Current Challenges and Future Directions in Data Assimilation and Reanalysis


Joint WCRP–WWRP Symposium on Data Assimilation and Reanalysis

What: Scientists from the three research areas—data assimilation, reanalyses, and observing systems—came together to discuss current progress and future challenges and to highlight the synergies between the communities.

When: 13–17 September 2021

Where: Online

KEYWORDS: Coupled models; Data assimilation; Reanalysis data; Machine learning

https://doi.org/10.1175/BAMS-D-21-0331.1
Corresponding author: Arianna Valmassoi, avalmass@uni-bonn.de

In final form 5 May 2022

© 2023 American Meteorological Society. This published article is licensed under the terms of a Creative Commons Attribution 4.0 International (CC BY 4.0) License.
The first Joint World Climate Research Programme (WCRP)–World Weather Research Programme (WWRP) Symposium on Data Assimilation and Reanalysis took place on 13–17 September 2021, and it was organized in conjunction with the ECMWF Annual Seminar on Observations. The last WCRP–WWRP-organized meetings were held separately for data assimilation and reanalysis in 2017 (Buizza et al. 2018; Cardinali et al. 2019). Since then, common challenges and new emerging topics have increased the need to bring these communities together to exchange new ideas and findings. Thus, a symposium involving the aforementioned communities was jointly organized by the German Meteorological Service (DWD), Hans-Ertel-Centre for Weather Research (HErZ), WCRP, WWRP, and the ECMWF Annual Seminar. Major goals were to increase diversity, provide early career scientists with opportunities to present their work and extend their professional network, and bridge gaps between the various communities.

The online format allowed more than 500 participants from over 50 countries to meet in a virtual setting, using the Gather platform (www.gather.town) as the central tool to access the meeting. A virtual conference center was created where people could freely move around and talk to other close-by participants. A lobby served as the main hub, and it connected the poster halls and the conference rooms for the oral presentations and the ECMWF seminar talks. The feedback from the participants was overwhelmingly positive.

Scientifically, the meeting offered opportunities to bring together the communities of Earth system data assimilation, reanalysis, and observations to identify current challenges, seek opportunities for collaboration, and discuss strategic planning on more integrated systems for the longer term. The contributions totaled 140 oral and over 150 poster presentations covering a large variety of topics with increased interest in Earth system approaches, machine learning, and increased spatial resolutions. Key findings of the symposium and the ECMWF Annual Seminar are summarized in the next section. The “Conclusions and remarks” section highlights the common and emerging challenges of these communities.

Topics

Operational data assimilation and infrastructure. Many participants were affiliated with operational numerical weather prediction centers and thus a significant portion of the program focused on related updates. One common theme was moving beyond medium-range
Forecasts. National centers have focused their data assimilation (DA) developments on both operational coupled Earth system models and new or improved high-resolution regional systems (cf. sections “Convective-scale data assimilation” and “Data assimilation developments”). Other areas of interest included the viability of higher-frequency updating and increased use of observations from the private sector (reported in sections “Convective-scale data assimilation” and “Observations”).

Several operational centers have embraced community model developments including data assimilation infrastructure, allowing more focus on scientific advancements and a significant reduction of duplication of efforts. The U.S.-based Joint Center for Satellite Data Assimilation (JCSDA), the Met Office, NOAA, NASA, and the U.S. Navy are collaborating on the Joint Effort for Data Assimilation Integration (JEDI; Auligne 2021), which received a lot of attention. However, the development of generic DA infrastructure is not limited to JEDI, as additional examples were shown featuring both the Parallel Data Assimilation Framework (PDAF; Nerger 2021) and the Data Assimilation Research Testbed (DART; Raeder et al. 2021). The progress presented at the symposium underlines the ongoing, successful efforts for collaborations among operational as well as nonoperational centers.

Reanalysis. Since its first implementations for the atmosphere, reanalyses have expanded to include more components of the Earth system (land, ocean, sea ice, composition) as well as regional specializations (continents, seas, and the Arctic).

Reanalyses represent the synthesis between models and observations to produce the best possible Earth system representation over multiple decades while assimilating observations at small spatiotemporal scales. Global reanalyses continue to be a foundation for global climate analysis and research, especially where observations are limited, such as the stratosphere with the Stratosphere–Troposphere Processes and Their Role in Climate (SPARC) Reanalysis Intercomparison Project (S-RIP; Fujiwara et al. 2021). Continuity in space and time is a great asset for users to study an ever-growing and diversified number of applications of physical or socioeconomic nature. Heat waves (Thomas et al. 2021a), droughts (Arshad et al. 2021), or weather/climate interactions with dust (Mytilinaios et al. 2021; Basart et al. 2021) and other aerosols (Franke et al. 2021) represent active research topics related to reanalyses. Further, renewable energy applications have lead to a growing interest in wind data from reanalyses (Morris et al. 2021; Niermann et al. 2021; Thomas et al. 2021b; Spangehl et al. 2021).

