JGR Atmospheres

RESEARCH ARTICLE
10.1029/2025JD044491

Special Collection:
Climate and weather extremes in a warming climate: Processes, Prediction and Projection

Key Points:

- Assimilating bias-corrected PWV data reduced errors and improved 4–6-hr monsoon MCS precipitation predictions
- A 100-km spacing between synthetic GPS stations effectively captured convective patterns, outperforming both the 50- and 200-km setups
- Results emphasize the need to optimize GPS density for MCS forecasts and identify terrain-related model limitations

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:
S. Shohan,
sshohan@arizona.edu

Citation:
Shohan, S., Koch, S. E., Castro, C. L., Arellano, A. F., Kay, J., Risanto, C. B., et al. (2025). Impact of assimilating GPS precipitable water vapor on simulations of two North American monsoon convective events using observing system simulation experiments. *Journal of Geophysical Research: Atmospheres*, *130*, e2025JD044491. https://doi.org/10.1029/2025JD044491

Received 5 JUN 2025
Accepted 14 AUG 2025

Impact of Assimilating GPS Precipitable Water Vapor on Simulations of Two North American Monsoon Convective Events Using Observing System Simulation Experiments

Samkeyat Shohan[1], Steven E. Koch[1], Christopher L. Castro[2], Avelino F. Arellano[1], Junkyung Kay[2], Christoforus Bayu Risanto[3], Tammy M. Weckwerth[2], James O. Pinto[2], and David K. Adams[4]

[1]Department of Hydrology & Atmospheric Science University of Arizona, Tucson, AZ, USA, [2]NSF National Center for Atmospheric Research, Boulder, CO, USA, [3]Vatican Observatory, Vatican City, Holy See, [4]Instituto de Ciencias de la Atmósfera y Cambio Climático, Universidad Nacional Autónoma de México, Mexico City, Mexico

Abstract This study evaluates the impact of assimilating precipitable water vapor (PWV) within an observing system simulation experiment (OSSE) framework to improve forecasts of monsoonal mesoscale convective systems (MCSs) in Arizona. Two contrasting case studies differing in convective forcing, longevity, intensity, and coverage are analyzed using a 40-member ensemble of 1.8-km resolution Weather Research and Forecasting (WRF) convective-permitting model (CPM) simulations including the Data Assimilation Research Testbed (DART) system. Synthetic PWV data are derived from a nature run (NR) and bias corrected using real GPS-derived PWV observations from a campaign during the North American monsoon (NAM) season 2021. These synthetic PWV are assimilated in an inferior model simulation called the control run (CR) to avoid the identical twin problem. Horizontal GPS station spacing experiments (e.g., superobbed, 50 km, 100 km, and 200 km) are conducted to identify configurations that maximize forecast skills. Assimilating the synthetic PWV reduces mean errors (∼2 mm) and dry bias during the first 4–6 hr of the predictions using analyses improved with PWV data assimilation. The 100-km GPS network optimally captures convective precipitation patterns, outperforming coarser (200-km) and finer (50-km) grids due to an improved representation of moisture and winds afforded by PWV data assimilation at the appropriate scales. Topography strongly influences moisture distribution, with elevation-dependent biases, overestimation in low elevations (0–500 m), underestimation in midelevations (500–2,000 m), and systematic high-elevation (>2,000 m) biases due to vertically integrated PWV constraints. This study provides actionable insights for optimizing GPS network design and improving convective-scale modeling in arid/semiarid regions.

Plain Language Summary The effect of assimilating precipitable water vapor (PWV) data on the prediction of monsoonal mesoscale convective systems (MCSs) in Arizona is the subject of the study. By perturbing 40-member ensemble of high-resolution WRF model simulations within the framework of an observing system simulation experiment (OSSE), we investigate the impact of varying GPS network density on forecast skill. Assimilation of PWV data reduces early forecast errors and dry biases notably, with best moisture and wind improvements found for 100-km GPS station spacing. Our results emphasize the role for observation network design in improving convective-scale prediction over complex terrain.

1. Introduction

The North American monsoon (NAM) is a period of enhanced precipitation that occurs in the southwestern United States during the summer months, lasting from approximately late June to mid-September. Moisture for the monsoon comes predominantly at lower levels from the Gulf of California and eastern Pacific while upper-level moisture contributions come from the Gulf of Mexico resulting in convective rainfall (Adams & Comrie, 1997). It has also been argued that local evapotranspiration additionally plays an important role, especially in the latter part of the monsoon (Dominguez et al., 2016). NAM precipitation is important in this region, for example, as a source of water for agriculture and human use and sustainability of ecosystems (Sheppard et al., 2002). The NAM is associated with the occurrence of severe weather from thunderstorms, with unique hazards, including microbursts (Willingham et al., 2011), haboobs (or dust storms) (Raman et al., 2014), flash floods, landslides (Feng et al., 2013; Ray et al., 2007), and lightning (Holle & Zhang, 2023). The overall objective of this study is to improve representation of these monsoon thunderstorms through better specification of moisture in the initial condition within a convective-permitting regional atmospheric modeling system for two NAM extreme weather events in Arizona during the 2021 monsoon.

The initiation and development of deep convection during the NAM is a result of dynamic and thermodynamic processes and moisture transport in complex terrain. In the context of convective-permitting models, predictability of monsoon thunderstorms is challenged by the relative lack of observed meteorological data at the meso-beta scale (10–100 s of km) and relative lack of knowledge of the upscale growth process of monsoon convection (Moker et al., 2018; Risanto et al., 2021). In this work, we address these two challenges. An ensemble system can capture uncertainties in initial conditions and model physics, leading to a more realistic range of possible outcomes (Leutbecher & Palmer, 2008). Our research aims to improve convective-permitting numerical weather prediction (NWP) model forecasts of extreme monsoon precipitation events by addressing data gaps and enhancing ensemble model system designs to better specify model initial conditions and forecast uncertainty. In this regard, multiscale ensemble prediction and data assimilation (DA) methods are essential. Comprehensive observations of water vapor distribution and mesoscale circulation systems, gathered from soundings, ground-based profiling, and airborne systems, can improve prediction of organized convection, including mesoscale convective systems (MCS) and its associated hazards (S. L. Benjamin et al., 2004; Caumont et al., 2016; Kay et al., 2022; Madhulatha et al., 2020; Nakamura et al., 2004; Pan et al., 2021; Risanto et al., 2021; Sun et al., 2020). While the assimilation of observations obtained with advanced ground-based and airborne sensor technologies has proven effective for improved prediction of MCSs over the Great Plains (Fast et al., 2019), similar studies that consider the NAM region in the southwestern U.S. are lacking.

Data assimilation (DA) systematically integrates real-time observational data into analyses used to initialize NWP models. The evolution of NAM thunderstorms depends on mountain-valley circulations in complex terrain, thermodynamic conditions, and synoptic-scale dynamics, with substantial spatiotemporal variability (Adams & Comrie, 1997; Adams & Souza, 2009; Bieda et al., 2009; Johnson et al., 2007). The interplay of these factors make forecasts of monsoon thunderstorms challenging, even at convective-permitting resolution (Higgins & Gochis, 2006; Hu et al., 2023; Navon, 2009). A commonly used DA technique is the ensemble Kalman filter (EnKF; J. L. Anderson, 2001). EnKF has been implemented in this study using the Weather Research and Forecasting (WRF-ARW; Skamarock et al., 2019) model in conjunction with the Data Assimilation Research Testbed (DART; J. Anderson et al., 2009). In addition, the presence of complex topography, such as in Arizona, impacts the representativeness of near-surface and surface observations. For instance, two stations at different heights may have very different atmospheric states over shorter horizontal distances. DART relies on the ensemble technique under the assumption of well-constrained observation error and background error, but in situations of high variability associated with terrain, this becomes increasingly hard to assure (Hacker et al., 2018).

GPS meteorology (GPS-Met) refers to the utilization of Global Positioning System (GPS) signals to retrieve atmospheric data, primarily water vapor information. Precipitable water vapor (PWV), or the total water vapor in a column of the atmosphere, is estimated through the measurement of GPS signal delay as it propagates through the atmosphere (Bevis et al., 1992; Guerova et al., 2016; Rocken et al., 1993). S. G. Benjamin et al. (2010) evaluated the impact of eight observation data types using the Rapid Update Cycle (RUC) model, focusing on short-range forecasts (3–12 hr) of wind, temperature, and relative humidity; the results showed most data types improved forecasts, but particularly GPS PWV data. Short-term NWP forecasts depend on high-resolution and frequently updated observational data to accurately capture complex meteorological phenomena during the NAM (Serra et al., 2016). Risanto et al. (2021) likewise showed improvement in short-term forecasts of moisture fields by assimilating GPS PWV into a CPM during NAM events.

Observing system simulation experiments (OSSEs) are a valuable tool for evaluating the benefits of new observational data in weather forecasting. Central to these experiments is the nature run (NR), which is designed to provide an accurate high-resolution depiction of atmospheric conditions (Zeng et al., 2020). Maintaining a clear distinction between the NR and the forecast model is vital, as the forecast model generally operates at a lower resolution and accuracy. Ensuring that simulated observations accurately reflect a real condition is also important, including coverage and measurement errors. “Quick OSSEs” (Hoffman & Atlas, 2016) target specific meteorological events, typically operate without calibration, and assume that uncertainties in the background and observation errors maybe derived from unbiased Gaussian distributions. Observation errors used in the OSSE should be tuned such that the standard deviation of the background error variance in observation space is similar to the standard deviation of the background ensemble variance plus observation error variance. This is termed the consistency ratio and is achieved by modifying the observation errors or the ensemble spread so that this ratio is nearly 1.0 (T. A. Jones et al., 2016).

OSSE calibration consists of adjusting the synthetic observation errors so that the accuracy and impact of the observing systems in the OSSE are comparable to the accuracies and impacts of the same observing systems in the real world, if known (Privé et al., 2023). Calibration using an observing system experiment (OSE) can help align the performance of observing systems in OSSEs with their real-world effectiveness (Boukabara et al., 2018; Hoffman & Atlas, 2016; Zeng et al., 2020), because actual observations are used in an OSE, rather than the numerically generated data produced by an OSSE NR. Validating an OSSE using an OSE offers a significant advantage in assessing how new observations impact short-term forecasting. This study benefits from the availability of an OSE for calibration, as discussed in Risanto et al. (2025). Calibrated OSSEs allow for controlled testing of hypothetical observing systems by assimilating simulated observations into an NWP model.

The purpose of this study is to determine whether the assimilation of PWV data sets in an OSSE framework can significantly improve the forecast accuracy of monsoonal MCSs in Arizona compared to simulations without these observations. The impacts are examined for two cases that differed significantly in terms of the nature of the forcing, longevity, precipitation intensities, and coverage of convection. Moreover, tests are performed in which the horizontal spacing between GPS stations is varied to obtain a configuration that maximizes forecast skill.

The present paper is structured as follows: We start with describing the step-by-step methodological approaches in Section 2, including setup of the model and data assimilation system, and the calibration processes whereby the OSE results of Risanto et al. (2025) are used to optimize the OSSE output and a description of the experiments performed. Then, in Section 3, we describe the aspects of the two chosen NAM cases, and in Section 4, we present our results with discussion. Finally, in Section 5, we summarize our major findings.

2. Data and Methods

2.1. GPS-Met Stations Network

A network of 31 GPS-Met instruments collected PWV observations throughout Arizona and adjacent states to capture the spatial variability of PWV during a field campaign during the 2021 NAM. Data collection includes GPS PWV estimates from eight campaign Arizona GPS receivers that were deployed between 1 July 2021 and 10 September 2021, in addition to measurements from 23 permanent GPS receivers in the southwestern United States affiliated with the Surface Observing Upper-air Meteorological Instrumentation (SUOMI) network. However, the distribution of these stations across the landscape was not uniform, both in terms of spatial placement and elevation. As illustrated in Figure 1a, the stations reveal two prominent clusters, marked by black ellipses, which exhibit significant variability in elevation. A strong motivation of the 2021 GPS-Met campaign was to cover strong gradients in elevation and assess the effects on data assimilation. However, this uneven distribution presents specific challenges for the DART system when attempting to apply a consistent horizontal cutoff radius, as suggested by J. L. Anderson (2012), to all the stations involved in the study.

To address the issues arising from this spatial and elevation variability and reduce the impact of correlated errors in the observations, we implemented "superobbing." *Superobbing* is a technique commonly used in DA for data reduction. This technique is applied where multiple high-resolution observations within a defined spatial and temporal window are present and systematically averaged to produce a single representative observation, known as a *superobservation*. This process helps reduce representativeness errors and observation density, especially in areas with clustered data (Geer & Bauer, 2010). To effectively apply superobbing, we first examined the frequency distribution of the minimum distances between the GPS stations. This analysis revealed a bimodal distribution: the first mode peaks at approximately 25 km, indicating a dense clustering of stations within close proximity of each other, while the second mode peaks at around 100 km, suggesting another set of stations situated further apart (Figure 1c). Utilizing these identified modes, we proceeded to average the observations of stations that fell within the 25-km distance bracket, enhancing the overall representativeness and reliability of the PWV measurements. The GPS stations that are averaged do not exceed ∼500 m of elevation differences. Consequently, the station distances are more uniformly spaced at distances of between 100 and 200 km (Figure 1d).

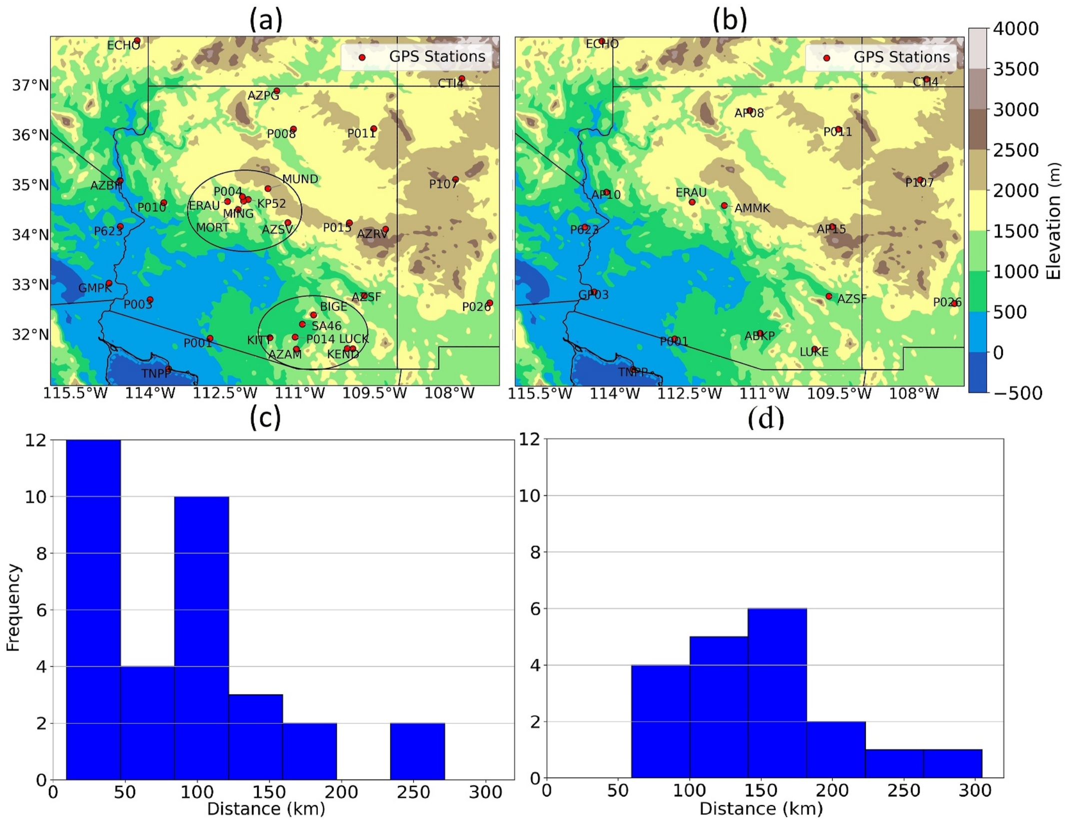

Figure 1. (a) GPS stations deployed during the 2021 field campaign in Arizona and nearby states, with the two mesoscale clusters of densely distributed stations shown by the ellipses, (b) "superobbed" stations using the average station distances, (c) frequency distribution of minimum distances between raw stations, and (d) frequency distribution of minimum distances between superobbed stations.

