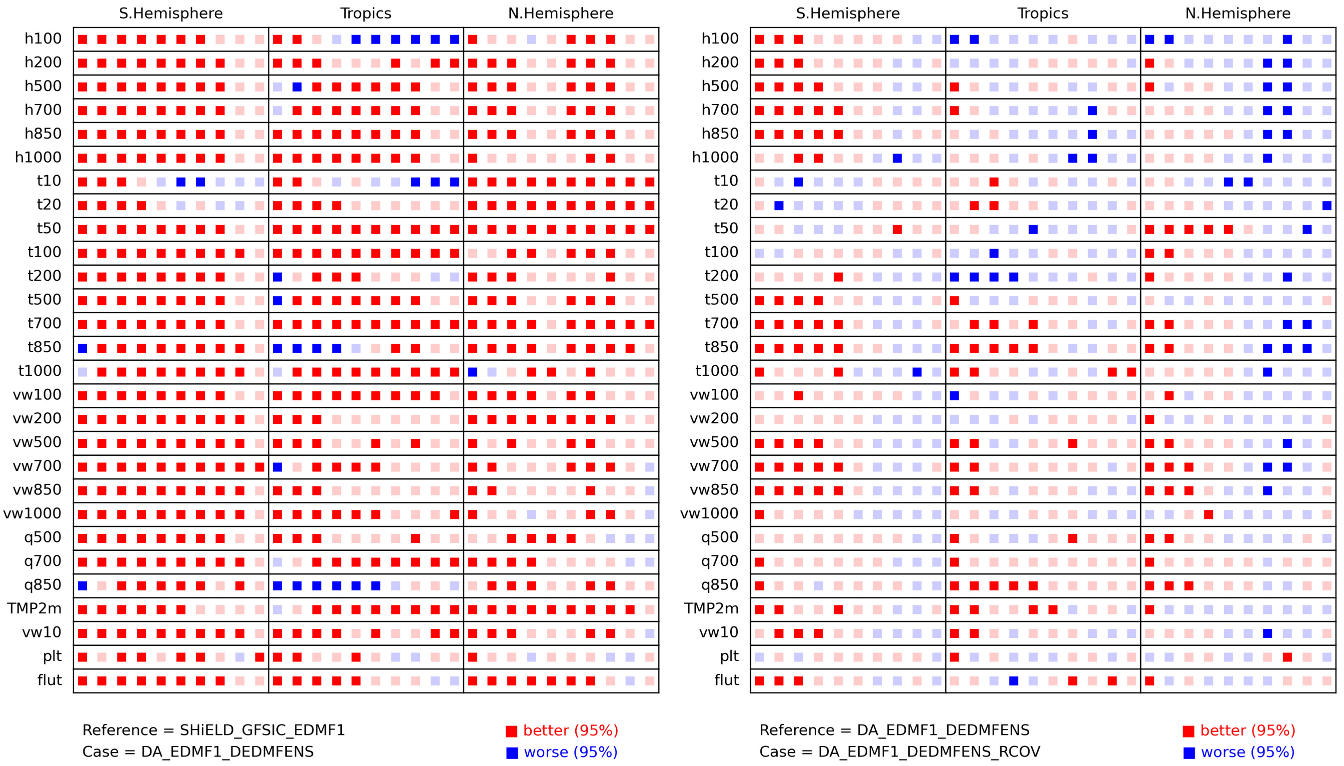

FIGURE 11 As Figure 6, but comparing DA_EDMF1_DEDMFENS_RCOV to DA_EDMF1_DEDMFENS to show the impact of using interchannel correlated observation error covariances for IASI and CrIS.

can further improve 850 hPa temperature and humidity forecast in the tropics, where the system struggles with the EDMF1 version of the TKE-EDMF scheme (also see Figure 8b,e).

5 | UNDERSTAND MODEL ERROR ASSOCIATED WITH THE TKE-EDMF SCHEME

Based on the experiments discussed in Section 3.3, the main source of analysis and forecast errors was identified to be within the TKE-EDMF scheme. In this section, we aim to gain a deeper understanding of the root causes of these errors. In order to do that, we first compared the differences between the two versions of the scheme and differences between the scheme implemented in SHiELD and that implemented in GFS. Sensitivity experiments (Tables 3 and 5) were conducted to understand the outcome of the differences. We would like to answer the following two questions through the investigation:

1. What changes in the TKE-EDMF scheme are responsible for the deficiencies in ensemble spread observed in both versions of the scheme but at different levels (Figures 3–5)?
2. What causes the large systematic temperature and humidity errors in the marine stratocumulus regions?

In the TKE-EDMF scheme (Han & Bretherton, 2019), the vertical turbulent flux of a field ϕ is written as:

$$\overline{w'\phi'} = -K_\phi \frac{\partial \overline{\phi}}{\partial z} + M_\mathrm{u}\left(\phi_\mathrm{u} - \overline{\phi}\right) - M_\mathrm{d}\left(\phi_\mathrm{d} - \overline{\phi}\right), \qquad (2)$$

where the terms on the right side of the equation are eddy diffusion, surface-driven updraft mass-flux mixing and stratocumulus-top-driven downdraft mass-flux mixing. The overbars denote horizontal averages across a grid cell and primes represent turbulent fluctuations. There are many subtle differences between the two versions of the TKE-EDMF scheme in the parameterization of eddy diffusivity K_ϕ, mass-flux for the updrafts M_u and mass-flux for the downdrafts M_d. The differences that can be used to explain the results and help answer the two questions are listed in Table 4.

5.1 | Impact of the parameterization on ensemble spread

We found that the ensemble spread is influenced by the strength of eddy diffusion in the TKE-EDMF scheme. The eddy diffusivity K_ϕ is a function of turbulent mixing length l_k and TKE (Equations 7 and 8 in Table 4). In the TKE-EDMF scheme, both turbulent

TABLE 5 Names and descriptions of SHiELD forecast sensitivity experiments run at C196 horizontal resolution with its original 91 vertical level configuration. All forecast experiments are initialized from operational GFS analysis at 0000 UTC on 1 July 2022.

Experiment	Description
EDMF0	SHiELD forecast with the EDMF0 version of the TKE-EDMF scheme
EDMF0_l0z	As EDMF0, but the background characteristic mixing length l_0 is computed using Equation (4) in Table 4
EDMF1	SHiELD forecast with the EDMF1 version of the TKE-EDMF scheme
EDMF1_tkemin	As EDMF1, but the minimum TKE threshold is set to be 1.0^{-9} as Equation (5) in Table 4.
EDMF1_radj	As EDMF1, but the eddy diffusivity at stratocumulus cloud top is computed with the additional $K_{\Delta R}$ as Equation (7) in Table 4
EDMF1_nonlocal	As EDMF1, but turn on non-local mixing for hydrometeors.

Abbreviations: EDMF, eddy-diffusivity mass-flux; GFS, Global Forecast System; SHiELD, System for High-resolution prediction on Earth-to-Local Domains; TKE, turbulent kinetic energy.

mixing length and TKE are bounded by a minimum threshold. We will demonstrate how the parameterization of the background (or minimum) characteristic mixing length and the minimum TKE threshold affect the strength of eddy diffusion and, consequently, the ensemble spread.

The background characteristic mixing length l_0 is a constant in EDMF0 (Equation 3 in Table 4) while it exponentially decreases with height in EDMF1 (Equation 4 in Table 4). When we change l_0 to Equation 4 in version EDMF0 in SHiELD forecast experiment EDMF0_l0z (Table 5), the resulting mixing length l_k, eddy diffusivity K_ϕ and eddy diffusion component of the turbulent flux $-K_\phi \frac{\partial \overline{\phi}}{\partial z}$ (using turbulent heat flux as an example) become smaller than those in the EDMF0 forecast, and the reductions increase with height (Figure 12a,b,d). The lack of ensemble spread in the upper troposphere and stratosphere in DA_CNTL (Figures 3a,b and 4h) is the result of using a vertically constant l_0 in the EDMF0 version of the scheme, which produces too much local turbulent mixing. When using a vertically decreasing l_0 in DA cycling with version EDMF0 (DA_CNTL_l0z), the ensemble spread is largely increased and the gain in ensemble spread also increases with height (Figure 3a,b). Above 350 hPa, there is little difference between DA_CNTL_l0z and DA_EDMF1 in zonal wind and temperature ensemble spread.