Major international centers continue both forward production of existing systems and developing new projects, given the user interests. Examples of atmospheric global reanalysis systems discussed in the symposium’s sessions are JRA-3Q at JMA (Kobayashi et al. 2021; Kosaka et al. 2021; Harada et al. 2021; Naoe et al. 2021), ERA5 and planned ERA6 at ECMWF (Hersbach et al. 2021; A. Bell et al. 2021; Muñoz-Sabater et al. 2021), as well as MERRA-2, the planned MERRA-3, and GEOS-R21C at NASA GMAO (Bosilovich et al. 2021; El-Akkraoui et al. 2021). Many of these current and future full observing system reanalyses include an ensemble component to also provide an uncertainty estimate. Further, they exhibit a progressively increased resolution for longer time periods up to three-quarters of a century (e.g., 20-km JRA-3Q, 30-km ERA5). With continuing data rescue efforts (Andersson et al. 2021), reconstructions of even longer periods become possible such as 20CRv3 from NOAA/CIRES/DOE (Slivinski et al. 2021) reaching back to 1836. Several ocean reanalysis products were also presented, such as the CMCC Global Ocean Reanalysis System (Banerjee et al. 2021) and the ECMWF ORAP6 ocean and sea ice reanalysis (Zuo et al. 2021).

Due to growing computing capabilities and ongoing research, regional reanalyses are expanding into many areas at higher resolutions. Examples covered were regional reanalysis activities at DWD (Kaspar et al. 2021), the Copernicus regional reanalysis (Schimanke et al. 2021), a 5-km regional reanalysis over Japan (Fukui et al. 2021), the Indian Monsoon Data
Assimilation and Analysis (IMDAA) regional reanalysis over the Indian monsoon region (Rani et al. 2021), a high-resolution reanalysis for the Mediterranean Sea (Escudier et al. 2021), and a 20-yr high-resolution Red Sea reanalysis (Sanikommu et al. 2021). These datasets allow for the in-depth evaluation and study of the local weather and climate, not readily captured in global systems. Further, the Arctic continues to be of interest for reanalysis efforts (e.g., C3S Copernicus Arctic Regional Reanalysis; Schyberg 2021), owing to the limited observations, changing environment and development of sea ice and glacial representations in models.

As regional reanalyses are reaching kilometer-scale resolutions, significant interactions not only with the reanalysis producers (e.g., Kaspar et al. 2021; Fourrie et al. 2021), but also with the convective-scale DA community are envisaged.

**Convective-scale data assimilation.** While convective-scale DA (CSDA) has been a focus of NWP research and developments for a while, the interaction between CSDA and reanalysis communities is expected to continue and grow. Further, CSDA is moving beyond regional applications as global NWP approaches the kilometer scales and high-resolution observational systems such as Doppler velocities (Lippi et al. 2021) can be utilized. CSDA continues to be improved and fostered by the development of regional NWP and the inclusion of corresponding observing systems at kilometer scales (e.g., Hu et al. 2021; Carley et al. 2021; Hernandez-Banos et al. 2021).

The symposium included discussions on specific challenges of CSDA, such as error representation of fine-resolution observations compared to coarse-resolution simulations, multiscale and non-Gaussian errors, and model errors that contribute to nonlinear error growth. Assimilating observations at CSDA scales, such as radar, satellite, and lightning observations, can improve analyses and forecasts (Miyoshi et al. 2021; Combarnous et al. 2021; Deppisch et al. 2021). Further, innovative observations, such as crowd-sourced data, have also shown promising results (Paschalidi et al. 2021, and presentations in the “Observations” section).

Convective-scale dynamics are driven by instabilities that require increased spatial resolutions $O(0.1–1)$ km and temporal resolution from hours to minutes. For nonlinear regimes, non-Gaussianity can become more severe, leading to suboptimal solutions in ensemble filters conditioned on Bayesian assumptions. In this regard, advanced DA strategies, like nowcasting objects, grafting look-alike storms, and warm bubbles, were proposed (Neef et al. 2021; Sodhi and Fabry 2021; Janjić et al. 2021).