2.2. Weather Research Forecasting (WRF) Model

This study utilizes the Advanced Research WRF (WRF-ARW) version 4.2 (Skamarock et al., 2019). The configuration of the model including physics parameterizations (Table 1) is nominally similar to a real-time forecasting system in the Southwest U.S.

This model uses initial and lateral boundary conditions from the High-Resolution Rapid Refresh (HRRR) version 4 to produce its forecasts. A two-way nested domain structure is employed, where the outer domain (d01), with a grid spacing of 5.4 km, encompasses the western United States (Figure 2a), and an inner domain (d02) spanning Arizona and New Mexico uses a mesh with 1.8-km grid spacing. The high spatial resolution of the inner domain qualifies it as a convective-permitting model capable of capturing the complex topographically driven dynamics associated with thunderstorms and heavy rainfall events (Moker et al., 2018). The model includes 71 vertical

Table 1
The WRF-ARW Model Physics Parameterization Schemes

Schemes	WRF physics options
Microphysics scheme	Thompson graupel scheme (Thompson et al., 2008)
Longwave radiation scheme	RRTMG scheme (Iacono et al., 2008)
Shortwave radiation scheme	RRTMG scheme (Iacono et al., 2008)
Land surface scheme	Noah-MP (Niu et al., 2011)
Planetary boundary layer scheme	MYNN 2.5 level TKE scheme (Nakanishi & Niino, 2009)
Cumulus scheme	d01: Kain-Fritsch (Kain, 2004); d02: none

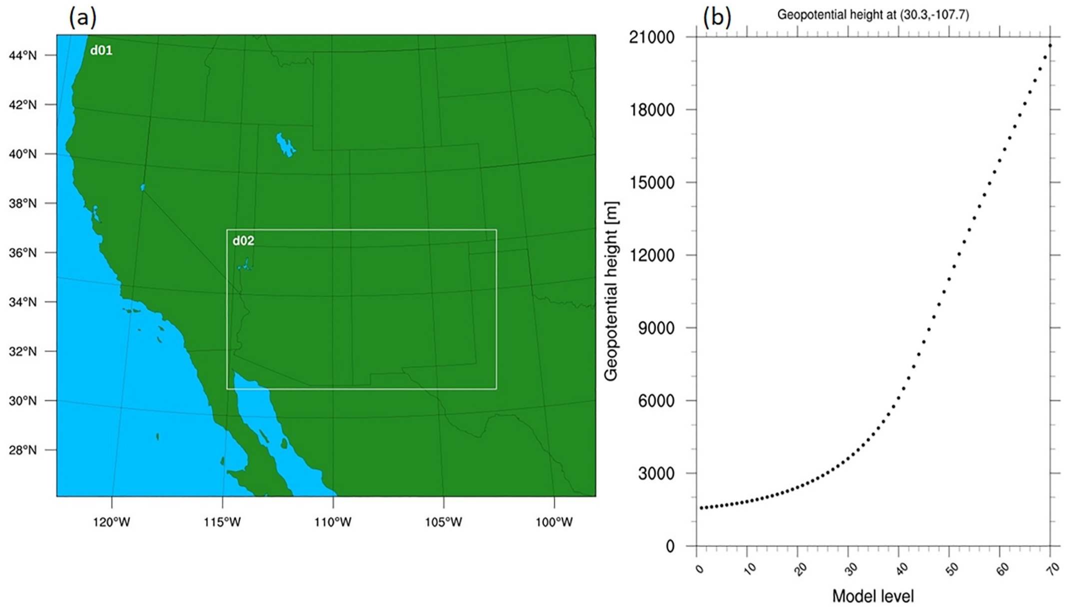

Figure 2. (a) The WRF-ARW model domains. Horizontal resolution, d01: $\Delta x = \Delta y = 5.4$ km (231 × 230), d02: $\Delta x = \Delta y = 1.8$ km (334 × 301) (b) The 71 vertical levels employed in the model are shown in geopotential height (m MSL).

levels, with the lowest levels more densely packed so boundary layer processes are adequately represented (Figure 2b) and for future OSSE work involving ground-based profiling systems.

2.3. Data Assimilation Research Testbed (DART)

This study utilizes the ensemble adjustment Kalman filter (EAKF; J. L. Anderson, 2001) algorithm implemented within Manhattan release (J. Anderson et al., 2009; DART, 2023). DART is an advanced open-source community software facility specifically designed for ensemble data assimilation to improve the accuracy and reliability of NWP. The EAKF algorithm employed in DART is an advanced statistical data assimilation method for integrating observed data into a model background field to enhance the accuracy and reduce uncertainties of the analyses used to initialize NWP models. DART has been successfully coupled with a variety of Earth system models that operate across scales, encompassing both regional and global dimensions (Romine et al., 2014; Schumacher & Clark, 2014; Sobash & Stensrud, 2015).

PWV is the vertically integrated water vapor in a column. DART uses a forward operator to convert model fields into simulated observation values for specific times and locations. The forward operator performs the integration over all vertical levels in the atmosphere for a given grid point, *i*, *j*:

$$\mathrm{PWV}_{i,j} = \frac{1}{\rho_w g} \sum_{\mathrm{k=l}}^{\mathrm{nk}-1} \frac{1}{2}\left(\mathrm{MR}_k + \mathrm{MR}_{k+1}\right)\left(p_k - p_{k+1}\right)$$

where, ρ_w is liquid water density, g is the acceleration due to gravity, MR is the water vapor mixing ratio where MR = QVAPOR/(1 − QVAPOR), and QVAPOR is the specific humidity. $k = 1$ denotes the lowest model level, nk indicates the total number of model levels, and p represents the model-equivalent pressure interpolated from domain d02. The observation error is set at 0.075 cm, based on the measurements from Adams et al. (2011). A similar observation error value was used in the study by Risanto et al. (2021).

Using observations too distant from a grid point can create false correlations due to the finite size of the ensemble in data assimilation. Covariance localization helps address this issue by limiting the influence of observations on state variables to nearby grid points. We apply a localization function based on the Gaspari and Cohn (1999) correlation function, using a half-width cutoff radius of 223 km in the horizontal direction and 3.5 km in the vertical direction. The horizontal localization radius value was chosen after conducting many experiments to find maximum positive forecast impacts given the GPS station network. The sensitivity experiments are conducted for

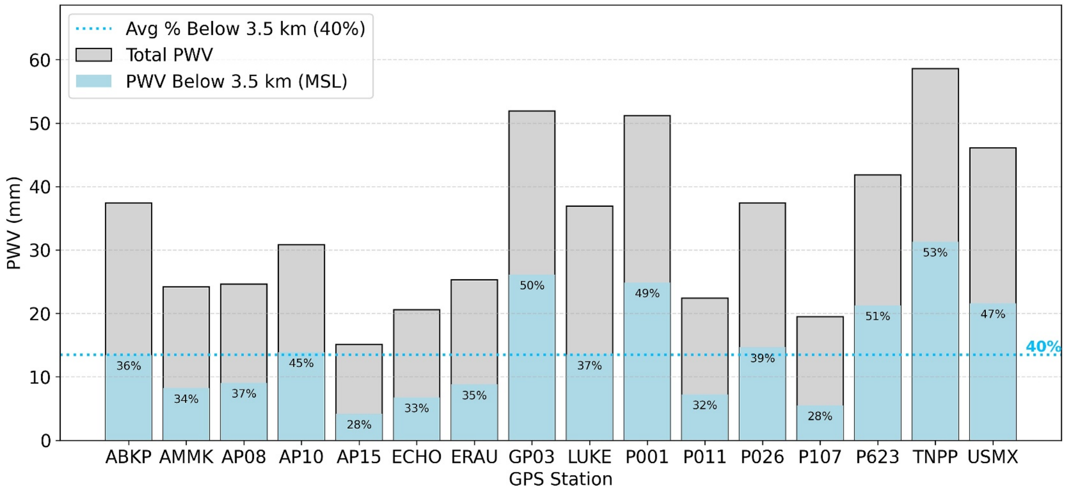

Figure 3. Total GPS PWV with percentages of moistures from the surface to a height of 3.5 km MSL at specific GPS stations as derived from the real GPS-Met observations at 2021-08-15 18UTC.

horizontal localization between 100 and 500 km by keeping all other configurations same. This range is deemed to be appropriate considering the extent of the observation locations. The vertical localization radius was set at 3.5 km MSL because nearly 40% of atmospheric moisture was found below this altitude (Figure 3). Other aspects of the DART configuration used in this study follow that described in Risanto et al. (2021).

2.4. Experimental Setup

This study focused on two mesoscale convective events during NAM season in August 2021, coincidentally occurring on consecutive days while the GPS-Met campaign network was operational. This observational network provides us a unique opportunity to validate our OSSE with an OSE (Risanto et al., 2025).

2.4.1. DA Configurations

The OSSE approach using GPS-equivalent synthetic PWV observations involves three key steps. First, we generated an NR by objectively choosing the best performing member among a 40-member ensemble forecasting system using randomly perturbed initial conditions (J. L. Anderson, 2001; Whitaker & Hamill, 2002) that produces the most representative state of the atmosphere as explained in Section 3. Second, we extracted synthetic PWV observations by applying bilinear interpolation (P. W. Jones, 1999) from the NR using the superobbed stations (Figure 1b) and then added random noise based on observation error specification. Third, we employed the WRF/DART system to assimilate the synthetic PWV observations extracted from the NR over a 6-hr window into a coarser model and subsequently executed an 18-hr "free forecast."

We selected two different operational forecast model data sets, namely the HRRR and the Global Forecast System (GFS), to serve as initial and lateral boundary conditions (ICBC) for driving the WRF model. The HRRR data set is used to generate the NR due to its ability to represent the true state of the atmosphere and is more compatible with CPM than the GFS. The NR consists of two domains: the outermost domain (d01) has grid spacings of 5.4 km, while the innermost domain (d02) has grid spacings of 1.8 km. Similarly, the GFS was used as the forcing data to produce the control run (CR), which also consists of two domains: the outermost domain (d01) has grid spacings of 9 km, while the innermost domain (d02) has grid spacings of 3 km. The number of vertical levels (71) is the same for both models. The idea is to extract synthetic observations from the NR and assimilate them into the CR to improve the forecast (Figure 4). Therefore, the CR and NR differ significantly in terms of horizontal grid spacing and boundary conditions, which helps to prevent identical twin problems that can affect interpretation of the results if the NR and the target DA system are too similar (Hoffman & Atlas, 2016). A CR of no data assimilation (NODA) is performed to evaluate the forecast improvement.

A 6-hr spin-up period was used for all runs, followed by 6 hr of assimilation from 12:00 to 18:00 UTC, and then an 18-hr "free deterministic forecast" was initialized by the ensemble mean analysis at 18:00 UTC (Figure 4). During

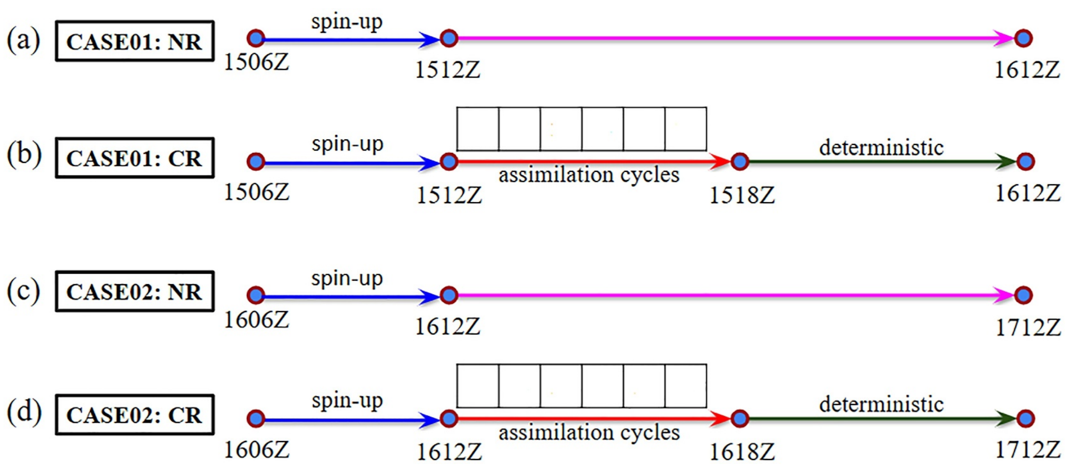

Figure 4. Schematic illustration of the experimental procedure used for the NR and CR in the two case studies involving a 6-hr spin-up period, a 6-hr data assimilation window using hourly updates for the CR, and an 18-hr free forecast. Note that the labels (ddhhZ) below each filled circle represent the day of August and the time of day in UTC.

each DA hourly cycle, we updated 20 state variables (e.g., U, V, W, PH, T, MU, QVAPOR, U10, V10, T2, Q2, PSFC, QCLOUD, QICE, QRAIN, QSNOW, QGRAUP, TH2, QNRAIN, and QNICE) by calculating regression coefficients between PWV data and state variables using the prior ensemble covariance. The quality of DA was evaluated at the end of the assimilation cycle.

Our primary goal is to assess the impact of different spacings and varying numbers of stations on monsoon convection forecasts. Thus, we also performed experiments with uniform station spacings of 50, 100, and 200 km over the state of Arizona (Figure 5). For this purpose, we established a horizontal localization radius of 2.23 times the spacing between the stations as employed in objective analysis schemes (Koch et al., 1983). This choice is made to address potential numerical issues to effective resolutions within the model, while also ensuring a consistent framework for our OSE and OSSE. Furthermore, we maintained a vertical localization distance of 3.5 km across all experiments. This kept a standardized environment for evaluating our results, allowing us to effectively isolate the influence of station spacing on forecast sensitivity and enabling us to draw more accurate conclusions from our findings.

2.4.2. NR Selection

We collected real PWV observations from the special GPS-Met network field campaign conducted in August 2021. This provided us an opportunity to select the best candidate for the NR from 40-member ensemble outputs and to apply bias correction to our extracted synthetic PWV observations from the NR before performing DA. The

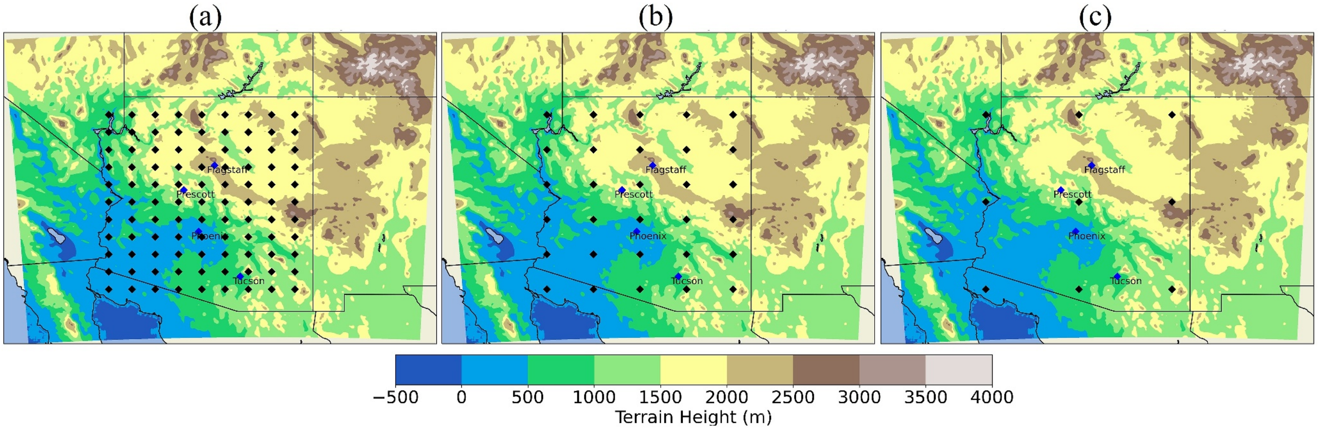

Figure 5. The uniform station experiments using site spacings of (a) 50, (b) 100, and (c) 200 km.