The minimum TKE threshold is applied during the integration of the TKE equation based on equations (37) and (38) in Han and Bretherton (2019). The minimum TKE ($e_{\min}$) is deduced from background diffusivity for momentum (Equation 6 in Table 4) in version EDMF1, while $e_{\min}$ is set to be 1.0^{-9} in version EDMF0. Figure 12c shows that the predicted TKE in the EDMF1 forecast experiment is much greater than that predicted in the EDMF1_tkemin forecast experiment (Table 5), where $e_{\min}$ is changed to be 1.0^{-9} in EDMF1, particularly in the lower to middle troposphere. Therefore, EDMF1 produces more eddy diffusion, which is even larger than EDMF0 below 400 hPa (Figure 12d). When we change $e_{\min}$ to 1.0^{-9} with version EDMF1 in the DA cycling, the ensemble spread increases in the lower to middle troposphere due to the reduced eddy diffusion there (not shown, but almost overlaps with DA_CNTL_l0z in Figure 3). We need to point out that changing the parameters of the vertical mixing scheme can help increase ensemble spread, but it does not necessarily lead to better forecast skills. The vertical mixing scheme has to be tuned together with other physics schemes to achieve desired forecast performance.

The sensitivity experiments show that weaker eddy diffusion tends to produce larger ensemble spread, and vice versa. The eddy diffusion component of the vertical turbulent mixing scheme represents down-gradient transport by small eddies (Köhler et al., 2011). It tends to reduce the gradient of momentum, heat and tracers. Larger ensemble spread is often found in areas where strong gradients exist. Our hypothesis is that the change of eddy diffusion affects the ensemble spread through the change of the gradient of the model state. To testify the hypothesis, we ran two ensemble forecasts with the EDMF0 version of the TKE-EDMF scheme but use different background characteristic mixing length l_0. Figure 13 shows the change of the temperature gradient norm averaged over the ensemble members and the change of the ensemble spread due to the change of l_0 from a constant (Equation 3) to exponentially decreasing with height (Equation 4), which leads to weaker eddy diffusion. As eddy diffusion becomes weaker, the temperature gradient increases. The increment of the gradient norm that includes both horizontal and vertical components (Figure 13b) highly matches the increment in ensemble spread (Figure 13c) in terms of the patterns. Although eddy diffusion represents local vertical mixing, the change in the vertical component of the gradient (Figure 13a) only partially explains the change in ensemble spread. The local vertical mixing also modifies the horizontal gradient of the model state. The experiments confirm that the change in model state gradient

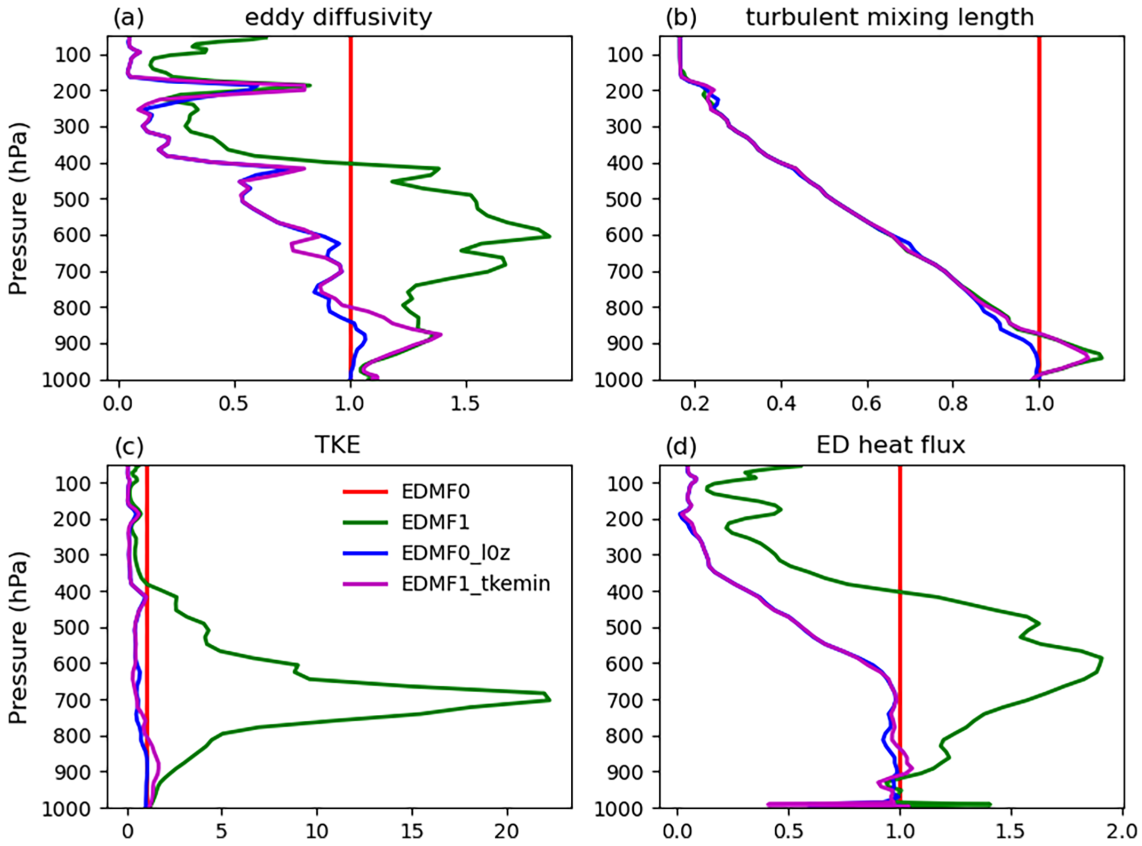

FIGURE 12 EDMF0-normalized global mean (a) eddy diffusivity for heat K_{h}; (b) turbulent mixing length l_k; (c) TKE; and (d) eddy diffusion component of heat flux $-K_{\mathrm{h}}\frac{\partial\overline{\theta}}{\partial z}$. The mean profiles are from System for High-resolution prediction on Earth-to-Local Domains (SHiELD) six-hour forecasts initialized from operational Global Forecast System (GFS) analysis at 0000 UTC on 1 July 2022. The descriptions of SHiELD forecast experiments are listed in Table 5.

largely explains the impact of changing eddy diffusion on ensemble spread.

5.2 | Systematic error in marine stratocumulus regions

The deficiency in predicting marine stratocumulus clouds is not unique to our model. The lack of representation of those low-level clouds is also found in the IFS and other model forecasts (e.g., Lopez et al., 2022; Nuijens et al., 2015). Our verification has been based on ERA5 reanalysis. Comparing the forecast against observations collected by NASA's Clouds and the Earth's Radiant Energy System (CERES) provides another independent validation. Figure 14 shows averaged 0–6-hour forecast top-of-atmosphere (TOA) net shortwave flux (incoming minus reflected shortwave radiation) biases verified against the CERES SYN1deg product (Doelling et al., 2013) for DA_CNTL and DA_EDMF1. We also include the validation for ERA5 reanalysis. Positive biases can be found near the western coastlines of the continents in both DA_CNTL and DA_EDMF1 forecasts, which indicate underestimation of shortwave radiation reflected by the low clouds and, therefore, underprediction of cloud amount or cloud cover in those areas. The problem also appears in ERA5 and is worse in DA_CNTL than in DA_EDMF1. Further away from the coastline over the tropical oceans, there are negative biases in DA_EDMF1, which suggests overprediction of cloud cover there. Lopez et al. (2022) noted that the area is dominated by open-cell convection with typical cloud sizes ranging from 2 to 20 km as revealed by NOAA's GOES image (their figures 9 and 11). This imposes challenges to properly represent those clouds even running IFS at a much higher resolution as mentioned in Lopez et al. (2022).