Another strategy that has proven to be effective is the use of high-frequency observations through smaller perturbations and more linear error growth (Ruiz et al. 2021). Along these lines Miyoshi et al. (2021) assimilated phased-array radar observations every $30$ s with 1,000 ensemble-members and at 500-m spatial resolution in real time. Advances have also been reported for non-Gaussian DA algorithms which may bring significant benefits for CSDA in years to come (see section “Data assimilation developments” for more details).

A major issue for high-resolution observations are significant observation error correlations. Yang et al. (2021) investigated this effect in Doppler velocities and found that including these correlations explicitly improves the representation of extreme precipitation events. Similar positive impacts have been reported by Fujita et al. (2021). To represent model errors in CSDA, other strategies were also proposed, including additive noise, uncertainties in physical parameterizations, and errors resulting from unresolved scales (Janjić et al. 2021; Waller et al. 2021).

Further, multiscale features of CSDA can be addressed by blending algorithms that combine large- and small-scale analyses from a global model and a regional model (Schwartz et al. 2021). In addition, adaptive localization approaches including scale-dependent localization, variable-dependent localization, and correlation-dependent localization as well as the impact of large ensembles on localization were discussed (H. Wang et al. 2021; Necker et al. 2021; Duc et al. 2021).
Data assimilation developments. A wide variety of topics were presented in the broader area of methodological advances in DA, from the ongoing research of ensemble approaches to machine learning methods to coupled systems. As in previous symposia, there were submissions focusing on more theoretical aspects of the DA generally split between algorithms and information flow. For NWP applications, Kalman filters were very present in ensemble DA [e.g., Pannekoucke et al. (2021) with the parametric Kalman filter, Raboudi et al. (2021) with the time-dependent observation noise inclusion, or X. Wang et al. (2021) with the multi-scale local Gain Form Ensemble Transform]. Also, hybrid DA algorithms were popular with applications such as integrated hybrid EnKF (Lei et al. 2021) or a 20-yr eddy-resolving reanalysis for the Red Sea (Sanikommu et al. 2021). Further, the non-Gaussian DA algorithms presented include non-Gaussian hybrid data assimilation (e.g., the maximum likelihood ensemble filter; Fletcher et al. 2021), non-Gaussian observation errors (Hu and Van Leeuwen 2021), and particle filters. The latter approach continues to show great progress and potential as it is transitioned to more complex model applications. Examples of particle filters were given by Schenk et al. (2021) with a 4D-localized version; Feng et al. (2021) with an improved local non-linear ensemble transform filter; Kawabata and Ueno (2021) with an implementation of non-Gaussian assimilation methods for error estimation at the convective-scale; and Kotsuki et al. (2021), who showed increased stability for a local particle filter with a Gaussian mixture extension in intermediate global circulation model.

The emergence of machine learning (ML) and deep learning (DL) techniques in recent years were also visible at the symposium. The major aim was to use ML to support specific DA aspects, accelerate/improve parts of the DA or to address model errors and parameter uncertainties. ML was used to support computationally expensive traditional DA methods by Barthelemy et al. (2021), who proposed to use a neural-network to emulate a high-resolution background from a low-resolution NWP. Further, ML was employed to improve error estimates (e.g., Farchi et al. 2021) or ensure more consistent error models also in non-Gaussian DA (Hossen et al. 2021). The more classical ML use of model parameter estimation was addressed by Legler and Janjić (2021) in a shallow-water convective-scale model, or Bocquet et al. (2021), who used the dual EnKF to learn the state, global and local parameters based on either covariance or domain localization.

Coupled DA in Earth system models was discussed with respect to strengths and current limitations. Bhargava et al. (2021) showed the benefit of weakly coupled atmosphere–ocean SST assimilation, Reichle et al. (2021) presented atmosphere–land DA with the assimilation of SMAP brightness temperature, Kleist et al. (2021) detailed the recent and future updates for weakly coupled DA in the NOAA GFS, and O’Kane et al. (2021) showed a 60-yr retrospective product using a strongly coupled DA with ocean–sea ice and biogeochemistry components. Other work presented focused on the important topic of coupled covariances within the different Earth system compartments (e.g., Tang et al. 2021; Smith et al. 2021). Several authors mentioned the problems of underdispersed land ensembles (e.g., Reichle et al. 2021), with Draper et al. (2021) presenting vegetation and soil parameter perturbations as a possible way forward. Several authors also outlined methodological considerations of the choices between strongly and weakly coupled DA, e.g., different time scales in the components (de Rosnay 2021) or tuning of the preexisting system degrading overall efficacy of coupled model state estimates (Reichle et al. 2021).