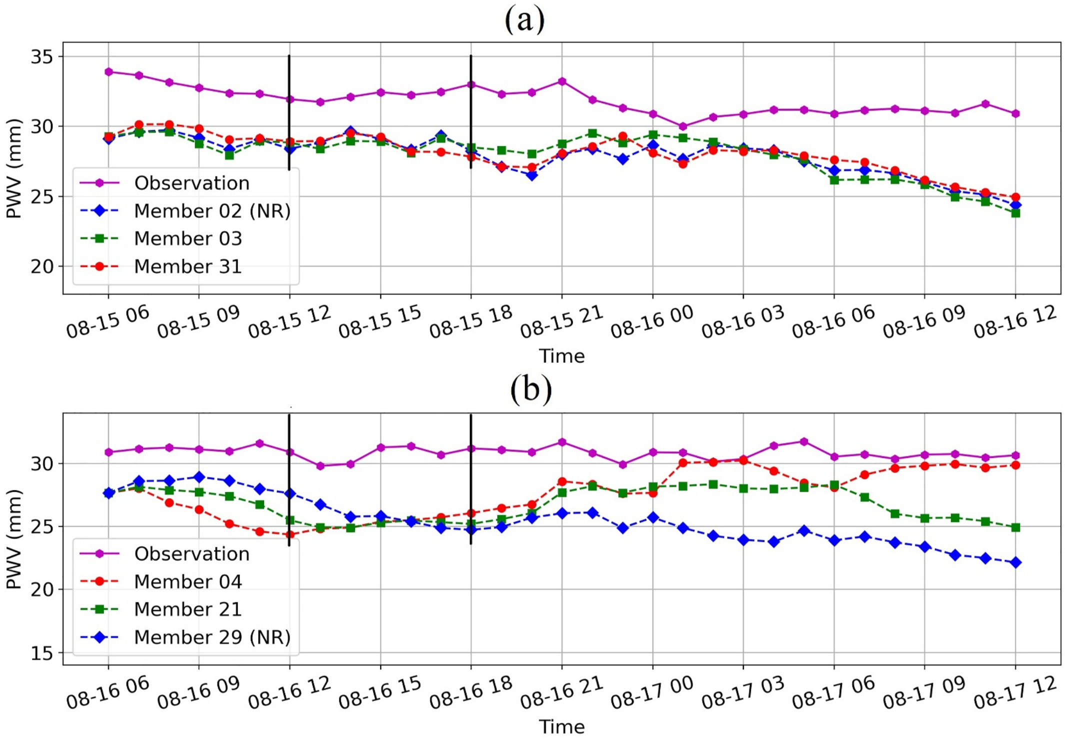

Figure 6. GPS station-averaged PWV time series for the three best performing NR candidates based on accumulated precipitation for (a) the 15 August 2021 event and (b) the 16 August 2021 event compared to actual GPS PWV observations (purple lines). The black vertical lines represent the 6-hr assimilation window for each case, used for determining the smallest bias in PWV.

selection process for the NR began by evaluating PWV fields from the 40-member ensemble and accumulated precipitation in comparison to observations from the multiradar/multisensor system (MRMS, NCEP, 2018). This evaluation allowed us to narrow down the candidates to three ensemble members for both OSSE cases for further analysis. Subsequently, we extracted synthetic PWV data based on the original GPS stations from the 2021 campaign.

As shown in Figure 6, all three of the best ensemble members underestimated the station-averaged PWV during the assimilation cycling period when compared to the real PWV observations for both cases. The mean biases for each of the three ensemble members chosen as a potential NR candidate are given in Table 2.

Table 2
Mean Biases in PWV (mm Water Vapor) for Both Cases During the 6-hr Assimilation Window for Three Ensemble Members Deemed to Be the Best Forecasts

Ensembles	Mean biases (mm)
CASE 01	
Member: 02 (NR)	−3.44
Member: 03	−3.57
Member: 31	−3.58
CASE 02	
Member: 04	−5.41
Member: 21	−5.44
Member: 29 (NR)	−4.82

Despite the dry bias present among all members of the ensemble, the three best performing members displayed the lowest mean bias statistics among the 40 members. Accordingly (Table 2), ensemble member 02, which exhibits a mean bias of −3.44 mm for case 01, and ensemble member 29, with a mean bias of −4.82 mm during the 6-hr assimilation time window, were selected as candidates for NR consideration.

The 30-hr accumulated precipitation forecasts from the selected NR (Figure 7) shows precipitation features comparable to those of MRMS. However, despite our best efforts, both underestimation and overestimation persist across the domain in both cases due to inherent model limitations such as model resolution, physical parameterizations, and the highly nonlinear and scale interactive of evolving deep convection.

2.4.3. Elevation-Dependent Bias Correction

Significant vertical gradients, resulting from the uneven distribution of atmospheric moisture, are effectively measured by the network of GPS stations

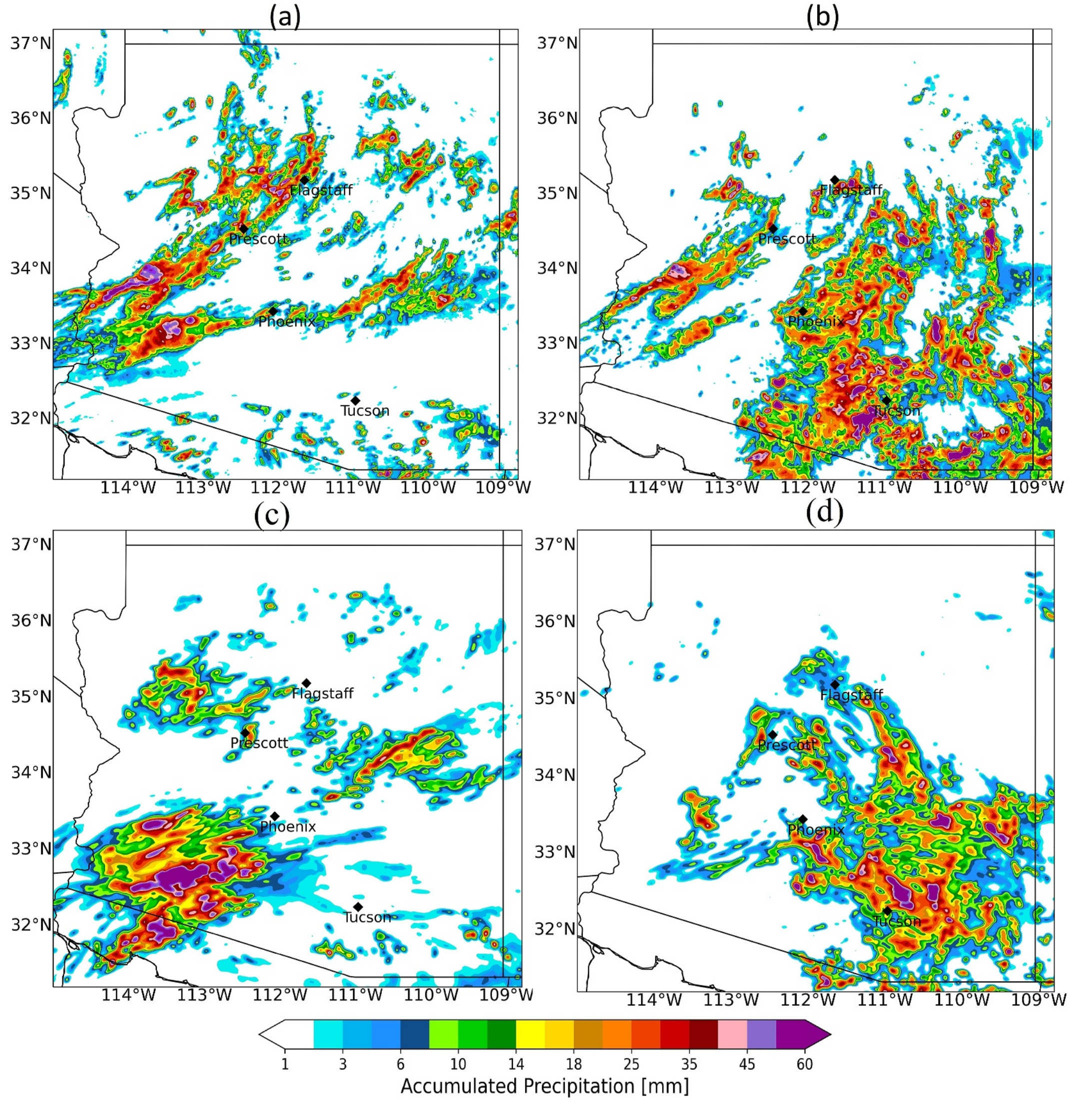

Figure 7. 30-hr accumulated precipitation for MRMS and selected NR. (a) Case 01 MRMS, (b) Case 02 MRMS, (c) Case 01 NR, and (d) Case 02 NR. Plots are valid for the accumulation periods of 20210815:06Z–20210816:12Z ((a) and (c)) and 20210816:06Z–20210817:12Z ((b) and (d)).

situated at various elevations, creating elevation-dependent biases. To address this issue, we conducted an elevation-dependent heterogeneous bias correction for the PWV values from the nature run. First, we identified mean biases in the NR values across different station elevation bins of 0.5 km from the surface to 3 km throughout the DA window (Table 3). Second, we corrected these elevation-dependent mean biases in the NR values and used these corrected values in all of the DA experiments by applying the terrain elevation-dependent bias corrections listed in Table 3.

Generally, these results indicated that negative (dry) biases increase from sea level up to 1.5 km and then decrease with increasing elevation up to ~2.5 km, above which the sign of the bias changes. A similar pattern is observed in the scatterplot (Figure 8), though case 02 exhibits significantly greater variations in biases for the observing sites at lower elevations compared to case 01.

Table 3
Mean Biases for Each 0.5 km Elevation Bin for Both OSSE Cases

Elevation bins (km MSL)	Mean biases (mm): Case 01	Mean biases (mm): Case 02
0.0–0.5	−2.24	−6.65
0.5–1.0	−2.27	−5.36
1.0–1.5	−6.38	−8.24
1.5–2.0	−5.28	−4.97
2.0–2.5	−0.98	−0.17
2.5–3.0	6.65	4.18

3. Synoptic Overview of the Two NAM Cases

3.1. Case 01: UTC 18 August 15

Three conditions are necessary to allow the development of deep moist convection: moisture, instability, and a lifting mechanism (Houze, 1993). In this case, both moisture and instability are ample. For instance, PWV values greater than 40 mm are observed over southern Arizona and northern Mexico and MUCAPE is more than 3,000 J/kg over the majority of southern and central Arizona, indicating a vigorous and very unstable atmosphere. But lifting still remained an issue. There was no indication of organized, widespread ascent, nor were there apparent mesoscale-scale lifting mechanisms (e.g., outflow boundaries or terrain-driven ascent) in evidence at this point in the forecast.

The large-scale synoptic setting is marked by a broad upper-level ridge across the western U.S. This setup allowed for rather weak dynamic forcing of ascent, with 500 hPa absolute vorticity values of moderate magnitude. An inverted trough (IVT) was noted just north of NW Mexico; however, at the time, the associated flow is weakly curved and not confidently supported by jet dynamics. Consequently, there are little deep vertical motion and any ascent more likely due to thermodynamic processes rather than by coherent synoptic-scale dynamics.

Deep shear (850–400 hPa bulk shear (Newman & Johnson, 2012)) is mostly weak to moderate (10–20 m s^{-1}) across AZ. Deep vertical shear is not, however, necessary for the onset of deep convection, but is essential for the development and maintenance of MCSs.

3.2. Case 02: UTC 18 August 16

A little more dynamically active synoptic situation is found for case 02. A broad trough axis from Ohio southward into western New Mexico provided the large-scale ascent, and a robust magnitude IVT prevailed over NW Mexico. As a result, there is a stronger and more organized dynamic forcing for ascent as than Case 01, but large-

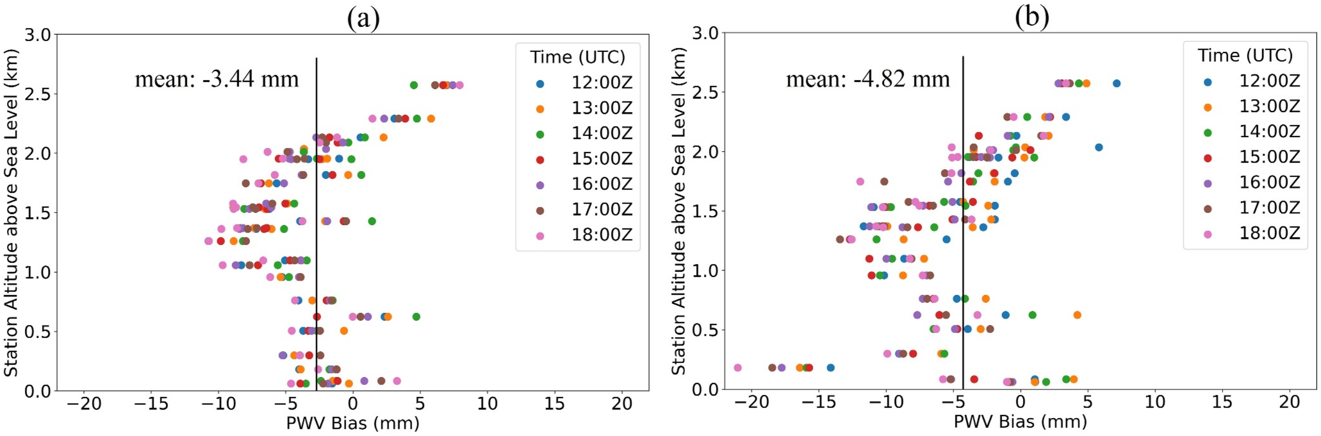

Figure 8. Scatterplot of PWV biases for each 0.5 km elevation bin during the 6-hr assimilation cycles for GPS stations operated for (a) Case 01 and (b) Case 02. Each color represents a distinct hour during the assimilation cycle. The solid black vertical lines represent the mean biases for each case.

scale vertical motion is still not the most dominant forcing. Once again, the three ingredients necessary for deep convection, moisture, instability, and lift, are there.

Moisture is a bit more abundant than the preceding day, with PWV values now greater than 50 mm in many locations across southern and central Arizona. The thermodynamic setup is largely unchanged; MUCAPE values in excess of 3,000 J/kg returned (especially in the vicinity of Phoenix and throughout southern Arizona). Crucially, in this situation, the instability is more colocated in space with the regions of maximum dynamic lift, and hence, the overlap of ingredients for convection is more favorable.

Deep layer shear is also stronger and more coherent in Case 02, notably over southwestern Arizona. The magnitude and direction-speed shear features of the 850–400 hPa bulk shear are much more intense than is previously observed. Although not a necessary condition for the onset of convection, this increased shear promoted a more favorable convective regime for the maintenance and organization of MCSs, with some potential for long-lived or rotating storms in southeastern Arizona (Figure 9).

As in the Case 01 study, large-scale lift is not the dominant factor in initiation, and thus, mesoscale lifting mechanisms are still critical. In Case 02, terrain-induced lifting over features such as the Mogollon Rim and White Mountains is probably instrumental for originating initial convection. Moreover, after the onset of initial convection, cold pool outflow boundaries are generated, which resulted in significant sources of mesoscale ascent needed to maintain and organize subsequent convective development throughout the event.

4. Results and Discussion

4.1. Evaluation of OSSE Forecasts

This section presents a comprehensive evaluation of the forecast outcomes for both OSSE cases, focusing specifically on synthetic PWV and accumulated precipitation metrics. Impacts are assessed relative to the NODA forecasts from the CR runs.