Now, we would like to understand how the changes in the vertical mixing scheme modify marine stratocumulus clouds or broadly speaking the structures of the stratocumulus-topped boundary layer (STBL). In order to understand the root cause of the large systematic temperature and humidity forecast errors of DA_EDMF1 in marine stratocumulus regions (Figures 5 and 8), we examined the vertical profiles of temperature, water vapor and

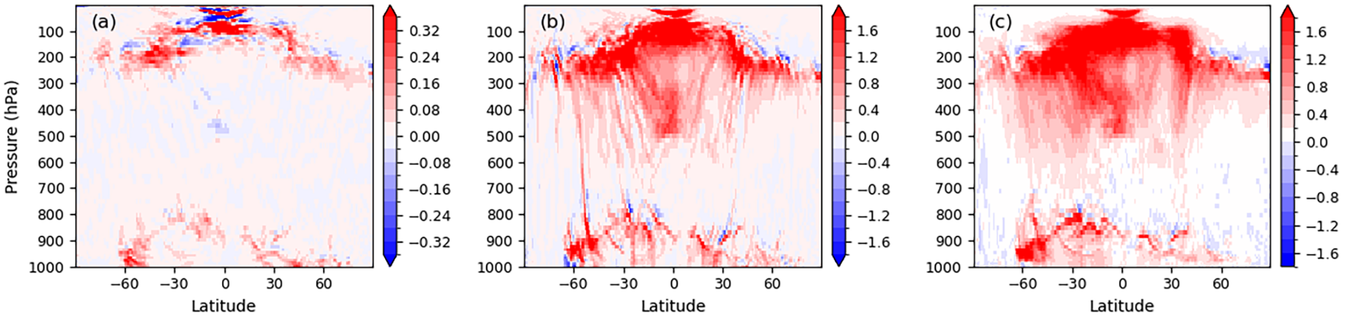

FIGURE 13 The change of zonal mean temperature (a) ensemble mean vertical gradient norm (10^{-2} K); (b) ensemble mean 3D gradient norm (10^{-2} K); (c) ensemble spread (10^{-2} K), when the background characteristic mixing length l_0 changes from a constant (Equation 3) to exponentially decreasing with height (Equation 4) in ensemble forecast using the EDMF0 version of the turbulent kinetic energy eddy-diffusivity mass-flux (TKE-EDMF) scheme. The gradient norms are first calculated for each ensemble member and then averaged over the 80 ensemble members. The gradients are calculated based on non-dimensional grid spacing. The data are from two six-hour ensemble forecasts initialized from operational Global Forecast System (GFS) ensemble analysis at 1800 UTC on 9 June 2022.

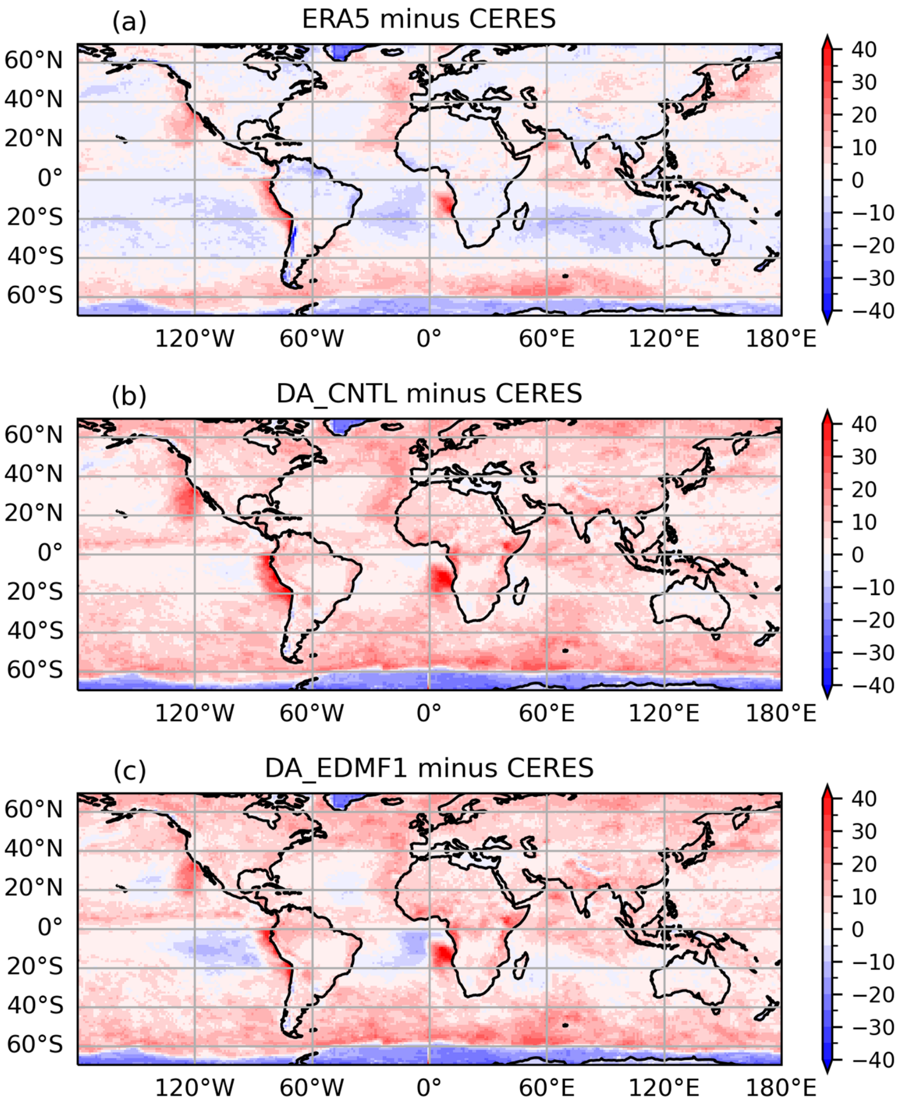

FIGURE 14 Six-hourly mean top-of-atmosphere (TOA) net shortwave flux relative differences (%) compared against Clouds and the Earth's Radiant Energy System (CERES) for (a) ERA5 reanalysis; (b) DA_CNTL; and (c) DA_EDMF1 0–6-hour forecast. The TOA net shortwave fluxes are averaged every six hours from 0000 UTC on 1 July to 1800 UTC on 31 July 2022.

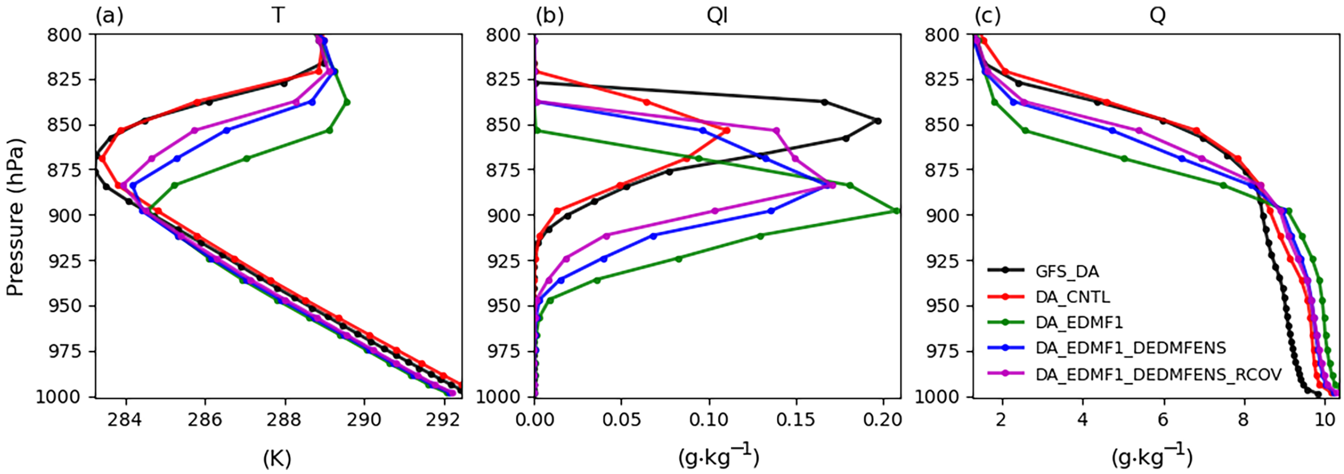

FIGURE 15 Vertical mean profiles of (a) temperature; (b) cloud liquid water specific mixing ratio; and (c) specific humidity within a 10° × 10° marine stratocumulus region centered at (15° S, 90° W). The profiles are from 6-hour forecasts initialized at 0000 UTC on 1 July 2022 of all the DA experiments listed in Table 2.

cloud liquid water mixing ratios averaged within a 10° × 10° marine stratocumulus region centered at (15° S, 90° W) off the west coast of the South American continent for the DA experiments (Figure 15). After 20 days of DA cycling, the stratocumulus cloud and inversion layer in the DA_EDMF1 forecast locate at a level much lower than DA_CNTL and GFS_DA (Figure 15b), which leads to much warmer and drier conditions around 850 hPa (Figure 15a,c). While the representation of the STBL with a number of pressure levels of the ERA5 data is rather poor, the temperature and humidity at 850 hPa show warm and dry biases in DA_EDMF1 compared with the ERA5 profile, and the biases in DA_CNTL are much smaller (not shown). The vertical position of the stratocumulus in DA_CNTL is roughly the same as that in GFS_DA, but the cloud amount is much less than GFS_DA and DA_EDMF1. The smaller cloud amount explains the much-reduced negative net TOA shortwave flux biases west of 90° W and 0° longitude over the tropical oceans in the DA_CNTL forecast but produces larger positive biases near the coastlines compared with DA_EDMF1 (Figure 14b).