Some contributions highlighted issues of assimilating new spatial observations in coupled DA (for more on observations, refer to section “Observations”). However, neglecting new observing systems in one compartment may imply benefits to other quantities. In fact, the assimilation of surface temperatures for the ocean (While et al. 2021) and land (Valmassoi and Keller 2021) show an improvement in the assimilated variable, but also an increase in biases in other parts of the system, at the base of the ocean mixed layer and the 2-m temperature, respectively.
**Observations.** A more detailed look at the state of observations for Earth system sciences was enabled by the joint aspects of the symposium with the concurrent ECMWF Annual Seminar, where recent Observing System Experiment results were shown for different Earth system components (Bormann 2021; Remy 2021; Kolassa 2021). Advances in data assimilation are enabling an even better utilization of observations, as exemplified by Laloyaux (2021) (estimating model error) and Prigent (2021) (surface observations). While observations for nonatmospheric components of the Earth system do not yet, in general, fully meet requirements, encouraging progress was reported for land, ocean, sea ice, snow, and composition by several Annual Seminar speakers (Inness 2021; Benedetti 2021; Charlton-Perez 2021).

Several presentations in the symposium discussed new or not yet fully exploited satellite observations. The assimilation of targeted geostationary hyperspectral sounder observations with high temporal resolution was shown to be beneficial for typhoon forecasts (Han et al. 2021) and visible satellite channels improved convective scale cloud and radiation forecasts (Scheck et al. 2021). Chandramouli et al. (2021) presented an online method for estimating nonlinear biases to improve the assimilation of all-sky satellite observations. The rising number and importance of GNSS observations was noted by Bormann (2021), while impressive results with the Aeolus research mission was shown as well (Koepken-Watts 2021). Aeolus lidar wind observations revealed to be in good agreement with Environment and Climate Change Canada (ECCC) forecasts and ERA5 data (Chou et al. 2021) with their assimilation in NWP exhibiting a significant positive impact at DWD (Cress 2021). Contributions also addressed the potential impacts on planned satellite missions such as the Meteosat Third Generation (MTG) and the potential for assimilation of the flash extent accumulation (Combarnous et al. 2021) or infrared and microwave channels (Villeneuve et al. 2021).

While 99% of observations assimilated in global NWP are from satellites (English 2021), multiple speakers at the ECMWF Annual Seminar noted that in situ observations remain very important, not least for the Earth’s surface (e.g., Siddorn 2021; Ingleby 2021; Sandells 2021). Recent developments using ground-based remote sensing in data assimilation were presented, especially to characterize the state of the atmospheric boundary layer. Here, an observation gap—especially for temperature, humidity, and wind profiles—remains, as satellites are not able to resolve the ABL to the required extent. Ceilometers, Doppler lidars (Kayser et al. 2021), and microwave radiometers (Knist et al. 2021) were shown to be on their way to becoming operationally assimilated. For Doppler lidar networks, Nomokonova et al. (2021) showed significant potential for short-term wind forecasts in the ABL, especially at wind turbine hub heights, with Observation System Simulation Experiments (OSSEs). Further, the EU COST action PROBE (Profiling the Atmospheric Boundary Layer at European Scale; www.probe-cost.eu; Loehnert et al. 2021) is coordinating efforts to set standards (Lehmann et al. 2021) and exploit (Merker et al. 2021) newly organized ground-based remote sensing networks for data assimilation applications.

There is also a rising interest in the assimilation of crowd-sourced or otherwise nonconventional observations, which are inexpensive, but may require more effort for quality control and bias correction. It was demonstrated that assimilating bias-corrected temperature and humidity observations from citizen weather stations leads to significant improvements in the first six forecast hours (Paschalidi et al. 2021). Vehicle-based sensor measurements (Acevedo et al. 2021; Z. Bell et al. 2021) and commercial wind turbine data (Kelbch and Keller 2021) were also discussed as promising observations for data assimilation and reanalysis.

The seminar reflected a large, growing, and increasingly diverse global observing system. With this increase of complexity, the number of observation anomalies also increases making quality control and cross calibration very important (Dahoui 2021; Zhang 2021; Bathmann 2021). A more continuous approach to data screening and assimilation may enable some problems to be spotted early (Lean 2021). Waller (2021) highlighted the need of a more detailed
look at sources of observations’ error correlation, notably the spatial one. The seminar pre-
sented also the expected future evolution of both the space sector (Coppens 2021; Donlon
2021; Donoho 2021) and in situ observations (Randriamampianina 2021).