4.1.1. Case 01

At the end of the DA cycles, we observe a noticeable positive impact resulting from the assimilation of synthetic PWV data, in terms of a considerable reduction in the mean errors (>2 mm) and halving of the dry bias (Figure 10b) compared to the NODA. This reduction in the errors is reflected in the root mean square error (RMSE) (Figure 10c), which was calculated relative to the actual GPS-PWV observations. Throughout the assimilation, most of the OSSE experiments exhibit a consistent trend of steady adjustments. However, the OSSE utilizing the sparse 200-km array diverges from this pattern, moving closer to the NODA state by UTC 18 August 15. Upon transitioning to the forecast phase, a continued reduction in bias is evident (Figure 10b). Further analysis of the magnitude of RMSE reveals that there is a gradual increase in error values across all OSSE experiments following the assimilation cycles (Figure 10c), such that by the eight-hour forecast interval, these errors are no less than those of the NODA counterpart. To summarize, these statistics indicate that the initial error in the OSSE system is drastically reduced by the end of the data assimilation window and that the synthetic PWV data continue to have a positive impact in terms of mean bias reduction, and to a lesser extent, the RMSE for nearly 8 hr. Thus, the assimilation of synthetic PWV data had an enduring positive influence on forecast performance.

In Figure 11a, the temporal evolution of convection seen in MRMS reveals two prominent clusters of precipitation accumulation (Figure 11a and 13): the first cluster being situated primarily to the west of Flagstaff in northern Arizona, while the second cluster is located to the east of Phoenix. In addition to these primary regions, there is also scattered precipitation observed along the Mexican border and in northwestern New Mexico. The NR forecast effectively captures both of the significant precipitation clusters identified by MRMS, showing a close alignment with the observed data, thus, supporting our decision regarding use of this particular member of the ensemble to represent nature. In contrast, the NODA poorly depicts these mesoscale precipitation clusters in terms of both coverage and intensity. Turning to the OSSE results, we first note that the OSSE-superobbed station run managed to capture most of the precipitation features indicated by MRMS, although with less pronounced representation, particularly in the region near Flagstaff, while over-predicting the convection in northwest New Mexico. The OSSE utilizing a 100-km array subjectively demonstrates superior performance in accurately capturing precipitation features across much of central and southern Arizona compared to the other two uniform

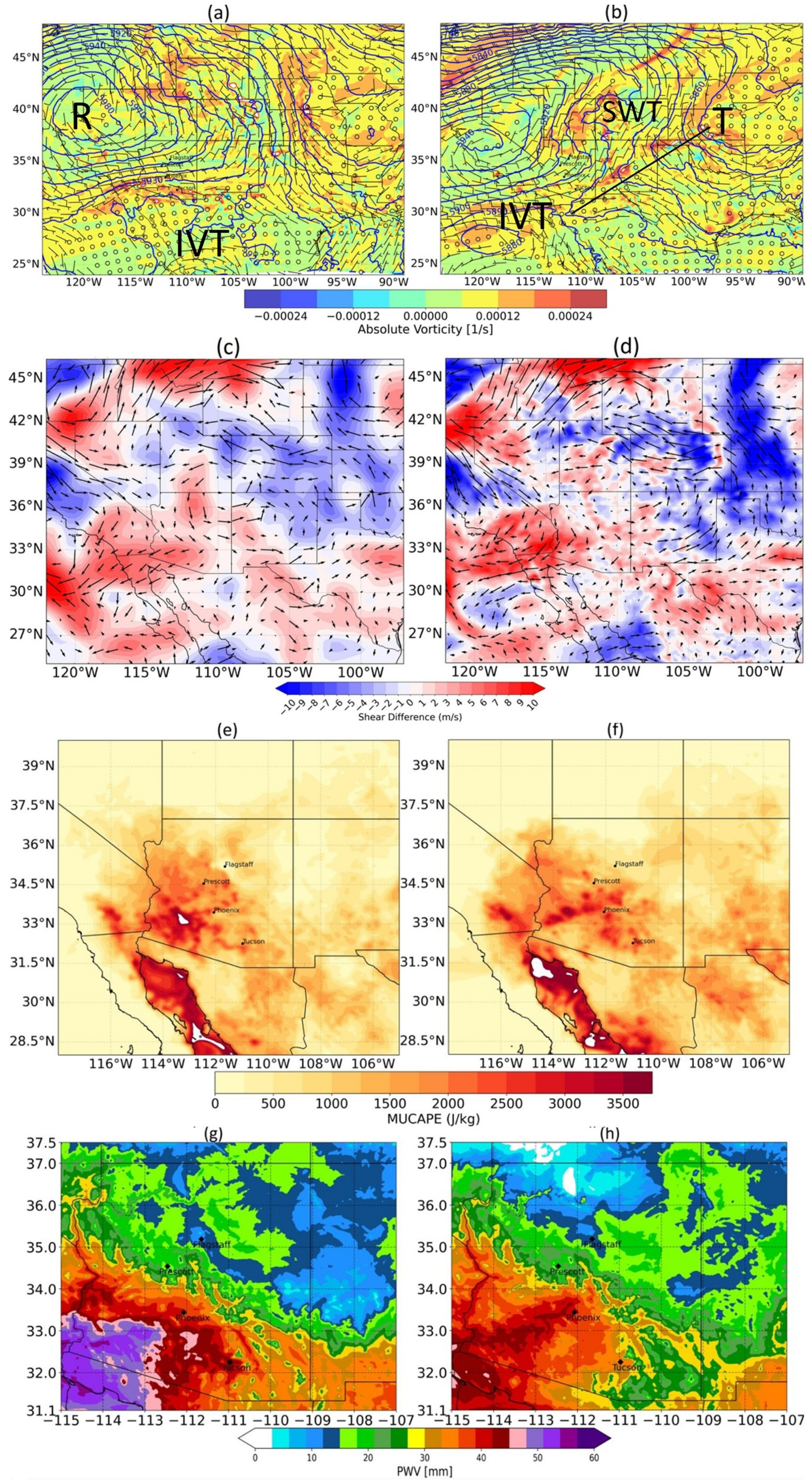

Figure 9. HRRR analyses of 500 hPa heights, absolute vorticity, and winds for (a) Case 01 at UTC 18 15 August 2021 and (b) Case 02 at UTC 18 16 August 2021. Identified synoptic-scale features: T = trough, R = ridge, IVT = inverted trough, and SWT = shortwave trough. (c) GFS and (d) HRRR 850–400 hPa bulk shear difference: case 02 at UTC 18 August 16—case 01 at UTC 18 August 15. HRRR MUCAPE fields (J/kg) are shown in panels (e) Case 01 at UTC 18 and (f) at Case 02 at UTC 18, and PWV fields (mm) in panels (g) Case 01 at UTC 18 and (h) Case 02 at UTC 18.

OSSE experiments, although none of the uniform array experiments performed well with the convection to the west of Flagstaff.

The OSSEs for both the 50-km array and the 100-km array exhibit significant overestimation of precipitation in southern Arizona, while the OSSE with a 200-km array illustrates a tendency toward underestimating precipitation levels in those areas even with higher precipitation thresholds (shown in bias plots of Figure 11b).

Fractional skill score (FSS) results depicted in Figure 11b indicate that the 50-km and 100-km arrays performed better for both the 5 and 10 mm precipitation thresholds during the initial hours than other OSSEs and maintained superiority to NODA until 9 hr into the forecasts. However, the 200-km array OSSE and superobbed stations OSSE both start performing better from 3 to 9 hr into the forecasts. All of the OSSEs exceeded the performance of the NODA control run at all times leading up to 9 hr, thus demonstrating the value added of the PWV data assimilation. Of note is that a significant underestimation of accumulated precipitation (negative Bias) is observed throughout the 12-hr forecasts by all the OSSE experiments, NODA, and NR with the exception of the denser arrays (50-km and 100-km) during the first 3 hr of accumulation, when bias was negligible.

In summary, the overall performance of all OSSE experiments indicates a consistent underestimation of precipitation in northern Arizona, but quite good forecasts overall of the mesoscale distribution of convective precipitation over central and southern Arizona especially by the denser GPS arrays.

4.1.2. Case 02

As shown in Figure 12, during the data assimilation window (spanning from UTC 12 August 16 to UTC 18 August 16), we observe a pronounced reduction in negative (dry) bias across all the OSSE experiments when compared to the NODA, similar to what was seen in Case 01. This positive impact persists for the next 4 hr (22 UTC) into the free forecasts. Afterward, there is a significant decline in forecast accuracy, as illustrated in Figures 12b and 12c. Interestingly, RMSE increased after 4 hr, with a uniform array showing higher values than NODA, while superobbed remains low. Overall, Case 02 exhibited a less persistent impact of DA, lasting only up to 4 hr in comparison to Case 01, which demonstrated effects that persisted for up to 8 hr.

The 3-hr temporal evolution of accumulated precipitation from MRMS observations in Figure 13a shows that a large area of significant precipitation developed over southeastern Arizona and western New Mexico. The NR exhibits notable similarity to the MRMS data, capturing the primary precipitation patterns and features that characterize this event. However, unlike Case 01, the skill of the NODA in Case 02 leaves less room for improvement for the DA experiments due to the influence of synoptic scales.

In Figure 13a, while none of the OSSE experiments outperform the NODA, there is significant variability in how they capture the extensive area of precipitation over southeastern Arizona and southwestern New Mexico. The OSSE-superobbed station experiment displays a significant deficiency, showing very little precipitation in eastern Arizona while correctly maintaining substantial activity over southwestern New Mexico. The OSSE-50 km uniform array experiment reveals a precipitation pattern similar to that of the OSSE-superobbed run, while the OSSE-200 km run demonstrates a serious deficiency in accurately representing the precipitation across the entire state of Arizona. By contrast, the OSSE-100 km case performs rather well, arguably better than any of the other OSSE results in successfully capturing a significant quantity of 3-hr accumulated precipitation in initial forecast hours over eastern Arizona, though lacking in accuracy over central Arizona.

According to Figure 13b, the best performing DA experiments in terms of the temporal evolution of FSS and bias are the superobbed spacing at 3 hr and 100 km array spacing at 6 hr. In this case, the positive impact only extends out to 6 hr forecast in some of the experiments (e.g., 100 km) except for superobbed that maintains skills until 9 hr. The higher FSS observed from the 100-km array spacing at 6 hr is probably due to the overestimation of precipitation, but this overestimation resulted in low FSS and high biases at UTC 21 August 16 and UTC 03 August 17. Analysis of the 2-m surface temperature field (not shown) reveals stronger temperature gradients east of central Arizona than the NODA and NR data sets. These gradients could lead to the early development of convection, resulting in heavy precipitation, but the resultant stabilization led to a rapid diminution of convection thereafter. Additionally, a strong outflow boundary (Figure 13a. 03 for 100 km spacing) is observed early in the forecasts (e.g., at UTC 21 August 16), which could trigger new thunderstorm development toward southeastern Arizona and southwestern New Mexico.

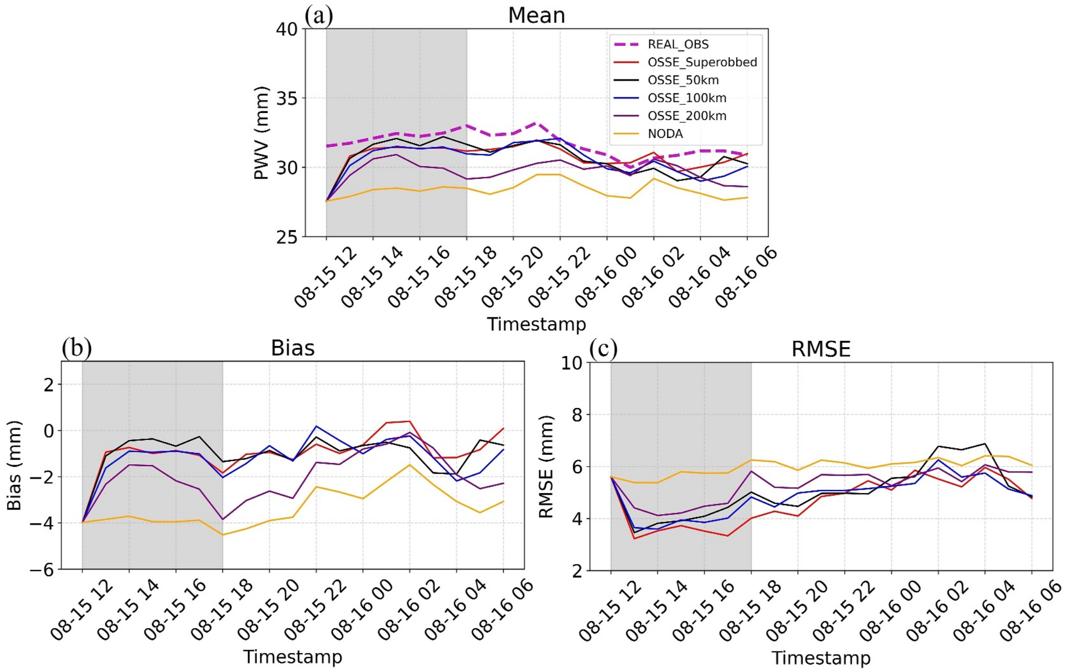

Figure 10. Case 01: Statistical comparison of real observations with different OSSE experiments from the superobbed (*red*), uniform arrays [50-km (black), 100-km (*blue*), and 200-km (*purple*)] and NODA simulations (*orange*). A network of 31 stations from the 2021 campaign from UTC 12 August 15 to UTC 12 August 16 are grouped by each timestamp, and statistical parameters are calculated. Shown are time series of (a) mean PWV averaged across all stations, (b) mean bias, and (c) RMSE. Real mean network values of PWV from GPS observation are depicted by the dashed line in (a). A gray shading indicates the period of DA cycles followed by 12-hr forecasts.

In summary, the improvement in initial moisture quickly diminishes compared to Case 01, failing to outperform NODA in the 12-hr forecasts. The various OSSE experiments show mixed responses regarding FSS and biases in this strongly forced case, making it challenging to identify the best DA experiments.

4.2. Understanding Moisture Variability

4.2.1. Spatial Moisture Pattern Analysis

A detailed evaluation of the PWV fields over 6-hourly forecast intervals is presented in Figure 14 for the NR, NODA, and OSSE forecasts in the first case. This analysis of difference fields aids in identifying areas of moisture improvement and reduction. At the end of the DA cycles at UTC 18 August 15, all OSSEs display a notable increase in moisture compared to the NODA run across nearly the entire modeling domain. However, there is a notable moisture gradient near the Arizona/New Mexico border for uniform arrays, which maybe due to artifacts from the lack of stations in New Mexico compared to superobbed stations. After 6 hr of the free forecast, the moisture begins to decrease in some areas relative to NODA, revealing pockets of dryness within the domain. By the end of the forecast valid at UTC 06 August 16, these dry pockets have expanded in areal coverage and intensity in all of the OSSEs. These comparisons indicate that the PWV in each of the DA experiments evolves slightly differently than that in the NoDA experiment with some regions ending up more moist and others drying out.