Figure 15 also shows that using both versions of the TKE-EDMF scheme in the ensemble (DA_EDMF1_DEDMFENS) helps offset the systematic errors with the cloud layer moved upward. Applying inter-channel correlated observation errors for IASI and CrIS (DA_EDMF1_DEDMFENS_RCOV) pushes the cloud layer further upward and reduces temperature and humidity errors furthermore around 850 hPa. To see the effect of observation error, Figure 16 presents the temperature and specific humidity analysis increment profiles obtained from DA_EDMF1_DEDMFENS_RCOV single-cycle analyses run from the background forecasts of four DA_EDMF1_DEDMFENS cycles. Cloud water increments are also produced by the DA system, but they are zeroed out before IAU is run to initialize the forecast (Kleist et al., 2023). The upgraded observation errors produce less warming and humidifying increments below 900 hPa and create more cooling and moisturizing increments above 900 hPa. Those increments are more in favor of allocating the cloud layer at higher levels during the forecast.

We ran forecast sensitivity experiments to help us understand how the changes in the vertical mixing scheme modify the vertical structures of the STBL. We were also curious whether the STBL differences found in the reduced resolution experiments could potentially appear in the DA cycling run at the model's original horizontal resolution of C768, which we plan to do in the future. Therefore, we ran forecast sensitivity experiments at both C192 and C768 resolutions. Most of the findings with the two horizontal resolutions are consistent, except that the non-local mixing component of the scheme possesses horizontal resolution dependency. All the forecast experiments were initialized from the operational GFS analysis.

The same vertical profiles shown in Figure 15 are plotted for the six-hour forecast sensitivity experiments in Figure 17. As expected, the profile differences shown in the DA experiments (GFS_DA, DA_CNTL and DA_EDMF1) are also shown in the free forecast experiments (GFS, EDMF0 and EDMF1), but at a much-reduced magnitude in terms of the vertical location of the cloud layer, because the systematic model error got reinforced through DA cycles. When the background characteristic mixing length l_0 in EDMF0 is computed using Equation (4) (EDMF0_l0z) and the scheme becomes less diffusive, the amount of cloud liquid water is increased. However, cloud liquid

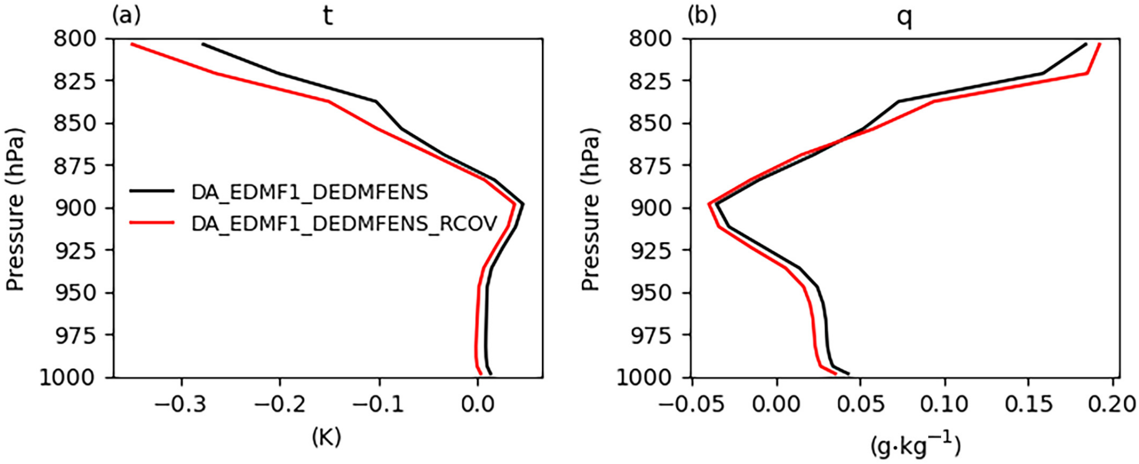

FIGURE 16 Vertical mean analysis increment profiles of (a) temperature and (b) specific humidity within the same 10° × 10° marine stratocumulus region as in Figure 15. The profiles are the mean of four random cycles from DA_EDMF1_DEDMFENS and four DA_EDMF1_DEDMFENS_RCOV single-cycle analyses run from the background forecasts of the four DA_EDMV1_DEDMFENS cycles.

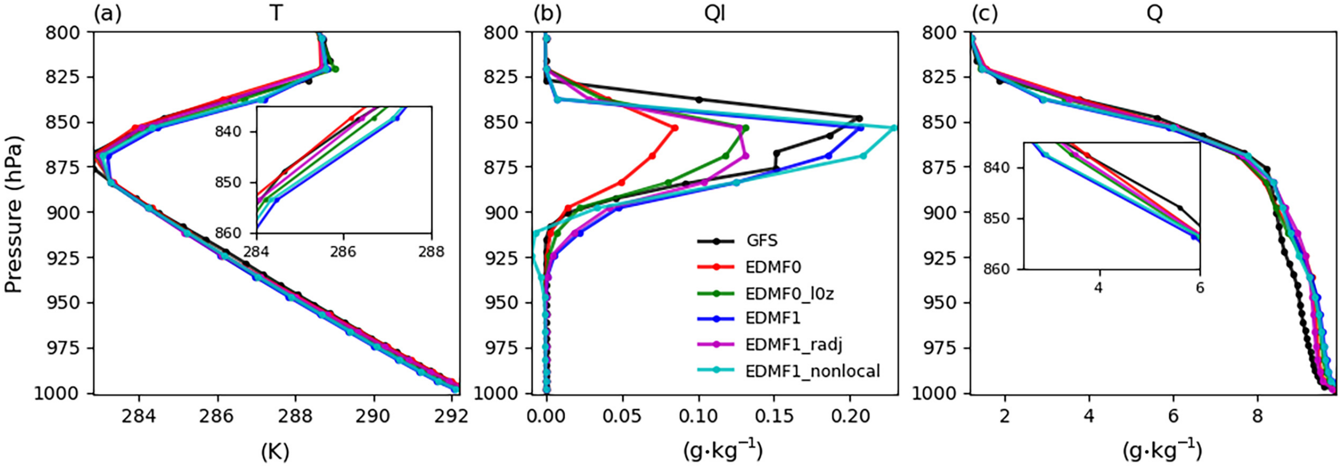

FIGURE 17 As Figure 15, but for Global Forecast System (GFS) and System for High-resolution prediction on Earth-to-Local Domains (SHiELD) free forecast experiments all initialized from operational GFS analysis at 0000 UTC on 1 July 2022. The descriptions of the SHiELD forecast experiments are listed in Table 5.

water in EDMF0_l0z is still less than that in EDMF1 and GFS. The reason is because the eddy diffusivity is enhanced based on radiative flux jump at stratocumulus cloud top in EDMF0 (Equation 7 in Table 4). When the same eddy diffusivity enhancement is applied in EDMF1 (EDMF1_radj), the cloud liquid water profile looks similar to that of EDMF0_l0z, and the temperature and water vapor profiles are much closer to EDMF0 and GFS. The total water tendency profile from a higher horizontal resolution (C768) simulation equivalent to EDMF1_radj (C768_L91_radj) in Figure 20d shows that enhancing eddy diffusivity at cloud top (Figure 20a) helps transport more water to higher levels at the cost of reducing water within the cloud layer as a result of down-gradient transport by local turbulent mixing. The mass-flux components in the TKE-EDMF scheme represent non-local transport by large eddies in the PBL (Köhler et al., 2011). When non-local mixing is turned on for hydrometeors in SHiELD forecast (EDMF1_nonlocal) as in GFS, it helps transport more cloud water from cloud base to upper levels. However, it produces a significant amount of negative cloud water mixing ratios and it has very minor impact on the cloud top and the inversion layer.