Conclusions and remarks
Through its design, the first Joint WCRP–WWRP Symposium on Data Assimilation and
Reanalysis together with the ECMWF Annual Seminar on observations allowed for the three
scientific communities to come closer together highlighting the impressive progress and
addressing common challenges.

A major theme of the symposium was a trend toward the consolidation of data assimila-
tion efforts with respect to shared infrastructure. With the increasingly growing complexity
of current DA, such approaches aim to share the costs and split the efforts for achieving
progress without significant increases in expenses by facilitating collaboration and exploit-
ing synergies in operational centers and research institutes. Therefore, the successful use of
observation types and methodological advances can be accelerated. However, we also found
that these efforts are not at its goal, as a major recommendation from the symposium was
the need for data-sharing infrastructure to facilitate expensive experiments that now only a
few centers are able to implement.

As model resolutions in global NWP and reanalysis are constantly increasing, the need
for methods to appropriately and efficiently handle the assimilation of data at multiple spa-
tiotemporal scales is growing further. Issues formerly related only to limited area models are
becoming more relevant also for global DA, such as multiscale features, non-Gaussian error
distributions, and the model errors which contribute to fast error saturation. To tackle these
challenges, several algorithms have been proposed and successfully applied. In particular,
efficient non-Gaussian methods such as localized particle filters have been rapidly maturing
in recent years and may soon begin to replace the first variational- or Kalman filter–based
assimilation systems.

Further, machine learning methods are being introduced to the DA community at an as-
tonishing speed with many promising results. However, more research is required to more
fully exploit these techniques and to address the high-dimensional problems for which inex-
pensive solutions are needed, such as the estimation of complex model errors or observation
error covariances.

The advances in data assimilation are enabling an even better utilization of observations,
both in their characterization and in high-frequency assimilation. Several contributions have
demonstrated that there is a significant potential for new observations and underutilized remote
sensing data to be exploited in data assimilation, especially at the convective scale and through
exploitation of “all-surface/all-sky” radiance assimilation within the context of coupled DA.

The popularity of reanalysis is continuing to increase, due to its invaluable contribution
to climate monitoring and weather forecasting as well as in research and the commercial
sector, and with new fields of applications still emerging. The applications are not limited to
atmospheric-related events, since several products are available for the ocean, land surface,
atmospheric composition, and cryosphere. Reanalyses are continuously improved not only
by increasing their spatial resolution and assimilating more observing systems/observation
types, but also by extending them from atmosphere-focused products to a more complete
Earth system. Events such as this symposium help to bridge the gaps between different
reanalyses, working toward the plan to provide coupled products in the coming years (e.g.,
ECMWF and NASA).

Coupled Earth system modeling and assimilation is also becoming more prevalent both
in research and operational applications related to NWP, especially in connection with the
pursuit of seamless prediction systems. The associated multiscale challenges have to be
addressed in DA: be it the necessity to adapt the algorithms to include observing systems for other Earth system components, the need for weakly or strongly coupled DA, or the challenges of properly initializing each component of the Earth system. In this regard, a close collaboration between NWP and reanalysis communities is essential to exploit existing and future synergies.

**Acknowledgments.** We thank the Hans Ertel Centre for Weather Research (HErZ), the German Meteorological Service (DWD), the World Weather Research Programme (WWRP) and the World Climate Research Programme (WCRP), and especially the European Centre for Medium-Range Weather Forecasts (ECMWF) for their support in organizing the symposium.

References


Laloyaux, P., 2021: Bias aware data assimilation.
Morris, M., P. Kushner, and O. Mercan, 2021: Using reanalysis to assess
Lei, L., Z. Wang, and Z. Tan, 2021: Integrated hybrid data assimilation for an
Lean, P., 2021: Practical challenges using observations in operational numerical
Prigent, C., 2021: Challenges in microwave radiative transfer parameterization of the Earth surface, passive and active, from low to high frequencies, for both atmosphere and surface characterizations. ECMWF Annual Seminar on Observations, Online, ECMWF, https://vimeo.com/605649028.


Waller, J., 2021: Observation error covariances, can we handle spatial OBS error correlations. ECMWF Annual Seminar on Observations, Online, ECMWF, https://vimeo.com/605543387.