The differences in the PWV for experiments run for Case 02 are shown in Figure 15. Clearly, the OSSE 100 km experiment, and to a lesser extent, the 50 km experiment, exhibits a remarkable level of early moistening in comparison to the NODA experiment, more so than any of the other OSSE experiments conducted, which is similar to what was seen in Case 01. Also similar is that this pronounced moistening was nearly entirely eliminated by the end of the cycle, by which time impressive drying has replaced the moistening over a considerable portion of the domain, by which time the cold pools from the early convection have grown. It is interesting to observe that the implications of this moistening relate to what was observed in Figure 13a. 03 100-km array at

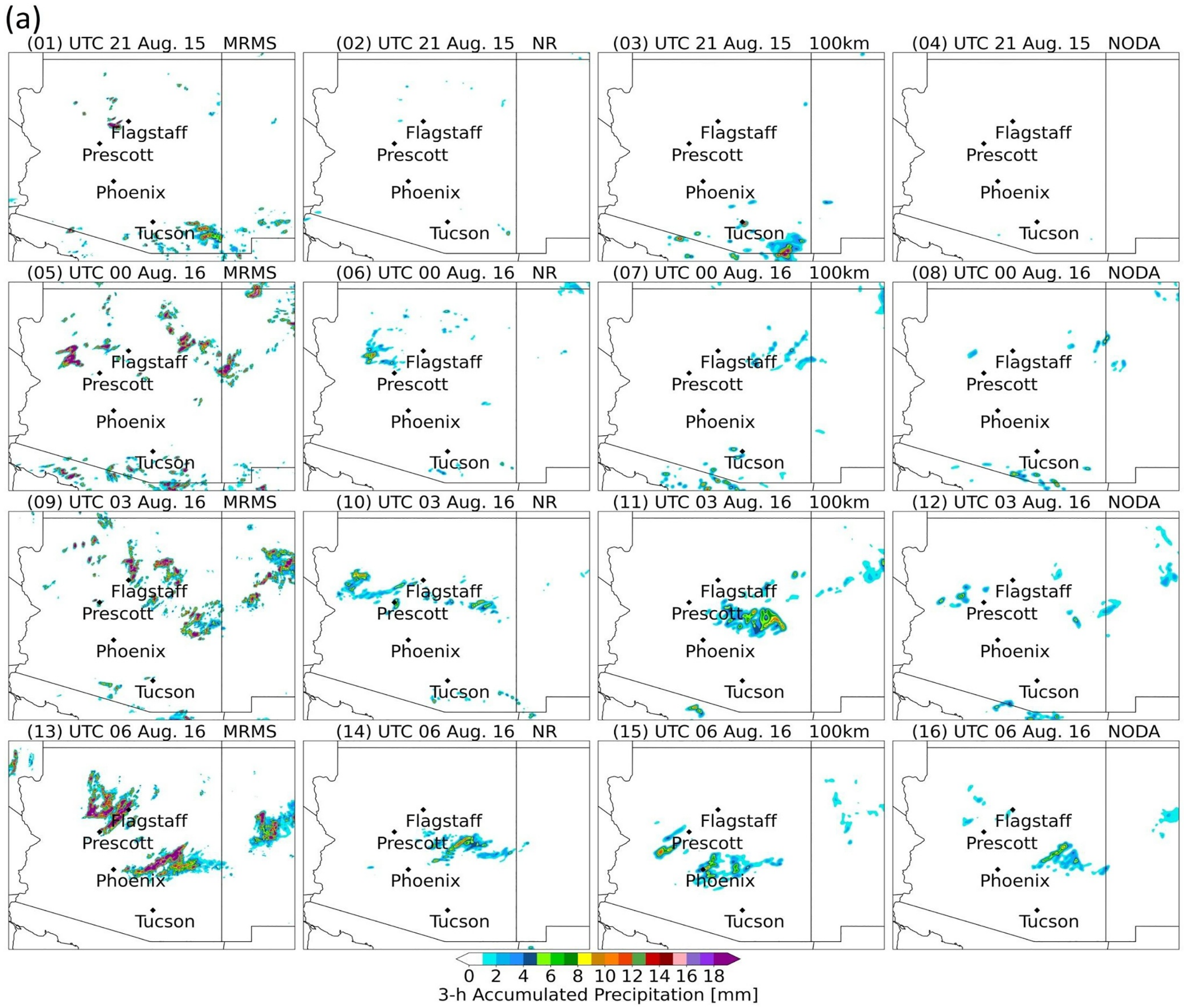

Figure 11. (a): 3-hr forecast accumulated precipitation over the Arizona region during NAM Case 01 precipitation from UTC 18 August 15 to UTC 06 August 16. Each row represents every 3 hr of accumulated precipitation, such as figure 01–04: UTC 21 August 15, 05–08: UTC 00 August 16, 09–12: UTC 03 August 16, and 13–16: UTC 06 August 16. Each column represents specific forecasts from the MRMS (figures: 01, 05, 09, and 13), in comparison to forecasts from NR (figures: 02, 06, 10, and 14), 100-km (figures: 03, 07, 11, 15), and NODA (figures: 04, 08, 12, 16). (b): Statistical evaluations of 3-hr accumulated precipitation for uniform array stations: *50-km*, *100-km*, *200-km,* and *superobbed* stations, *NODA (line),* and *NR (line)* against MRMS observations. The plot shows two metrics in percentages (%), first, and the left *y*-axis depicts temporal evolution of FSS (solid colors) using sampling-scale equivalent to ∼100 km and multiple precipitation thresholds, including 5, 10, and 15 mm. A higher percentage means higher skill. Second, in a similar way, the right *y*-axis depicts bias (dashed colors) where positive values represent overestimation and negative values indicate underestimation.

UTC 21 August 16, that is, substantial precipitation occurred in the vicinity where the most significant moistening took place at the end of assimilation cycles.

When comparing the differences in PWV fields for the OSSEs shown in Figures 14 and 15, a significant dryness (greater than −8 mm) appears in northern-central Arizona, particularly over major urban centers such as Phoenix and Tucson in Case 02 compared to Case 01 after 6 hr (valid at UTC 00 August 17) of the forecast, effectively reversing the initial heavy moistening. However, the OSSE 100-km array shows slightly less dryness than other OSSEs in these regions and maintains a strong moistening in southeastern Arizona bordering New Mexico that can be reflected on the temporal evolution of 3-hr accumulated precipitation map on Figure 13a 100-km array.

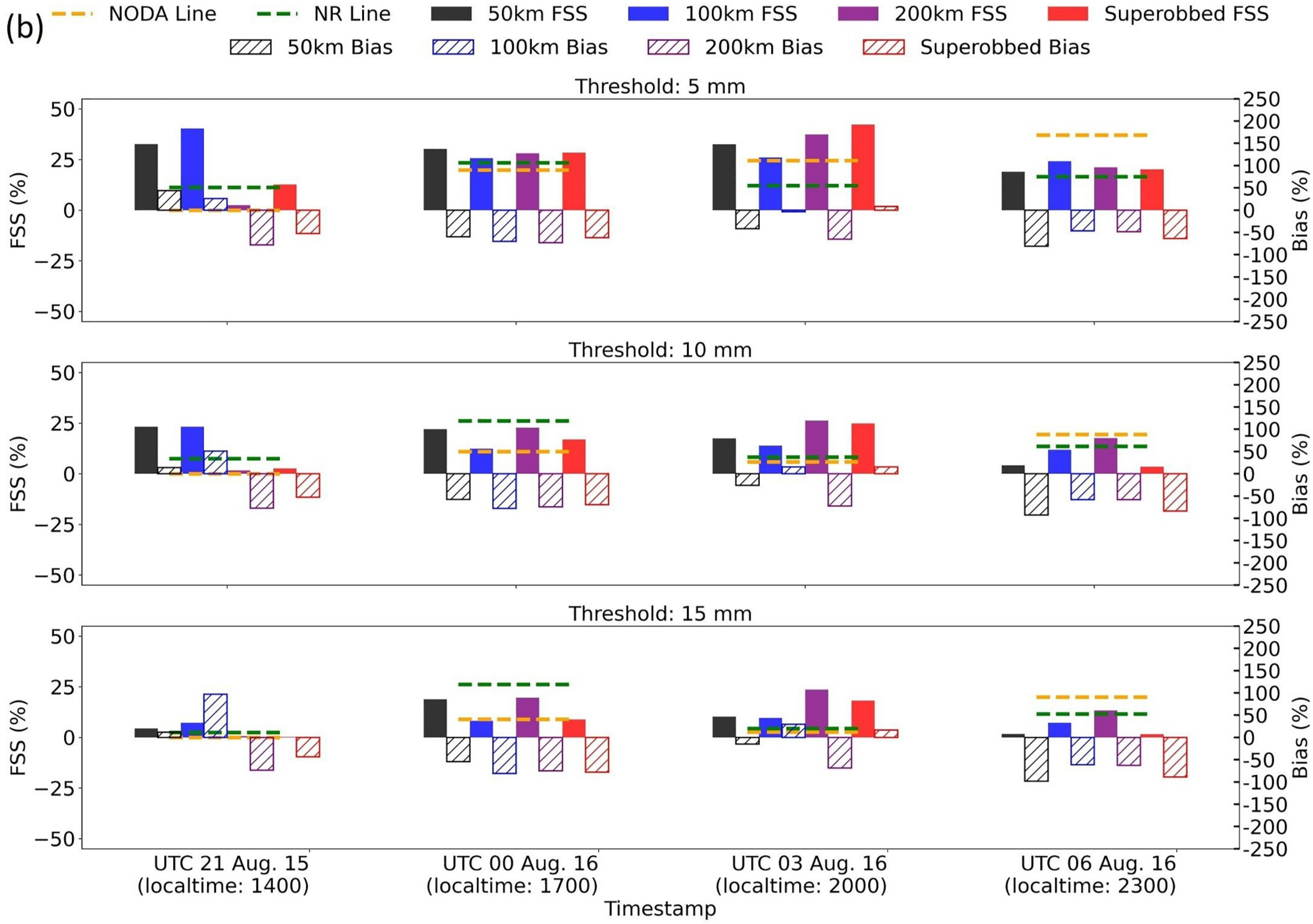

Figure 11. (Continued)

Moreover, in Case 02, a notable increase in moisture develops with time along the southern border of Arizona and Mexico, extending into the central and eastern parts of Arizona, with a pronounced drying pattern over central Arizona. This is the opposite of what is seen in Case 01, which hints to the movement of moisture to the southeast of Arizona as we proceed from Case 01 to Case 02. To understand this phenomenon, we investigated moisture flux convergence (MFC) for the OSSE experiments, from the initial state (UTC 18 August 16) to 6 hr into the forecast (UTC 17).

As shown in Figures 16c and 16d, all OSSE experiments for Case 02 show overall net divergence of moisture (MFC < 0) near central Arizona (blue circles), meaning more moisture is being removed from this region than is being supplied. This result explains the dryness (>−8 mm) that occurred for all the OSSE experiments in the plots of Figure 15 on UTC 17 August 16 in central Arizona when compared with NODA. This phenomenon potentially occurred due to the presence of convective downdrafts or compensating subsidence, where drier air descends and suppresses cloud formation (Houze, 1993). In the NAM region, thunderstorms often lead to moistening due to strong moisture convergence, while surrounding areas experience drying from downdrafts and evaporative cooling effects (Adams & Comrie, 1997; Higgins et al., 1997). For example, Figures 16c and 16d clearly shows MFC > 0 over northwestern and southeastern parts of Arizona, indicating net moistening that results in rising air motion, condensation, and cloud formation, while contributing to subsidence of the dry air in central Arizona.

The 100 km array (Figure 16d) shows a positive MFC near precipitation hotspots, particularly east of Phoenix and Tucson. This phenomenon aligns with the data presented in Figures 15e.i and 15e.ii, indicating that as the forecast progresses, moist air converges in these areas, leading to increased moisture and precipitation during the early hours, as illustrated in Figure 13a 100-km array temporal evolution.

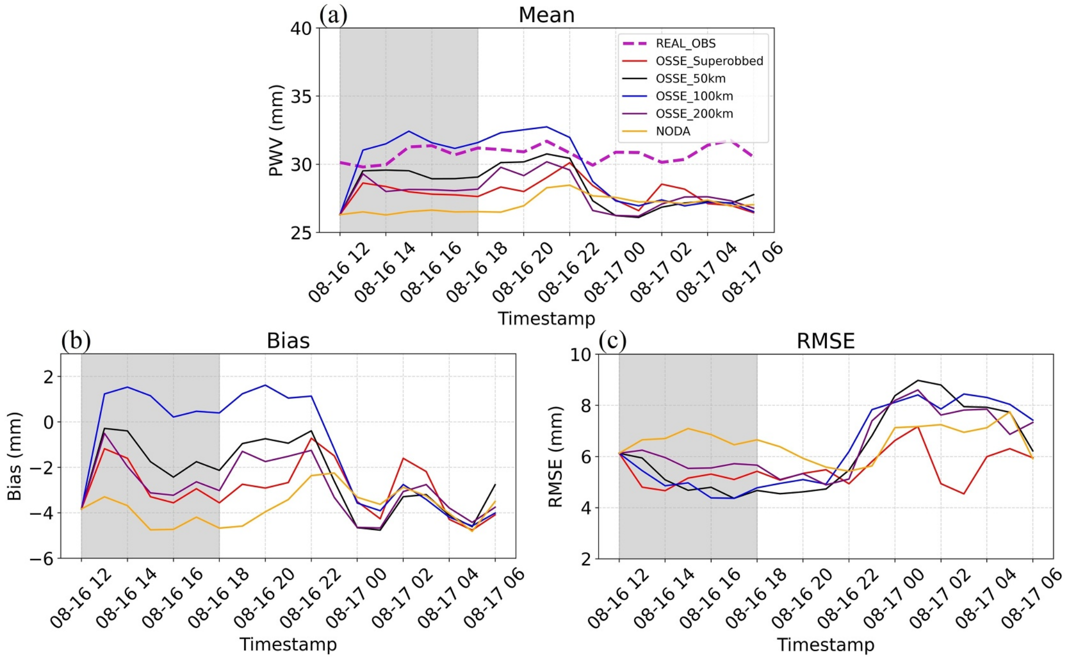

Figure 12. As in Figure 10, except for Case 02, for the period of UTC 12 August 16 to UTC 06 August 17.

Figures 16a and 16b highlight variability in MFC east of Flagstaff, suggesting the presence of both updrafts and downdrafts associated with frontal convection. Additionally, strong easterly winds (Figure 9b) have been observed entering southern Arizona from Mexico at the initial measurement on UTC 18 August 16. After 6 hr of forecasting at UTC 17, these easterlies intensify and move eastward in all OSSE experiments except for the 100-km array. This phenomenon may account for the shift in precipitation hotspots toward the southeast of Arizona and into New Mexico over the course of the forecast period. It could also explain the contrasting moisture advection to the southeast for Case 02, versus the drying effects observed in southern Arizona for Case 01.

4.2.2. Impact of Elevation on Moisture Forecasts

Figure 17 illustrates the variation of PWV biases over a 12-hr forecast period for both OSSE cases. It compares all the OSSE experiments against real PWV observations across different elevation bands: 0–500 m, 500–1,000 m, 1,000–1,500 m, 1,500–2,000 m, and above 2,000 m. Generally, the biases range from approximately −10 to +10 mm, with distinct fluctuations observed in each elevation band throughout the forecast period. Case 01 shows smaller overall variability, while Case 02 exhibits larger fluctuations across all OSSEs in all elevation bands. For both cases, the biases at observation sites located in the highest elevation band (e.g., 2000+ m) tend to be close to zero, with the lines closely following each other. In contrast, the observations obtained at sites within the lowest elevation band of 0–500 m indicate an overall trend of positive bias for all OSSEs. Overall, there is a gradual trend of negative biases as the elevation band increases from 0–500 m to 1,500–2,000 m for the OSSEs of both cases, until positive biases reappear for the elevation bands above 2,000 m. These bias profiles indicate the challenges associated with assimilating a vertically integrated quantity in cases when bias varies with height.

In summary, the elevation-dependent PWV biases could result from a combination of inadequate surface moisture representation at low elevations, misrepresented orographic processes at midelevations, and upper-level physics errors at high elevations. Addressing these issues requires improvements in terrain resolution, better parameterizations of the surface and microphysics, enhanced assimilation of high-elevation data, and tailored adjustments to convective and boundary layer schemes according to specific meteorological scenarios.