Besides the differences in the implementation of the TKE-EDMF scheme, the vertical structure of the STBL in the forecast was also found to be sensitive to model vertical resolution. To demonstrate the sensitivity, we ran SHiELD forecast experiments with four different vertical level configurations, shown in Figure 18, where L91 is the SHiELD configuration and L127 is the GFS configuration. Figure 19 shows vertical profiles of the SHiELD six-hour forecasts run at C768. As model vertical resolution increases, the stratocumulus cloud and its inversion layer move upward. The dependency of the TKE-EDMF scheme on vertical resolution is illustrated in Figure 20. Between 900 and 825 hPa, where stratocumulus clouds are located, the upper bound of the downdraft mass-flux contribution to the total water flux reaches to a higher level with higher vertical resolution (Figure 20c). The peak of the eddy diffusion within the cloud layer also shifts upward as vertical resolution increases (Figure 20a). As a result, more water vapor and cloud water are transported to higher levels with higher vertical resolution as indicated by the total water tendency in Figure 20d. Bechtold et al. (1996) tested the sensitivity of a 1D turbulence model to vertical resolution in the simulation of the STBL. They suggested that a minimum of typically three vertical levels inside the stratocumulus layer, which is about 300 m thick, is required to crudely simulate the cloud-layer dynamics. The vertical grid sizes of the L91 configuration above 925 hPa are greater than 100 m, while the number of vertical levels of the L127 configuration almost double that of the L91 within the cloud layer. The non-local mixing for hydrometeors is sensitive to both vertical and horizontal resolution. The experiments with 'nonlocal' labels in Figures 17b and 19b show that as both horizontal and vertical resolution increase, the non-local mixing produces much less negative cloud water mixing ratios. The non-local mixing is turned off for hydrometeors in SHiELD, because it produces negative values of the hydrometeors, which is worse in SHiELD than in GFS due to lower vertical resolution of SHiELD.

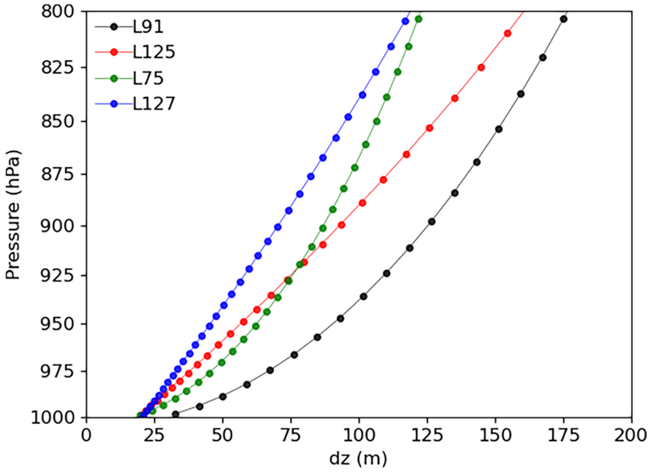

FIGURE 18 Vertical grid sizes below 800 hPa of four vertical level configurations available in System for High-resolution prediction on Earth-to-Local Domains (SHiELD).

6 | SUMMARY

A global DA system has been implemented for the GFDL SHiELD model by leveraging the GSI based GFS DA capabilities. This allows us to demonstrate the prediction skill of the model run with DA, wherein the SHiELD short-range forecast partially determines the quality of the analysis. The evaluation and the enhancement of the DA system in terms of background error (ensemble) and observation error suggests the need to adjust the DA system for different modeling systems to achieve optimal results and the advantage of developing SHiELD's own DA system. As illustrated by this study, the DA system also serves as a great tool to help us understand the model and model errors. It also provides us with research opportunities to

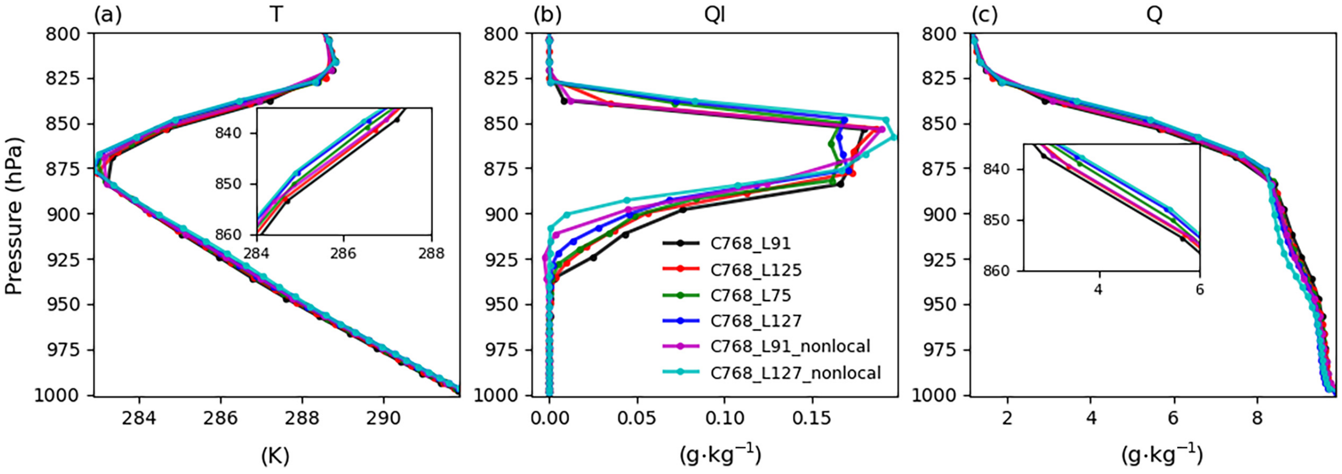

FIGURE 19 As Figure 17, but for System for High-resolution prediction on Earth-to-Local Domains (SHiELD) free forecast with the EDMF1 version of the turbulent kinetic energy eddy-diffusivity mass-flux (TKE-EDMF) scheme at C768 horizontal resolution with different vertical level configurations shown in Figure 18. Non-local mixing for hydrometeors is turned on in C768_L91_nonlocal and C768_L127_nonlocal.