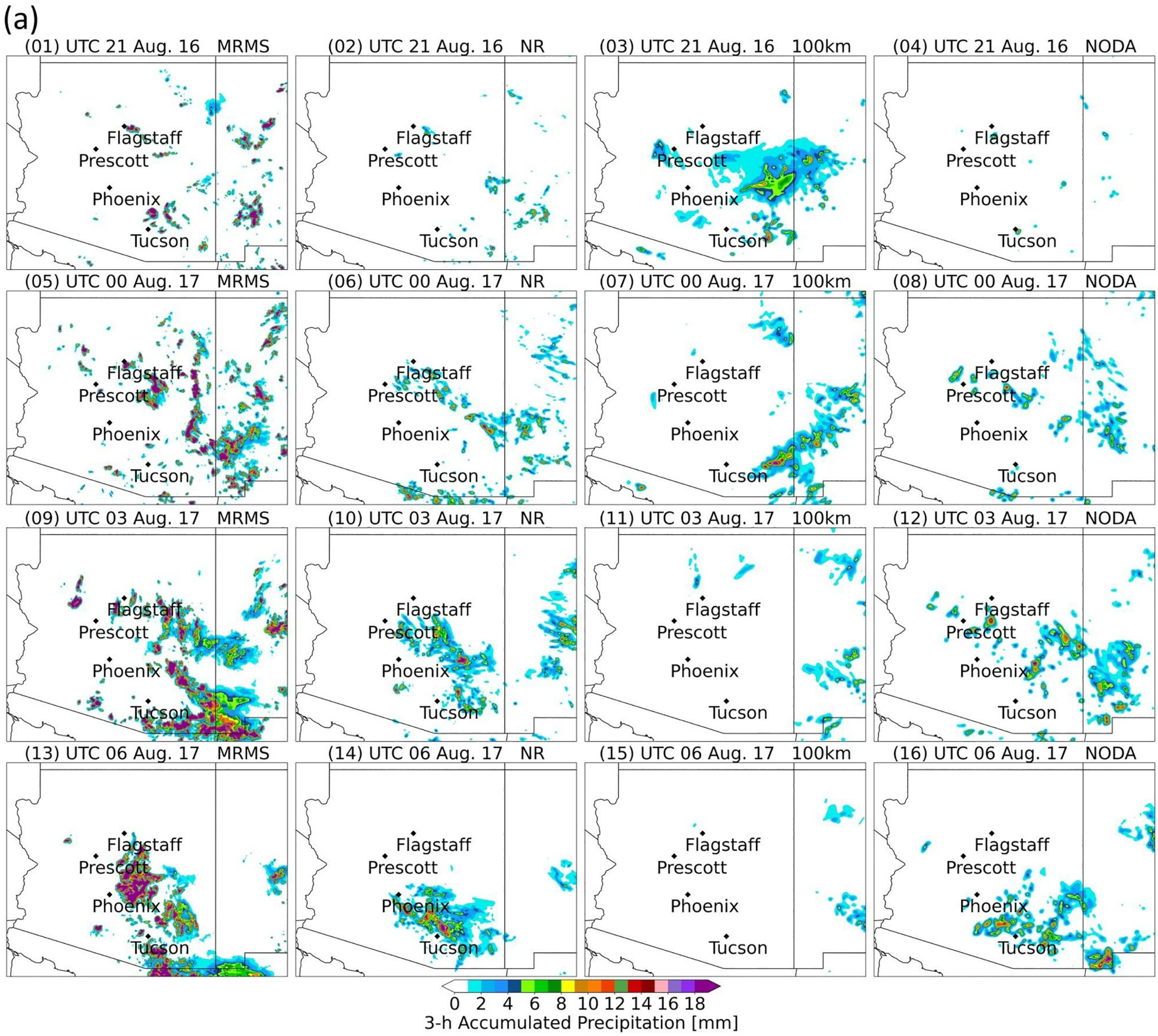

Figure 13. (a): As in Figure 11, except for Case 02, depicting 3-hr accumulated precipitation from UTC 18 August 16 to UTC 06 August 17. (b): As in Figure 11, except for Case 02, depicting FSS and bias for 3-hr accumulated precipitation from UTC 18 August 16 to UTC 06 August 17.

4.3. Convective Dynamics in OSSE Experiments

Figure 18 highlights significant effects of the observation network on MUCAPE and CIN across southern Arizona with enhanced moisture presence. MUCAPE is highest near Tucson and Phoenix, while areas in Flagstaff and northern Arizona show elevated CIN. The 50-km grid reveals spatially variable MUCAPE changes, whereas the 200-km grid shows broader modifications in both MUCAPE and CIN, indicating substantial shifts in convective suppression patterns. In this OSSE, the 50-km array serves as a baseline to evaluate the loss of predictive skill when using a 200-km array. The significant differences observed in the 200-km array highlight the crucial role of resolution in capturing localized convective processes, even in an idealized setup. Although the 50-km array still contains errors due to limitations in its physics, its finer grid allows it to better replicate the CAPE/CIN structure, especially in areas of complex terrain like Arizona. This emphasizes the importance of high-resolution models for

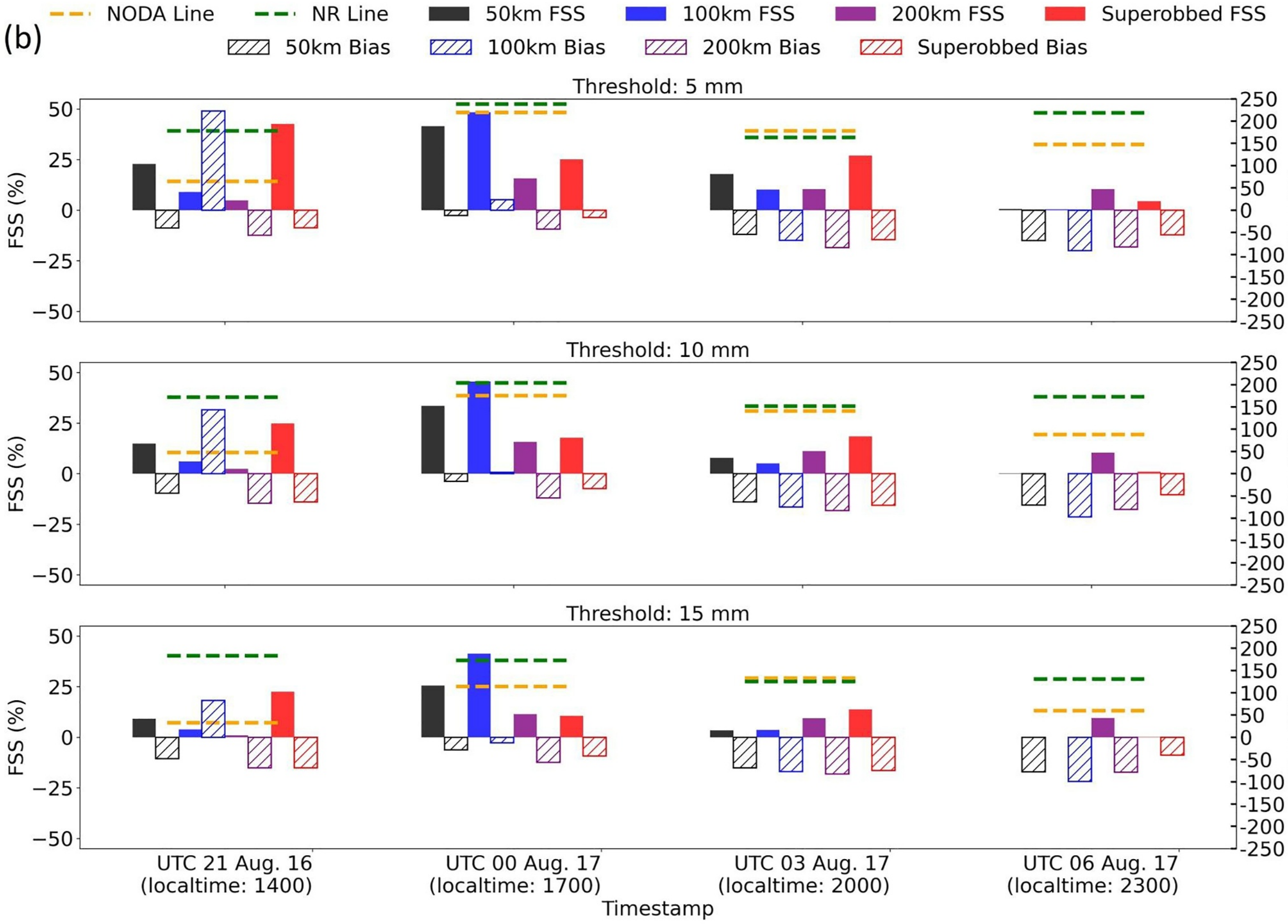

Figure 13. (Continued)

accurate storm prediction, as they can resolve the fine-scale dynamics forced by terrain variations that drive convection dynamics that are often missed by coarser arrays.

5. Conclusion

This study examines the integration of PWV data in an OSSE framework to improve forecasts of monsoonal MCSs in Arizona. Using a 40-member ensemble with a 1.8-km resolution WRF model, synthetic PWV data are created from an NR and bias corrected with real observations from an exclusive GPS-Met campaign in 2021 during NAM season. The research tests various GPS station spacing configurations to optimize forecast accuracy for MCS dynamics using a model configured similar to operational forecast.

The assimilation of synthetic bias-corrected PWV data led to a significant improvement in PWV forecast accuracy, resulting in reduced mean errors and smaller dry bias during the 6-hr assimilation window. During the forecast period, there was an initial reduction in errors, however, forecast accuracy decreased over time, with forecast skill returning to NODA levels by the 4–8-hr mark (Risanto et al., 2021). This would add value for extreme weather situations, in terms of improved advance warning capability.

In terms of precipitation forecasting, the impacts of PWV OSSEs varied across different observation network configurations. Overall, a 100-km array network is found to be the most effective for capturing key atmospheric

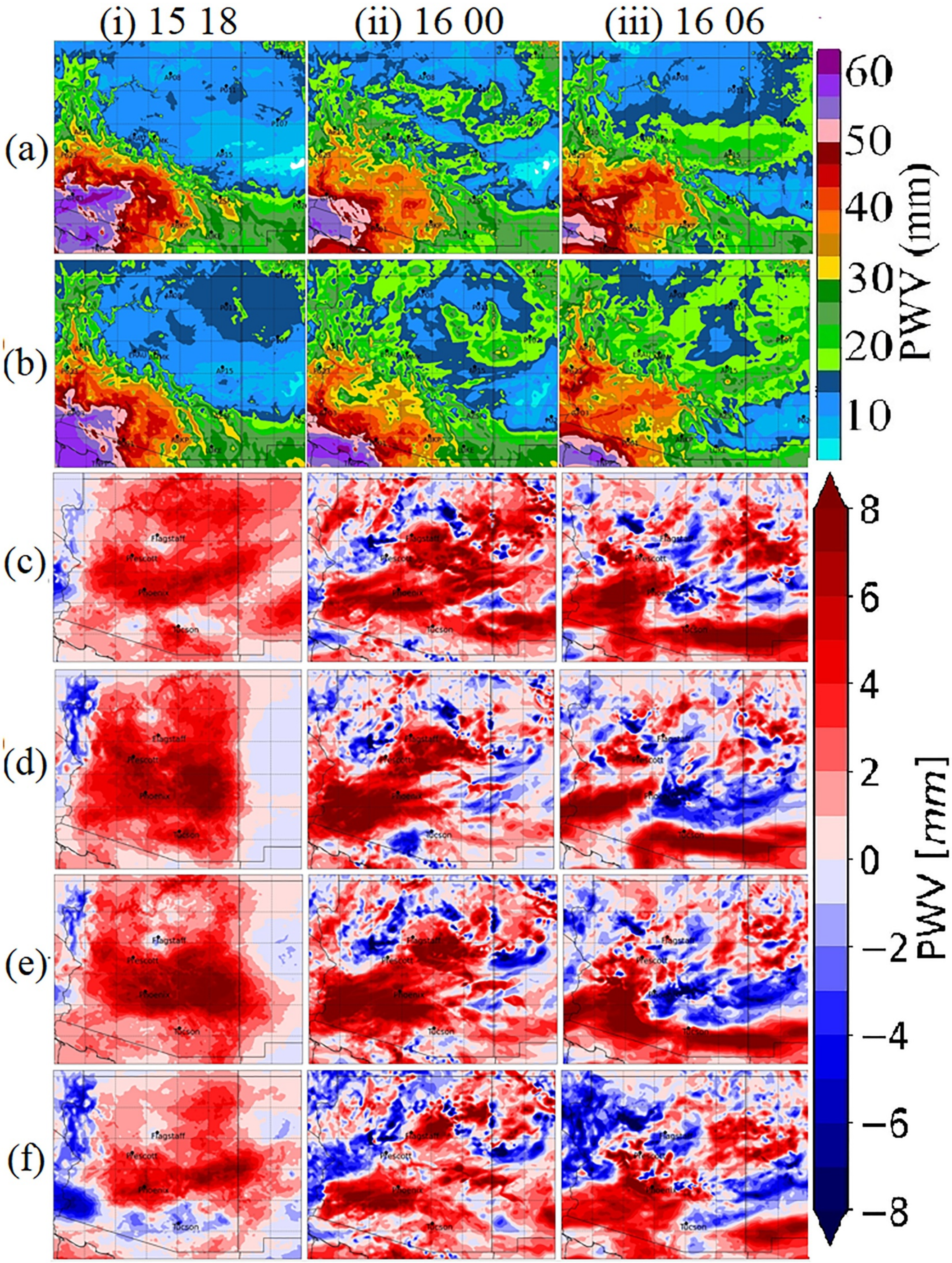

Figure 14. Case 01: rows from top to bottom: (a) NR, (b) NODA, (c) difference (OSSE superobbed-NODA), (d) difference (OSSE 50 km-NODA), (e) difference (OSSE 100 km-NODA), and (f) difference (OSSE 200 km-NODA). Columns represent PWV fields every 6 hr (i) initial state at UTC 18 August 15 and forecasts at (ii) UTC 00 August 16 and (iii) UTC 06 August 16.

processes in Arizona and New Mexico, balancing information distribution and model resolution for the synoptically weakly forced case; however, there is a mixed response across experiments for more synoptically forced case. The value added by assimilating local PWV observations is conditioned by the strength of synoptic forcing. In weakly forced environments, where convection is more sensitive to mesoscale thermodynamic details, PWV DA exerts a larger positive impact, which lasted up to 9 hr whereas 6 hr for strongly forced case. The availability

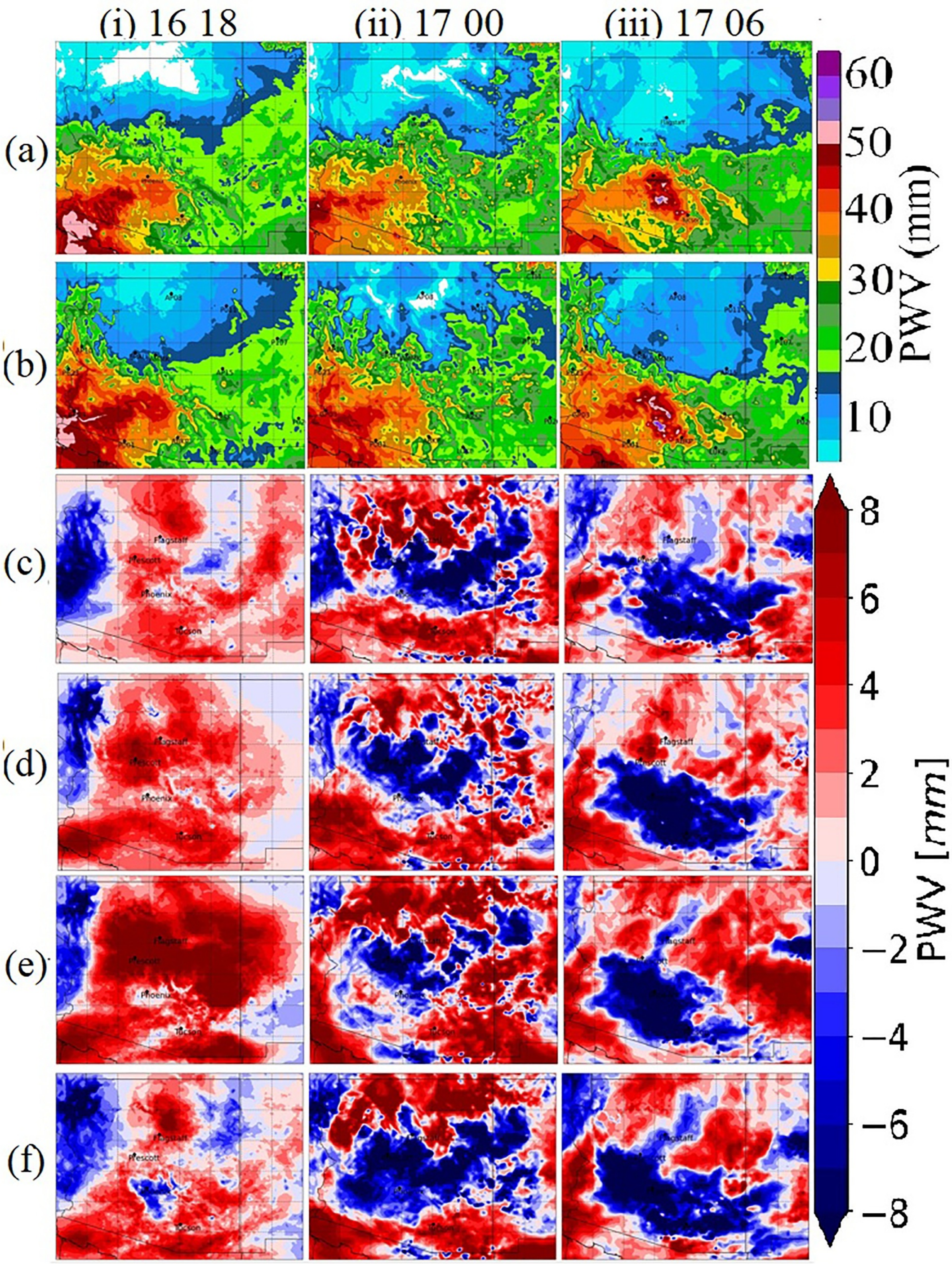

Figure 15. As in Figure 14, except for Case 02.

of moisture affects convection initiation, intensity, and organization, with topography playing a significant role in moisture transport and distribution. Mountainous areas like the Mogollon Rim enhance precipitation through orographic lifting, while lower-elevation deserts depend on synoptic and mesoscale moisture surges.

A denser 50-km observation network does not necessarily perform better than a sparser 100-km network. This can be correlated to observation errors; these errors can degrade analysis quality. Hunt et al. (2007) found that higher-

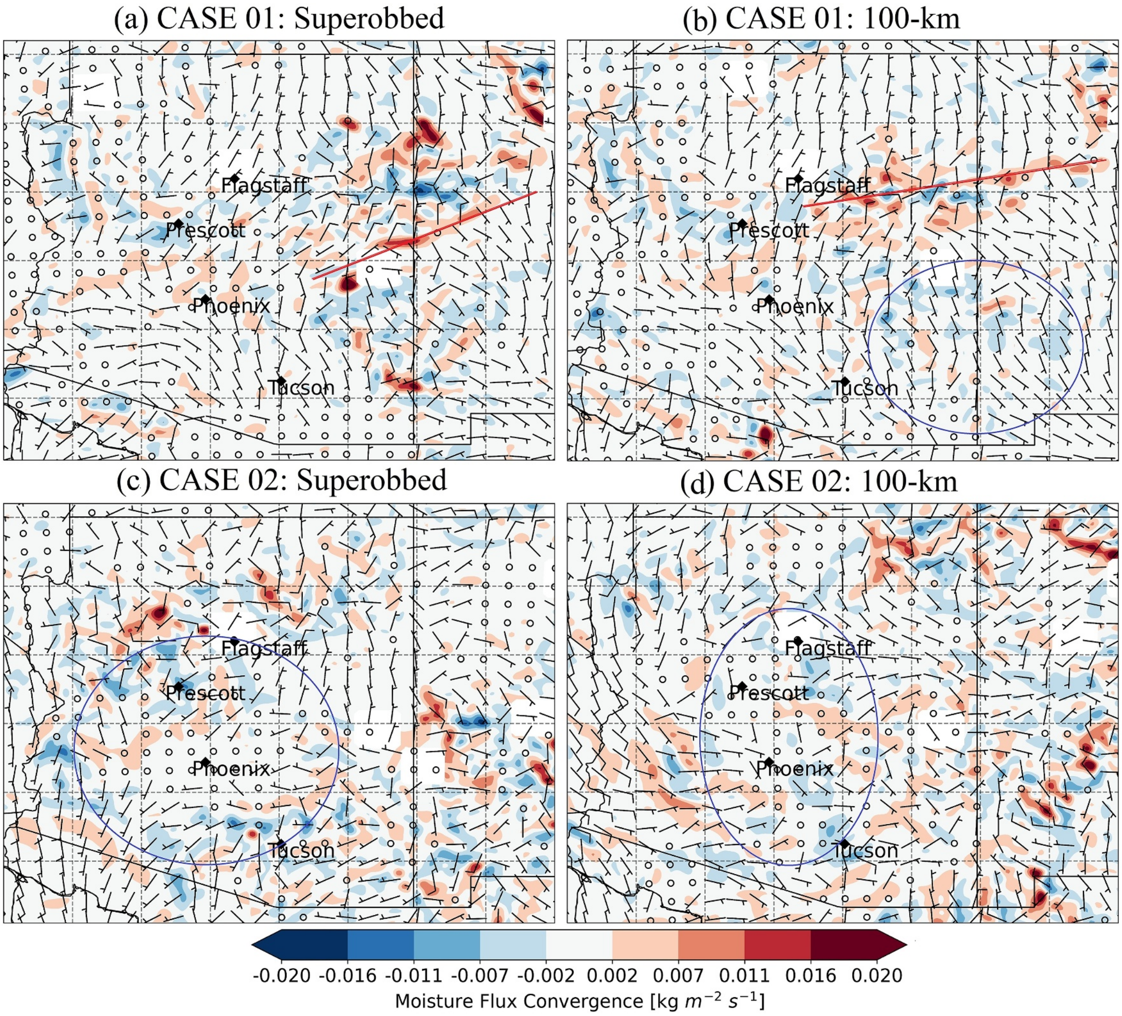

Figure 16. Moisture flux convergence (kg m^{-2} s^{-1}) superimposed with 10-m wind barbs at 6 hr into forecast from the initialization calculated at 700 hPa level applying low-pass Gaussian filter (a) CASE 01: superobbed on UTC 16 August 15, (b) CASE 01: 100 km array on UTC 16 August 15, (c) Case 02: superobbed on UTC 17 August 16, and (d) Case 02: 100 km array on UTC 17 August 16. Eclipses represent the regions where moisture is driven out. Red lines represent fronts.

density networks can amplify the effects of these errors if localization parameters are misconfigured. These errors can spread through the model, distorting forecasts and offsetting the potential advantages of higher observation density. Bannister (2008) highlighted that very dense coverage without proper error filtering can reduce forecast skill. This implies that the 100-km network may balance observation coverage and error management more effectively than the 50 km network, which may introduce stronger error correlations without sufficient correction.

Topography also influenced moisture forecasts, leading to varying biases across different elevation bands. Lower elevations (0–500 m) exhibited positive biases due to overestimated surface moisture, midelevations (500–2000 m) showed negative biases stemming from misrepresented orographic processes, and high elevations (above 2,000 m) had systematic positive biases likely due to overestimation of upper-level moisture. These biases are a result of limitations in model physics, challenges associated with vertically spreading information contained in a vertically integrated quantity in DA systems, and constraints in terrain resolution. Therefore, improving terrain representation, microphysics, and data assimilation techniques is essential for enhancing moisture forecast accuracy. Finally, DA experiments revealed significant effects on MUCAPE and CIN between 50 and 200 km arrays, highlighting the importance of high-resolution models for accurate storm predictions due to the loss of detail in coarser grids.

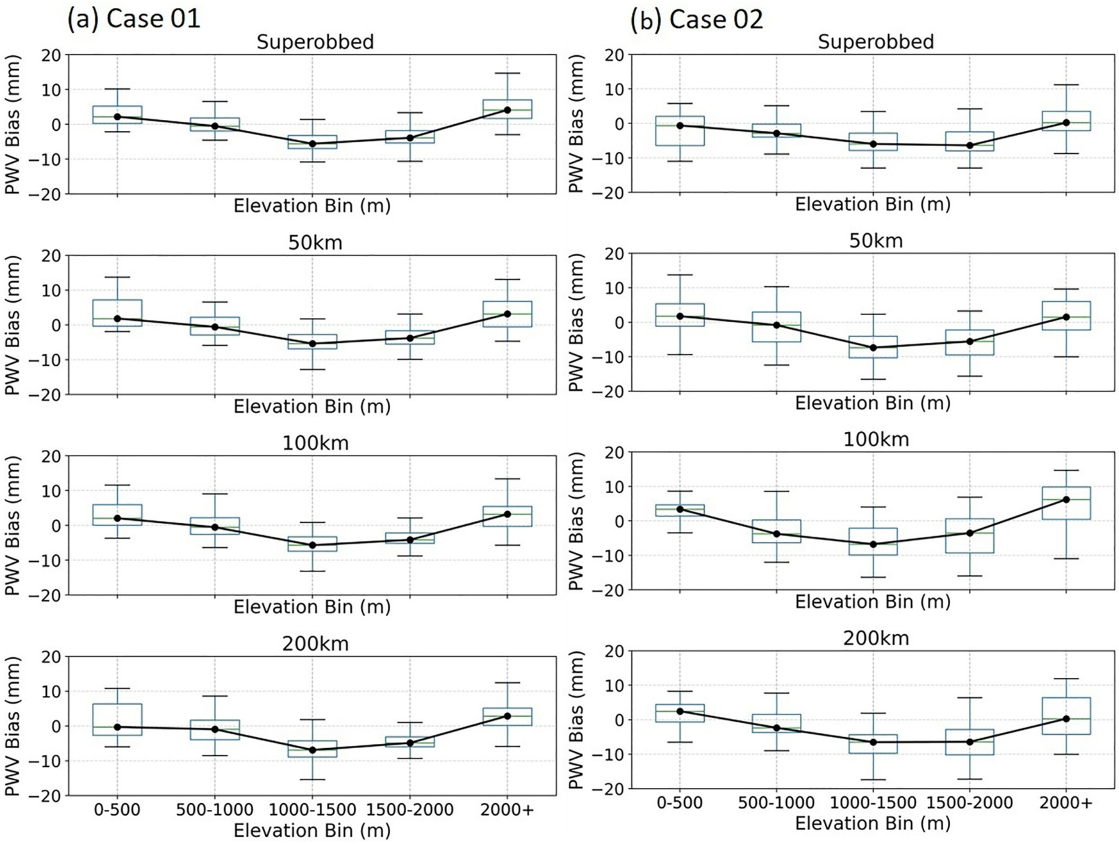

Figure 17. The boxplot shows the temporal-averaged PWV bias of 12 hr forecasts compared against real PWV from 2021 campaign for both OSSE cases aggregated by GPS-met site elevation bins: 0–500 m (4 stations), 500–1,000 m (4 stations), 1,000–1,500 m (6 stations), 1,500–2,000 m (8 stations), and 2,000+ m (5 stations). OSSE experiments: superobbed, 50-km, 100-km, and 200-km arrays for (a) Case 01 and (b) Case 02.

In future studies, integrating moisture profiles would add significant value to this research and may help mitigate some of the elevation biases since it is a limitation of the nature of vertically integrated measurements like GPS PWV. For instance, combining GPS PWV and MicroPulse DIAL (MPD) absolute humidity (AH) could enhance elevation-dependent biases by incorporating the distribution of moisture through atmospheric column. Moreover, analyzing many more weakly and strongly forced cases throughout an entire NAM season could add value to decision-making. Finally, due to the complex chaotic nature and processes that occur during the upscale growth of convective systems in regions of complex terrain, further studies using ensembles are needed to fully evaluate the impact of observing systems on NAM storm predictions.

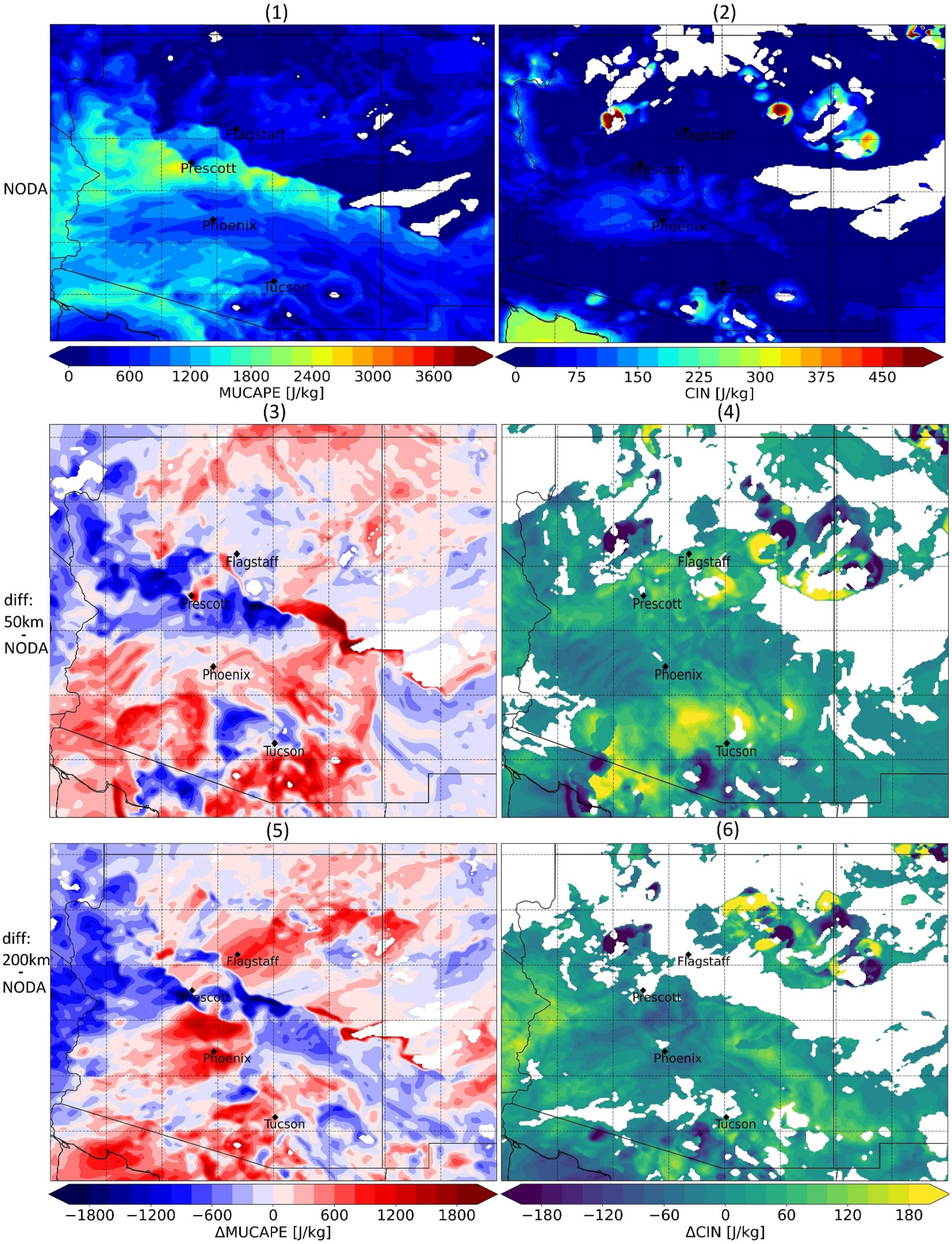

Figure 18. Case 01 MUCAPE and CIN through 6 hr of forecast at UTC 16, August 15. (1) MUCAPE (J/kg) NODA, (2) CIN (J/kg) NODA, (3) MUCAPE difference (50 km-NODA), (4) CIN difference (50 km-NODA), (5) MUCAPE difference (200 km-NODA), and (6) CIN difference (200 km-NODA).

Data Availability Statement

All relevant data and scripts used in this manuscript are available on the University of Arizona Research Data Repository at Shohan (2025).

Acknowledgments

This research was funded in part by NSF Award Nos. 2308409 and 2308410 and by NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977. This work was also funded in part by the NASA University Leadership Initiative (Award 80NSSC20M0162). We further acknowledge high-performance computing support from Derecho (https://doi.org/10.5065/qx9a-pg09) provided by the NSF National Center for Atmospheric Research (NCAR), sponsored by the National Science Foundation. We are grateful to Ronny Schroeder at Embry-Riddle University for help with the GPS-Met campaign. Finally, we thank the three reviewers who helped improve this manuscript.