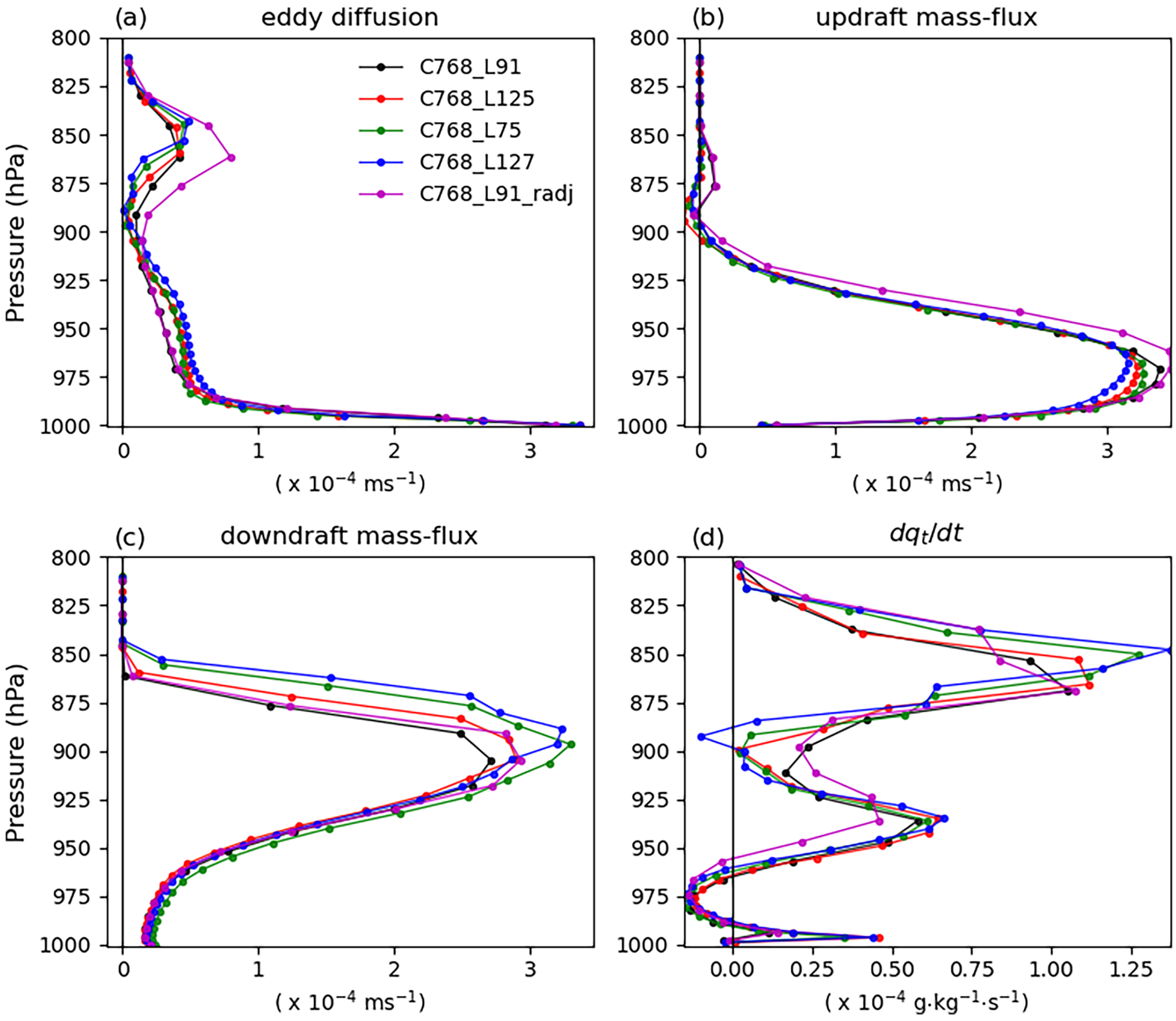

FIGURE 20 Six-hour averaged vertical mean profiles of (a) eddy diffusion; (b) updraft mass-flux; (c) downdraft mass-flux contributions to the total water flux $\overline{w'q_t'}$; and (d) total water tendency due to the turbulent kinetic energy eddy-diffusivity mass-flux (TKE-EDMF) scheme over the same 10° × 10° region as described in Figure 15. The profiles are computed from six-hour EDMF1 forecast experiments at C768 horizontal resolution with different vertical level configurations shown in Figure 18. C768_L91_radj: Same as C768_L91, but the eddy diffusivity at stratocumulus cloud top is computed with the additional $K_{\Delta R}$ as in Equation (7).

explore new DA methods and assimilation of new observations for new applications.

The performance of the SHiELD DA cycling system is evaluated by comparing it with the GFS DA cycling system run at a reduced horizontal resolution. A SHiELD replay forecast using the experimental GFS analysis was also run to understand the relative impact of model formulation and initial condition on short-range forecast. The comparison of the OmB statistics from the DA and replay experiments shows that the short-range forecasts within DA cycles are further improved where the forecast model shows better performance. However, there are exceptions, for example, in some regions covered by GPS RO bending angles. Further investigation based on ensemble reliability budget analysis reveals that the issue is attributed to the ensemble deficiencies in those problematic regions, especially in the upper troposphere and stratosphere. The problem was found to have arisen from the TKE-EDMF vertical turbulent mixing scheme. Switching to a newer version of the scheme improves the ensemble spread above 500 hPa but introduces ensemble deficiency in the lower troposphere and large systematic forecast errors in marine stratocumulus regions. To boost ensemble spread in order to make the analysis better constrained by the observations, we include both the initial and the newer versions of the TKE-EDMF in the ensemble, which significantly improves the background departure statistics.

We explored introducing interchannel error correlations for IASI and CrIS to enhance the assimilation of the radiances from the two hyperspectral IR sounders. The error covariance matrices were derived using the Desroziers et al. (2005) method. Instead of inflating the diagonal values of the diagnosed error covariance matrices as how it is done in GFS, we inflate the whole covariance matrices with the optimal inflation factors determined from ensemble replay experiments, where the pre-existing ensemble is used to run the deterministic component of the 4DEnVar system. Applying the upgraded error covariances in the full 4DEnVar system improves the short-range forecast fit to most other independent observations.

The SHiELD medium-range weather forecasts initialized from its self-cycled analysis outperform the forecast initialized from the experimental GFS analysis in most of the evaluated parameters verified against the ERA5 reanalysis. The improvement is especially remarkable in the southern hemisphere with the prediction skill of large-scale flow represented by the 500 hPa geopotential height RMSE improved more than 5% out to day 6. Enhancing the ensemble by using two versions of the TKE-EDMF scheme further improves the forecast, especially in the very few areas that are degraded with the original DA setup. The systematic forecast errors in marine stratocumulus regions are largely reduced. Upgrading observation errors to interchannel correlated error covariances for IASI and CrIS significantly improves forecast up to day 5. It further reduces the systematic forecast errors in marine stratocumulus regions.

The DA system provides a more comprehensive picture of the SHiELD model in terms of both its strengths and weaknesses. We took this opportunity to dive deeper to understand the analysis and forecast errors and their root causes. Our investigation focused on the implementation details of the TKE-EDMF scheme accompanied with sensitivity forecast and DA experiments. Our findings can be summarized as follows:

- The ensemble spread is largely affected by the strength of eddy diffusion in the TKE-EDMF scheme. Stronger local turbulent mixing, as a result of using a constant minimum turbulent mixing length throughout the vertical column in the EDMF0 version or a larger minimum TKE deduced from background diffusivity in the EDMF1 version, leads to an under-dispersive ensemble affecting analysis at different vertical levels.
- The impact of eddy diffusion on ensemble spread can be explained by its impact on the gradient of the model state.
- The systematic temperature and humidity forecast errors in marine stratocumulus regions are associated with a much lower stratocumulus cloud layer and inversion layer near the cloud top.
- Increasing eddy diffusivity within the stratocumulus cloud layer as a result of using a constant minimum turbulent mixing length or near the cloud top based on radiative flux jump tends to decrease cloud amount/-cover but helps maintain the cloud layer or cloud top at a higher level due to the down-gradient transport of local turbulent mixing.
- The vertical location of the stratocumulus cloud layer is sensitive to model vertical resolution. Water vapor and cloud water can be transported to higher levels with higher vertical resolution within the STBL.
- Turning on non-local mixing for hydrometeors helps transport more cloud water to higher levels and increases cloud cover. However, it produces negative hydrometeor mixing ratios, and the problem is more severe for lower horizontal and vertical resolutions. The negative tracer issue has been addressed in a recent upgrade of the TKE-EDMF scheme (Han et al., 2022). It would be interesting to test this upgrade in SHiELD in the future.

This study lays the foundation for future improvement of the SHiELD prediction system in terms of both model physics and the DA system. The insight gained from our system may have practical applications for GFS, as many of the software and scientific components utilized by both modeling systems are the same. The representation of model uncertainties and the handling of systematic model errors are topics that deserve further research. Stochastic process-level representation of model uncertainties such as the Stochastically Perturbed Parametrizations (SPP) scheme (e.g. Ollinaho et al., 2017) can be explored and combined with currently existing schemes in the future. Developing a bias-aware assimilation and forecast system with model error estimation and correction via methods such as machine learning are also of interest to us.

ACKNOWLEDGEMENTS

We would like to thank the NCEP/EMC DA team for discussions and answering questions about the GFS global DA system. We would also like to thank Jongil Han for upgrading the TKE-EDMF scheme and providing consultation for the scheme. Chih-Chi Hu and Yongfei Zhang are acknowledged for their comments and suggestions on an internal review of the manuscript. Several authors of this work are supported by the NOAA Research Global-Nest Initiative. Zhou and Gao are supported by awards NA18OAR4320123 and NA23OAR4320198 from NOAA, U.S. Department of Commerce.

DATA AVAILABILITY STATEMENT

The ERA5 reanalysis data, including the uncertainty estimation, are available from https://cds.climate.copernicus.eu/#!/search?tex-t=ERA5&type=dataset. The CERES data can be obtained from https://ceres.larc.nasa.gov/data/. The experimental data are available from the first author upon request.

ENDNOTES

[i] A quality control method called gross error check is normally used in data assimilation systems. Gross error check screens out observations when the difference between the observations and their background counterparts (OmB) are greater than a predefined threshold times the estimated observation error. A more accurate forecast produces smaller OmBs, which permits more observations to pass the gross error check.

[ii] The observational climatology data are 1982 to 2012 monthly mean SST data provided by NCEP.

ORCID

Mingjing Tong https://orcid.org/0000-0001-8451-7653
Xi Chen https://orcid.org/0000-0002-9952-9048

REFERENCES

Bathmann, K. & Collard, A. (2020) Surface-dependent correlated infrared observation errors and quality control in the FV3 framework. *Quarterly Journal of the Royal Meteorological Society*, 147, 408–424. Available from: https://doi.org/10.1002/qj.3925

Bauer, P., Thorpe, A. & Brunet, G. (2015) The quiet revolution of numerical weather prediction. *Nature*, 525, 47–55. Available from: https://doi.org/10.1038/nature14956

Bechtold, P., Krueger, S., Lewellen, W., van Meijgaard, E., Moeng, C.-H., Randall, D. et al. (1996) Modeling a stratocumulus-topped PBL: Intercomparison among different one-dimensional codes and with large-eddy simulation. *Bulletin of the American Meteorological Society*, 77, 2033–2042. Available from: https://doi.org/10.1175/1520-0477-77.9.2033

Berner, J., Shutts, G.J., Leutbecher, M. & Palmer, T.N. (2009) A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *Journal of the Atmospheric Sciences*, 66, 603–626. Available from: https://doi.org/10.1175/2008JAS2677.1

Bloom, S.C., Takacs, L.L., Silva, A.M.d. & Ledvina, D. (1996) Data assimilation using incremental analysis updates. *Monthly Weather Review*, 124(6), 1256–1271. Available from: https://doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2

Bormann, N. & Bauer, P. (2010) Estimates of spatial and inter-channel observation-error characteristics for current sounder radiances for NWP. Part I: methods and application to ATOVS data. *Quarterly Journal of the Royal Meteorological Society*, 136, 1036–1050. Available from: https://doi.org/10.1002/qj.616

Bormann, N., Bonavita, M., Dragani, R., Eresmaa, R., Matricardi, M. & McNally, A. (2016) Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quarterly Journal of the Royal Meteorological Society*, 142, 1767–1780. Available from: https://doi.org/10.1002/qj.2774

Bormann, N., Geer, A. & Bauer, P. (2011) Estimates of observation error characteristics in clear and cloudy regions for microwave imager radiances from NWP. *Quarterly Journal of the Royal Meteorological Society*, 137, 2014–2023. Available from: https://doi.org/10.1002/qj.833

Buizza, R., Miller, M. & Palmer, T.N. (1999) Stochastic representation of model uncertainties in the ECWMF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 125, 2887–2908. Available from: https://doi.org/10.1002/qj.49712556006

Chen, J.H., Lin, S.J., Magnusson, L., Bender, M., Chen, X., Zhou, L. et al. (2019) Advancements in hurricane prediction with NOAA's next-generation forecast system. *Geophysical Research Letters*, 46, 4495–4501. Available from: https://doi.org/10.1029/2019GL082410

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

Doelling, D.R., Loeb, N.G., Keyes, D.F., Nordeen, M.L., Morstad, D., Nguyen, C. et al. (2013) Geostationary enhanced temporal interpolation for CERES flux products. *Journal of Atmospheric and Oceanic Technology*, 30(6), 1072–1090. Available from: https://doi.org/10.1175/JTECH-D-12-00136.1

Ek, M., Mitchell, K., Lin, Y., Rogers, E., Grunmann, P., Koren, V. et al. (2003) Implementation of Noah land surface model advances in the national centers for environmental prediction operational mesoscale eta model. *Journal of Geophysical Research*, 108, 8851. Available from: https://doi.org/10.1029/2002jd003296

Eresmaa, R., Letertre-Danczak, J., Lupu, C., Bormann, N. & McNally, A. (2017) The assimilation of cross-track infrared sounder radiances at ECMWF. *Quarterly Journal of the Royal Meteorological Society*, 143, 3177–3188. Available from: https://doi.org/10.1002/qj.3171

Fairall, C.W., Bradley, E.F., Godfrey, J.S., Wick, G.A., Edson, J.B. & Young, G.S. (1996) Cool-skin and warm-layer effects on sea surface temperature. *Journal of Geophysical Research*, 101(C1), 1295–1308. Available from: https://doi.org/10.1029/95JC03190

Gao, K., Harris, L., Zhou, L., Bender, M. & Morin, M. (2021) On the sensitivity of hurricane intensity and structure to horizontal tracer advection schemes in FV3. *Journal of the Atmospheric Sciences*, 78(9), 3007–3021. Available from: https://doi.org/10.1175/jas-d-20-0331.1

Han, J. & Bretherton, C.S. (2019) TKE-based moist Eddy-diffusivity mass-flux (EDMF) parameterization for vertical turbulent mixing. *Weather Forecasting*, 34, 869–886. Available from: https://doi.org/10.1175/WAF-D-18-0146.1

Han, J., Wang, W., Kwon, Y., Hong, S., Tallapragada, V. & Yang, F. (2017) Updates in the NCEP GFS cumulus convection schemes with scale and aerosol awareness. *Weather and Forecasting*, 32, 2005–2017. Available from: https://doi.org/10.1175/Waf-D-17-0046.1

Han, J., Wei, L., Yang, F., Strobach, E., Zheng, W. & Sun, R. (2021) *Updates in the NCEP GFS cumulus convection, vertical turbulent mixing, and surface layer physics*, Vol. 505. College Park, MD: Office note (National Centers for Environmental Prediction [U.S.]). Available from: https://doi.org/10.25923/cybh-w893

Han, J., Yang, F., Montuoro, R., Li, W. & Sun, R. (2022) *Implementation of a positive definite mass-flux scheme and a method for removing the negative tracers in the NCEP GFS planetary boundary layer and cumulus convection schemes*. Office Note, Vol. 506. College Park, MD: National Centers for Environmental Prediction [U.S.]. Available from: https://doi.org/10.25923/5051-3r70

Harris, L., Chen, X., Putman, W., Zhou, L. & Chen, J. (2021) A scientific description of the GFDL finite-volume cubed-sphere dynamical Core. NOAA Technical Memorandum OAR GFDL (U.S.), 2021-001. doi: 10.25923/6nhs-5897

Harris, L., Chen, X., Zhou, L., Chen, J.-H. (2020): The nonhydrostatic solver of the GFDL finite-volume cubed-sphere dynamical Core. NOAA Technical Memorandum OAR GFDL; 2020-003. https://doi.org/10.25923/9wdt-4895.

Harris, L., Zhou, L., Lin, S.-J., Chen, J.-H., Chen, X., Gao, K. et al. (2020) GFDL SHiELD: A unified system for weather-to-seasonal prediction. *Journal of Advances in Modeling Earth Systems*, 12, e2020MS002223. Available from: https://doi.org/10.1029/2020MS002223

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049. Available from: https://doi.org/10.1002/qj.3803

Hollingsworth, A. & Lönnberg, P. (1986) The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The Wind Field. *Tellus*, 38A, 111–136. Available from: https://doi.org/10.1111/j.1600-0870.1986.tb00460.x

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

Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M. et al. (2010) Ensemble of data assimilations at EC-MWF. *ECMWF Technical Memoranda*, 636, 1–41.