References

Adams, D. K., & Comrie, A. C. (1997). The North American Monsoon. *Bulletin America Meteorology Social*, *78*(10), 2197–2214. https://doi.org/10.1175/1520-0477(1997)078<2197:TNAM>2.0.CO;2

Adams, D. K., Fernandes, R. M. S., & Maia, J. M. F. (2011). GNSS precipitable water vapor from an Amazonian rain forest flux tower. *Journal of Atmospheric and Oceanic Technology*, *28*(10), 1192–1198. https://doi.org/10.1175/JTECH-D-11-00082.1

Adams, D. K., & Souza, E. P. (2009). CAPE and convective events in the Southwest during the North American monsoon. *Monthly Weather Review*, *137*(1), 83–98. https://doi.org/10.1175/2008mwr2502.1

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., & Arellano, A. (2009). The data assimilation research testbed: A community facility. *Bulletin of the American Meteorological Society*, *90*(9), 1283–1296. https://doi.org/10.1175/2009BAMS2618.1

Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, *129*(12), 2884–2903. https://doi.org/10.1175/1520-0493(2001)129,2884:AEAKFF.2.0.CO;2

Anderson, J. L. (2012). Localization and sampling error correction in ensemble Kalman filter data assimilation. *Monthly Weather Review*, *140*(7), 2359–2371. https://doi.org/10.1175/MWR-D-11-00013.1

Bannister, R. N. (2008). A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, *134*(637), 1951–1970. https://doi.org/10.1002/qj.339

Benjamin, S. G., Jamison, B. D., Moninger, W. R., Sahm, S. R., Schwartz, B. E., & Schlatter, T. W. (2010). Relative short-range forecast impact from aircraft, profiler, radiosonde, VAD, GPS-PW, METAR, and Mesonet observations via the RUC hourly assimilation cycle. *Monthly Weather Review*, *138*(4), 1319–1343. https://doi.org/10.1175/2009MWR3097.1

Benjamin, S. L., Schwartz, B. E., Szoke, E. J., & Koch, S. E. (2004). The value of wind profiler data in U.S. weather forecasting. *Bulletin America Meteorology Social*, *85*(12), 1871–1886. https://doi.org/10.1175/BAMS-85-12-1871

Bevis, M., Businger, S., Herring, T. A., Rocken, C., Anthes, R. A., & Ware, R. H. (1992). GPS meteorology: Remote sensing of atmospheric water vapor using the Global Positioning System. *Journal of Geophysical Research*, *97*(D14), 15787–15801. https://doi.org/10.1029/92JD01517

Bieda, S. W., III., Castro, C. L., Mullen, S. L., Comrie, A. C., & Pytlak, E. (2009). The relationship of transient upper-level troughs to variability of the North American monsoon system. *Journal of Climate*, *22*(15), 4213–4227. https://doi.org/10.1175/2009jcli2487.1

Boukabara, S.-A., Ide, K., Zhou, Y., Shahroudi, N., Hoffman, R. N., Atlas, R., et al. (2018). Community global observing system simulation experiment (OSSE) package (CGOP): Assessment and validation of the OSSE system using an OSSE–OSE intercomparison of summary assessment metrics. *Journal of Atmospheric and Oceanic Technology*, *35*(10), 2061–2079. https://doi.org/10.1175/JTECH-D-18-0061.1

Caumont, O., Cimini, D., Löhnert, U., Alados-Arboledas, L., Bleisch, R., Buffa, F., et al. (2016). Assimilation of humidity and temperature observations retrieved from ground-based microwave radiometers into a convective-scale NWP model. *Quarterly Journal of the Royal Meteorological Society*, *142*, 2692–2704. https://doi.org/10.1002/qj.2860

Data Assimilation Research Testbed (Version 11.1.0-alpha). (2023). Data assimilation research testbed (Version 11.1.0-alpha) [Software]. *Boulder, Colorado: UCAR/NSF NCAR/CISL/DAReS*. https://doi.org/10.5065/D6WQ0202

Dominguez, F., Miguez-Macho, G., & Hu, H. (2016). WRF with water vapor Tracers: A study of moisture sources for the North American Monsoon. *Journal of Hydrometeorology*, *17*(7), 1915–1927. https://doi.org/10.1175/JHM-D-15-0221.1

Fast, J. D., Berg, L. K., Alexander, L., Bell, D., D'Ambro, E., Hubbe, J., et al. (2019). Overview of the HI-SCALE field campaign: A new perspective on shallow convective clouds. *Bulletin of the American Meteorological Society*, *100*(5), 821–840. https://doi.org/10.1175/BAMS-D-18-0030.1

Feng, X., Bosilovich, M., Houser, P., & Chern, J.-D. (2013). Impact of land surface conditions on 2004 North American monsoon in GCM experiments. *Journal of Geophysical Research: Atmospheres*, *118*(2), 293–305. https://doi.org/10.1029/2012JD018805

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757. https://doi.org/10.1002/qj.49712555417

Geer, A. J., & Bauer, P. (2010). Enhanced use of all-sky microwave observations sensitive to water vapour, cloud and precipitation. In *EUMETSAT/ECMWF Fellowship Programme Research Rep* (Vol. 20, p. 43). Retrieved from https://www.ecmwf.int/node/9506

Guerova, G., Jones, J., Douša, J., Dick, G., de Haan, S., Pottiaux, E., et al. (2016). Review of the state of the art and future prospects of the ground-based GNSS meteorology in Europe. *Atmospheric Measurement Techniques*, *9*(11), 5385–5406. https://doi.org/10.5194/amt-9-5385-2016

Hacker, J., Draper, C., & Madaus, L. (2018). Challenges and opportunities for data assimilation in mountainous environments. *Atmosphere*, *9*(4), 127. https://doi.org/10.3390/atmos9040127

Higgins, R. W., Yao, Y., & Wang, X. L. (1997). Influence of the North American monsoon system on the U.S. summer precipitation regime. *Journal of Climate*, *10*(10), 2600–2622. https://doi.org/10.1175/1520-0442(1997)010<2600:IOTNAM>2.0.CO;2

Higgins, W., & Gochis, D. (2006). Synthesis of results from the North American monsoon experiment (NAME) process study. *Journal of Climate*, *19*(9), 1601–1607. https://doi.org/10.1175/JCLI4081.1

Hoffman, R. N., & Atlas, R. (2016). Future observing system simulation experiments. *Bulletin America Meteorology Social*, *97*(9), 1601–1616. https://doi.org/10.1175/BAMS-D-15-00200.1

Holle, R. L., & Zhang, D. (2023). *Flashes of brilliance: The science and wonder of Arizona lightning* (p. 231). Springer International Publishing.

Houze, R. A., Jr. (1993). *Cloud dynamics* (p. 573). Academic Press.

Hu, G., Dance, S. L., Bannister, R. N., Chipilski, H. G., Guillet, O., Macpherson, B., et al. (2023). Progress, challenges, and future steps in data assimilation for convection-permitting numerical weather prediction: Report on the virtual meeting held on 10 and 12 November 2021. *Atmospheric Science Letters*, *24*(1), e1130. https://doi.org/10.1002/asl.1130

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*(1–2), 112–126. https://doi.org/10.1016/j.physd.2006.11.008

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*(D13), D13103. https://doi.org/10.1029/2008JD009944

Johnson, R. H., Ciesielski, P. E., McNoldy, B. D., Rogers, P. J., & Taft, R. K. (2007). Multiscale variability of the flow during the North American monsoon experiment. *Journal of Climate*, *20*(9), 1628–1648. https://doi.org/10.1175/jcli4087.1

Jones, T. A., Knopfmeier, K., Wheatley, D., Creager, G., Minnis, P., & Palikonda, R. (2016). Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast System. Part II: Combined radar and satellite data experiments. *Weather and Forecasting*, *31*(1), 297–327. https://doi.org/10.1175/WAF-D-15-0107.1

Jones, P. W. (1999). First- and second-order conservative remapping schemes for grids in spherical coordinates. *Monthly Weather Review*, *127*(9), 2204–2210. https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2

Kain, J. S. (2004). The Kain–Fritsch convective parameterization: An update. *Journal of Applied Meteorology*, *43*(1), 170–181. https://doi.org/10.1175/1520-0450(2004)043,0170:TKCPAU.2.0.CO;2

Kay, J., Weckwerth, T. M., Lee, W., Sun, J., & Romine, G. (2022). An OSSE study of the impact of micropulse differential absorption lidar (MPD) water vapor profiles on convective weather forecasting. *Monthly Weather Review*, *150*(10), 2787–2811. https://doi.org/10.1175/MWR-D-21-0284.1

Koch, S. E., desJardins, M., & Kocin, P. J. (1983). An interactive Barnes objective map analysis scheme for use with satellite and conventional data. *Journal of Applied Meteorology and Climatology*, *22*(9), 1487–1503. https://doi.org/10.1175/1520-0450(1983)022<1487:AIBOMA>2.0.CO;2

Leutbecher, M., & Palmer, T. N. (2008). Ensemble forecasting. *Journal of Computational Physics*, *227*(7), 3515–3539. https://doi.org/10.1016/j.jcp.2007.02.014

Madhulatha, A., Rajeevan, M., Mohan, T. S., & Thampi, S. B. (2020). Observational aspects of tropical mesoscale convective systems over southeast India. *Journal of Earth System Science*, *129*(1), 65. https://doi.org/10.1007/s12040-019-1300-9

Moker, J. M., Jr., Castro, C. L., Arellano, A. F., Serra, Y. L., & Adams, D. K. (2018). Convective-permitting hindcast simulations during the North American monsoon GPS transect experiment 2013: Establishing baseline model performance without data assimilation. *Journal of Applied Meteorology and Climatology*, *57*(8), 1683–1710. https://doi.org/10.1175/JAMC-D-17-0136.1

Nakamura, H., Koizumi, K., Mannoji, N., & Seko, H. (2004). Data assimilation of GPS precipitable water vapor to the JMA mesoscale numerical weather prediction model and its impact on rainfall forecasts. *Journal of the Meteorological Society of Japan*, *82*(1B), 441–452. https://doi.org/10.2151/jmsj.2004.441

Nakanishi, M., & Niino, H. (2009). Development of an improved turbulence closure model for the atmospheric boundary layer. *Journal of the Meteorological Society of Japan*, *87*(5), 895–912. https://doi.org/10.2151/jmsj.87.895

National Centers for Environmental Prediction (NCEP). (2018). *Multi-radar/multi-sensor (MRMS) precipitation data. Version 1.0*. UCAR/NCAR—Earth Observing Laboratory. https://doi.org/10.26023/R5QJ-VEPF-P00D

Navon, I. M. (2009). Data assimilation for numerical weather prediction: A review. In S. K. Park & L. Xu (Eds.), *Data assimilation for atmospheric, oceanic and hydrologic applications*. Springer. Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-71056-1_2

Newman, A., & Johnson, R. H. (2012). Mechanisms for precipitation enhancement in a north American monsoon upper-tropospheric trough. *Journal of the Atmospheric Sciences*, *69*(6), 1775–1792. https://doi.org/10.1175/JAS-D-11-0223.1

Niu, G.-Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, *116*(D12), D12109. https://doi.org/10.1029/2010JD015139

Pan, X., Lu, Y., Zhao, K., Huang, H., Wang, M., & Chen, H. (2021). Improving nowcasting of convective development by incorporating polarimetric radar variables into a deep-learning model. *Geophysical Research Letters*, *48*(21), e2021GL095302. https://doi.org/10.1029/2021GL095302

Privé, N. C., McGrath-Spangler, E. L., Carvalho, D., Karpowicz, B. M., & Moradi, I. (2023). Robustness of observing system simulation experiments. *Tellus A: Dynamic Meteorology and Oceanography*, *75*(1), 309–333. https://doi.org/10.16993/tellusa.3254

Raman, A., Arellano, A. F., Jr., & Brost, J. J. (2014). Revisiting haboobs in the southwestern United States: An observational case study of the 5 July 2011 Phoenix dust storm. *Atmospheric Environment*, *89*, 179–188. https://doi.org/10.1016/j.atmosenv.2014.02.026

Ray, A. J., Garfin, G. M., Wilder, M., Vásquez-León, M., Lenart, M., & Comrie, A. C. (2007). Applications of monsoon research: Opportunities to inform decision-making and reduce regional vulnerability. *Journal of Climate*, *20*(9), 1608–1627. https://doi.org/10.1175/JCLI4098.1

Risanto, C. B., Arellano, A. F., Koch, S., Castro, C. L., Shohan, S., Adams, D. K., et al. (2025). Improving monsoon precipitation forecast accuracy over complex terrain in Arizona with GPS-PWV data assimilation. *Monthly Weather Review*. (MWR-D-25-0101).

Risanto, C. B., Castro, C. L., Arellano, A. F., Jr., Moker, J. M., Jr., & Adams, D. K. (2021). The impact of assimilating GPS precipitable water vapor in convective-permitting WRF-ARW on North American monsoon precipitation forecasts over northwest Mexico. *Monthly Weather Review*, *149*, 3013–3035. https://doi.org/10.1175/MWR-D-20-0394.1

Rocken, C., Ware, R. H., Hove, T. V., Solheim, F., Johnson, C. A. J., Bevis, M., & Businger, S. (1993). Sensing atmospheric water vapor with the global positioning system. *Geophysical Research Letters*, *20*, 2631–2634. https://doi.org/10.1029/93GL02935

Romine, G. S., Berner, J., Fossell, K. R., Snyder, C., Anderson, J. L., & Weisman, M. L. (2014). Representing forecast error in a convection-permitting ensemble system. *Monthly Weather Review*, *142*, 4519–4541. https://doi.org/10.1175/MWR-D-14-00100.1

Schumacher, R. S., & Clark, A. J. (2014). Evaluation of ensemble configurations for the analysis and prediction of heavy-rain producing mesoscale convective systems. *Monthly Weather Review*, *142*(11), 4108–4138. https://doi.org/10.1175/MWR-D-13-00357.1

Serra, Y. L., Adams, D. K., Minjarez Sosa, C. M., Quintanar, A. I., Vazquez, L. M., & Figueroa Rodriguez, A. (2016). The North American monsoon GPS transect experiment 2013. *Bulletin America Meteorology Social*, *97*(11), 2103–2115. https://doi.org/10.1175/BAMS-D-14-00250.1

Sheppard, P. R., Comrie, A. C., Packin, G. D., Angersbach, K., & Hughes, M. K. (2002). The climate of the Southwest. *Climate Research*, *21*(3), 219–238. https://doi.org/10.3354/cr021219

Shohan, S. (2025). Dataset and software for "Impact of assimilating GPS precipitable water vapor on simulations of two North American monsoon convective events using observing system simulation experiments" [Dataset]. *University of Arizona Research Data Repository*. https://doi.org/10.25422/azu.data.29622020.v1

Skamarock, W., Klemp, J., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., et al. (2019). A description of the advanced research WRF model version 4.1. https://doi.org/10.5065/1dfh-6p97

Sobash, R. A., & Stensrud, D. J. (2015). Assimilating surface Mesonet observations with the EnKF to improve ensemble forecasts of convection initiation on 29 May 2012. *Monthly Weather Review*, *143*(9), 3700–3725. https://doi.org/10.1175/MWR-D-14-00126.1

Sun, Q., Vihma, T., Jonassen, M. O., & Zhang, Z. (2020). Impact of assimilation of radiosonde and UAV observations from the southern ocean in the polar WRF model. *Advances in Atmospheric Sciences*, *37*(5), 441–454. https://doi.org/10.1007/s00376-020-9213-8

Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review*, *136*(12), 5095–5115. https://doi.org/10.1175/2008MWR2387.1

Whitaker, J. S., & Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, *130*(7), 1913–1924. https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2

Willingham, K. M., Thompson, E. J., Howard, K. W., & Dempsey, C. L. (2011). Characteristics of Sonoran Desert microbursts. *Weather and Forecasting*, *26*(1), 94–108. https://doi.org/10.1175/2010WAF2222388.1

Zeng, X., Atlas, R., Birk, R. J., Carr, F. H., Carrier, M. J., Cucurull, L., et al. (2020). Use of observing system simulation experiments in the United States. *Bulletin of the American Meteorological Society*, *101*(8), E1427–E1438. https://doi.org/10.1175/BAMS-D-19-0155.1