Kleist, D.T., Carley, J.R., Collard, A., Emily, L., Liu, S., Martin, C.R. et al. (2023) *Current state of data assimilation capabilities at NCEP's environmental modeling center*. Office Note (College Park, MD: National Centers for Environmental Prediction (U.S.)), p. 514. Available from: https://doi.org/10.25923/pjs0-4j42

Kleist, D.T. & Ide, K. (2015) An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Monthly Weather Review*, 143, 452–470. Available from: https://doi.org/10.1175/MWR-D-13-00350.1

Kleist, D.T., Parrish, D.F., Derber, J.C., Treadon, R., Wu, W.-S. & Lord, S. (2009) Introduction of the GSI into the NCEP global data assimilation system. *Weather Forecasting*, 24, 1691–1705. Available from: https://doi.org/10.1175/2009WAF2222201.1

Köhler, M., Ahlgrimm, M. & Beljaars, A. (2011) Unified treatment of dry convective and stratocumulus-topped boundary layers in the ECMWF model. *Quarterly Journal of the Royal Meteorological Society*, 137, 43–57. Available from: https://doi.org/10.1002/qj.713

Lei, L. & Whitaker, J.S. (2016) A four-dimensional incremental analysis update for the ensemble Kalman filter. *Monthly Weather Review*, 144(7), 2605–2621. Available from: https://doi.org/10.1175/MWR-D-15-0246.1

Lei, L., Whitaker, J.S. & Bishop, C. (2018) Improving assimilation of radiance observations by implementing model space localization in an ensemble Kalman filter. *Journal of Advances in Modeling Earth Systems*, 10, 3221–3232. Available from: https://doi.org/10.1029/2018MS001468

Leutbecher, M., Lock, S., Ollinaho, P., Lang, S.T.K., Balsamo, G., Bechtold, P. et al. (2017) Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143, 2315–2339. Available from: https://doi.org/10.1002/qj.3094

Li, X. & Derber, J.C. (2008) Near sea surface temperatures (NSST) analysis in NCEP GFS. http://data.jcsda.org/Workshops/6th-workshop-onDA/Session-4/JCSDA&uscore;2008&uscore;Li.pdf

Lin, S.-J. (1997) A finite-volume integration method for computing pressure gradient force in general vertical coordinates. *Quarterly Journal of the Royal Meteorological Society*, 123(542), 1749–1762. Available from: https://doi.org/10.1002/qj.49712354214

Lin, S.-J. (2004) A "vertically Lagrangian" finite-volume dynamical core for global models. *Monthly Weather Review*, 132(10), 2293–2307. Available from: https://doi.org/10.1175/1520-0493(2004)132<2293:avlfdc>2.0.co;2

Lin, S.-J. & Rood, R.B. (1996) Multidimensional flux-form semi-Lagrangian transport schemes. *Monthly Weather Review*, 124(9), 2046–2070. Available from: https://doi.org/10.1175/1520-0493(1996)124<2046:MFFSLT>2.0.CO;2

Lopez, P., Matricardi, M., Fielding, M. (2022) *Validation of IFS+ RTTOV/MFASIS solar reflectances against GOES-16 ABI obs*. Reading, UK: ECMWF. Available from: https://www.ecmwf.int/en/elibrary/81292-validation-ifsrttovmfasis-solar-reflectances-against-goes-16-abi-obs

Lorenc, A.C., Bowler, N., Clayton, A., Pring, S. & Fairbairn, D. (2015) Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. *Monthly Weather Review*, 143, 212–229. Available from: https://doi.org/10.1175/MWR-D-14-00195.1

Magnusson, L., Chen, J.-H., Lin, S.-J., Zhou, L. & Chen, X. (2019) Dependence on initial conditions versus model formulations for medium-range forecast error variations. *Quarterly Journal of the Royal Meteorological Society*, 145, 2085–2100. Available from: https://doi.org/10.1002/qj.3545

Mouallem, J., Harris, L. & Chen, X. (2023) Implementation of the novel duo-grid in GFDL's FV3 dynamical core. *Journal of Advances in Modeling Earth Systems*, 15, e2023MS003712. Available from: https://doi.org/10.1029/2023MS003712

Nuijens, L., Medeiros, B., Sandu, I. & Ahlgrimm, M. (2015) The behavior of trade-wind cloudiness in observations and models: the major cloud components and their variability. *Journal of Advances in Modeling Earth Systems*, 7(2), 600–616. Available from: https://doi.org/10.1002/2014MS000390

Ollinaho, P., Lock, S.J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozza, A. et al. (2017) Towards process-level representation of model uncertainties: stochastically perturbed parametrisations in the ECMWF ensemble. *Quarterly Journal of the Royal Meteorological Society*, 143, 408–422. Available from: https://doi.org/10.1002/qj.2931

Orbe, C., Oman, L.D., Strahan, S.E., Waugh, D.W., Pawson, S., Takacs, L.L. et al. (2017) Large-scale atmospheric transport in

GEOS replay simulations. *Journal of Advances in Modeling Earth Systems*, 9, 2545–2560. Available from: https://doi.org/10.1002/2017MS001053

Palmer, T.N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G.J. et al. (2009) *Stochastic parametrization and model uncertainty*, Vol. 598. Reading, UK: ECMWF. Available from: https://doi.org/10.21957/ps8gbwbdv

Parrish, D.F. & Derber, J.C. (1992) The National Meteorological Center's spectral statistical-interpolation system. *Monthly Weather Review*, 120, 1747–1763. Available from: https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2

Polland, R.T., Rhines, P.B. & Thompson, R. (1973) The deepening of the wind-mixed layer. *Geophysical Fluid Dynamics*, 3, 381–404.

Putman, W.M. & Lin, S.-J. (2007) Finite-volume transport on various cubed-sphere grids. *Journal of Computational Physics*, 227, 55–78. Available from: https://doi.org/10.1016/j.jcp.2007.07.022

Rodwell, M.J., Lang, S.T.K., Ingleby, N.B., Bormann, N., Hólm, E., Rabier, F. et al. (2015) Reliability in ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 142, 443–454. Available from: https://doi.org/10.1002/qj.2663

Rodwell, M.J. & Wernli, H. (2023) Uncertainty growth and forecast reliability during extratropical cyclogenesis. *Weather and Climate Dynamics*, 4, 591–615. Available from: https://doi.org/10.5194/wcd-4-591-2023

Shutts, G., Leutbecher, M., Weisheimer, A., Stockdale, T., Isaksen, L. & Bonavita, M. (2011) Representing model uncertainty: stochastic parameterization at ECMWF. *ECMWF Newsletter*, 129, 19–24. Available from: https://doi.org/10.21957/fbqmkhv7

Takacs, L.L., Suárez, M.J. & Todling, R. (2018) The stability of incremental analysis update. *Monthly Weather Review*, 146, 3259–3275. Available from: https://doi.org/10.1175/MWR-D-18-0117.1

Tompkins, A.M. & Berner, J. (2008) A stochastic convective approach to account for model uncertainty due to unresolved humidity variability. *Journal of Geophysical Research*, 113, D18101. Available from: https://doi.org/10.1029/2007JD009284

Tong, M., Zhu, Y., Zhou, L., Liu, E., Chen, M., Liu, Q. et al. (2020) Multiple hydrometeors all-sky microwave radiance assimilation in fv3gfs. *Monthly Weather Review*, 148, 2971–2995. Available from: https://doi.org/10.1175/Mwr-D-19-0231.1

Weston, P.P., Bell, W. & Eyre, J.R. (2014) Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, 140, 2420–2429. Available from: https://doi.org/10.1002/qj.2306

Whitaker, J.S. & Hamill, T.M. (2012) Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, 140, 3078–3089. Available from: https://doi.org/10.1175/MWR-D-11-00276.1

Zhou, L. & Harris, L. (2022) Integrated dynamics-physics coupling for weather to climate models: GFDL SHiELD with in-line microphysics. *Geophysical Research Letters*, 49, e2022GL100519. Available from: https://doi.org/10.1029/2022GL100519

Zhou, L., Harris, L., Chen, J.-H., Gao, K., Guo, H., Xiang, B. et al. (2022) Improving global weather prediction in GFDL SHiELD through an upgraded GFDL cloud microphysics scheme. *Journal of Advances in Modeling Earth Systems*, 14, e2021MS002971. Available from: https://doi.org/10.1029/2021MS002971

Zhu, Y., Gayno, G., Purser, R.J., Su, X. & Yang, R. (2019) Expansion of the all-sky radiance assimilation to ATMS at NCEP. *Monthly Weather Review*, 147, 2603–2620. Available from: https://doi.org/10.1175/MWR-D-18-0228.1

Zhu, Y., Liu, E., Mahajan, R., Thomas, C., Groff, D., Delst, P.V. et al. (2016) All-sky microwave radiance assimilation in NCEP's GSI analysis system. *Monthly Weather Review*, 144, 4709–4735. Available from: https://doi.org/10.1175/MWR-D-15-0445.1

Zhu, Y., Todling, R. & Arnold, N. (2022) Observation impact and information retention in the lower troposphere of the GMAO GEOS data assimilation system. *Monthly Weather Review*, 150, 2187–2205. Available from: https://doi.org/10.1175/MWR-D-21-0334.1

How to cite this article: Tong, M., Zhou, L., Gao, K., Harris, L., Kaltenbaugh, A., Chen, X. *et al.* (2025) Improved weather predictions through data assimilation for GFDL SHiELD. *Quarterly Journal of the Royal Meteorological Society*, 151:e4930. Available from: https://doi.org/10.1002/qj.4930