A Probit-transformed Rank Histogram Filter for All-sky Infrared Radiance Assimilation

Haoxing Ju [1,2] Lili Lei [1,2]

[1]Key Laboratory of Mesoscale Severe Weather, Ministry of Education, Nanjing University, Nanjing, China [2]School of Atmospheric Sciences, Nanjing University, Nanjing, China

Introduction

Assimilation of all-sky satellite-measured radiance is challenging. For clear and cloudy regions usually have different background error statistics, and the exact location of clouds is still hard to predict, so there exist mismatch between observation and first guess. The heteroscedasticity and cloud mismatch lead to a skewed and big-tailed prior distribution, which heavily breaks the Gaussian assumption of classic ensemble filters like EnKF or EnSRF.

Despite the non-Gaussian prior distribution, the strong non-linear relationship between radiance observation and state variables could also make EnKF-like algorithm sub-optimal,as the conventional algorithm using a linear regression to spread observation increment to other unobserved variables.The non-linear relationship would be more evident when model resolution is getting finer.

Recently, a Quantile-Conserving Ensemble Filter Framework(QCEFF) is proposed, filters within this framework will ensure each members to stay at the same quantile in prior/posterior PDF when updating an observed variable[Anderson 2022]. For updating an unobserved variable, the regression of observation increments is performed in a space where variables are transformed by the probit and probability integral transforms[Anderson 2023]. This results in a non-linear regression increment for unobserved variables in physical space. Thus a non-Gaussian filters like Rank Histogram Filter (RHF) within QCEFF could be expect to better handle non-Gaussian prior and non-linear relation situations like all-sky radiance assimilation. To explore the potential of QCEFF, the traditional EAKF, RHF, and RHF within QCEFF (QCF_RHF) is evaluated here in an idealized WRF TC case with OSSE.

QCEFF update for unobserved variables

Note prior, posterior(analysis) ensemble of observed variable as y_n^p, y_n^a, prior ensemble of state variable as x_n^p. Assume y_n^p, x_n^p have continuous CDF F_y^p, F_x^p. The probit function is $\Phi^{-1}(Z)$. Variable in probit space (noted with tilde) is transformed using probit function and CDF:

$$\widetilde{x_n^p} = \Phi^{-1}\left[\mathrm{F}_x^p\left(x_n^p\right)\right], \widetilde{y_n^p} = \Phi^{-1}\left[F_y^p\left(y_n^p\right)\right], \widetilde{y_n^a} = \Phi^{-1}\left[F_y^p\left(y_n^a\right)\right] \quad (1)$$

A linear regression is conducted in probit space to get state variable probit posterior, and then transform back to physical space:

$$x_n^a = (F_x^p)^{-1}[\Phi(\widetilde{x_n^a})] \quad (2)$$

Single Observation Experiment

- **Point A**: Cloudy obs, skewed PDF caused by mismatch of clear model priors.
- **Point B**: Clear obs, tail outliers caused by mismatch of cloudy model priors.
- **Point C**: Clear obs, Gaussian-like PDF.

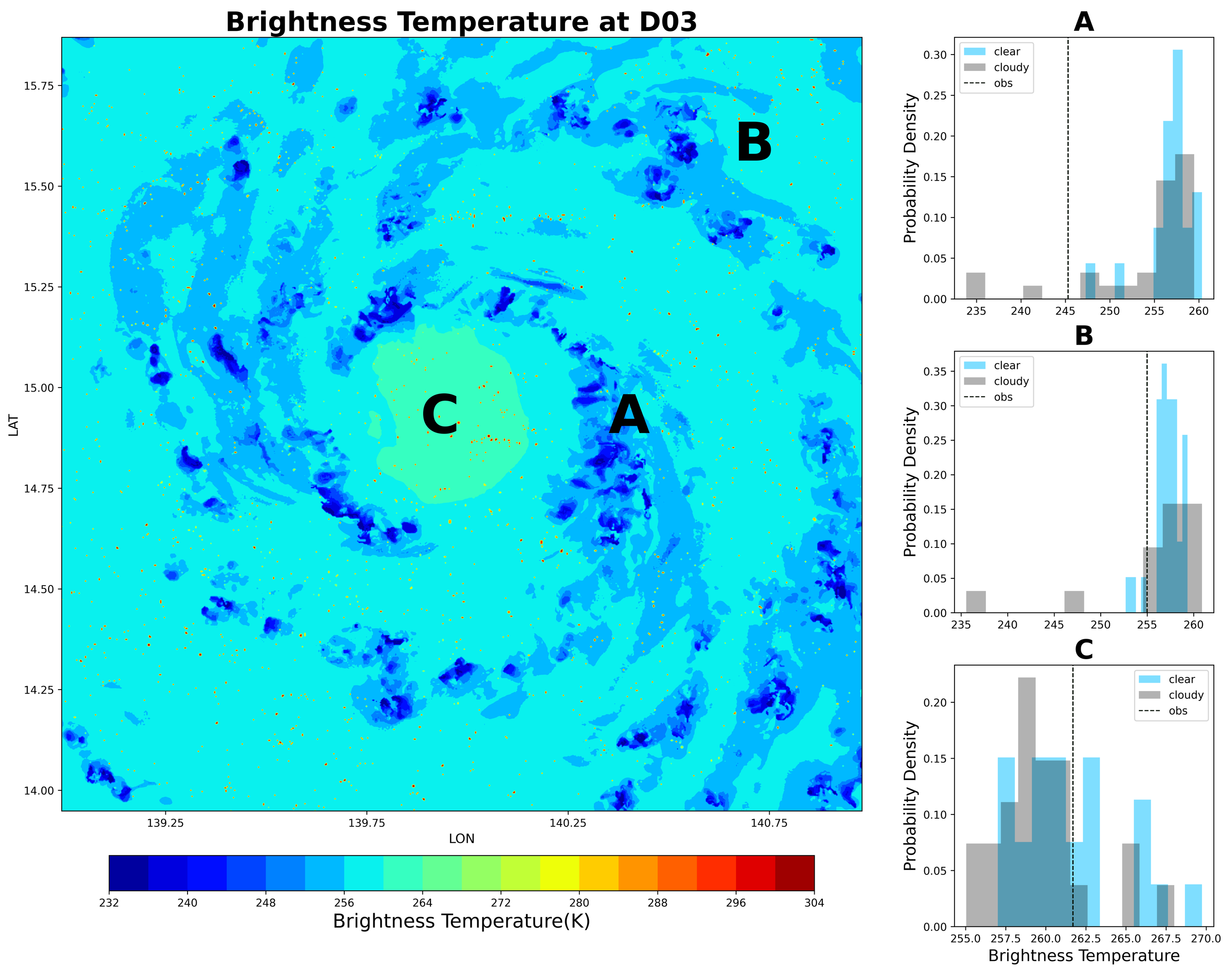

Figure 1. (a) Nature Run brightness temperature(K) at 040030 D03(300m). A, B, and C are selected point observation assimilated in single observation experiment. (b-d) The probability density histogram of selected points A, B, and C, respectively. The azure and gray bars represent PDFs of clear-sky and cloudy members, respectively. The dashed line indicate synthetic BT observation.

Singe Observation Experiment: Point A Results

- Firstguess need large QVAPOR increment in eyewall region. QCF_RHF has the largest increment.
- For EAKF and RHF, the analysis mean can only move along the linear regression line.
- QCF_RHF updates QVAPOR with non-linear regression for each members, the analysis mean is above regression line and more close to NR.

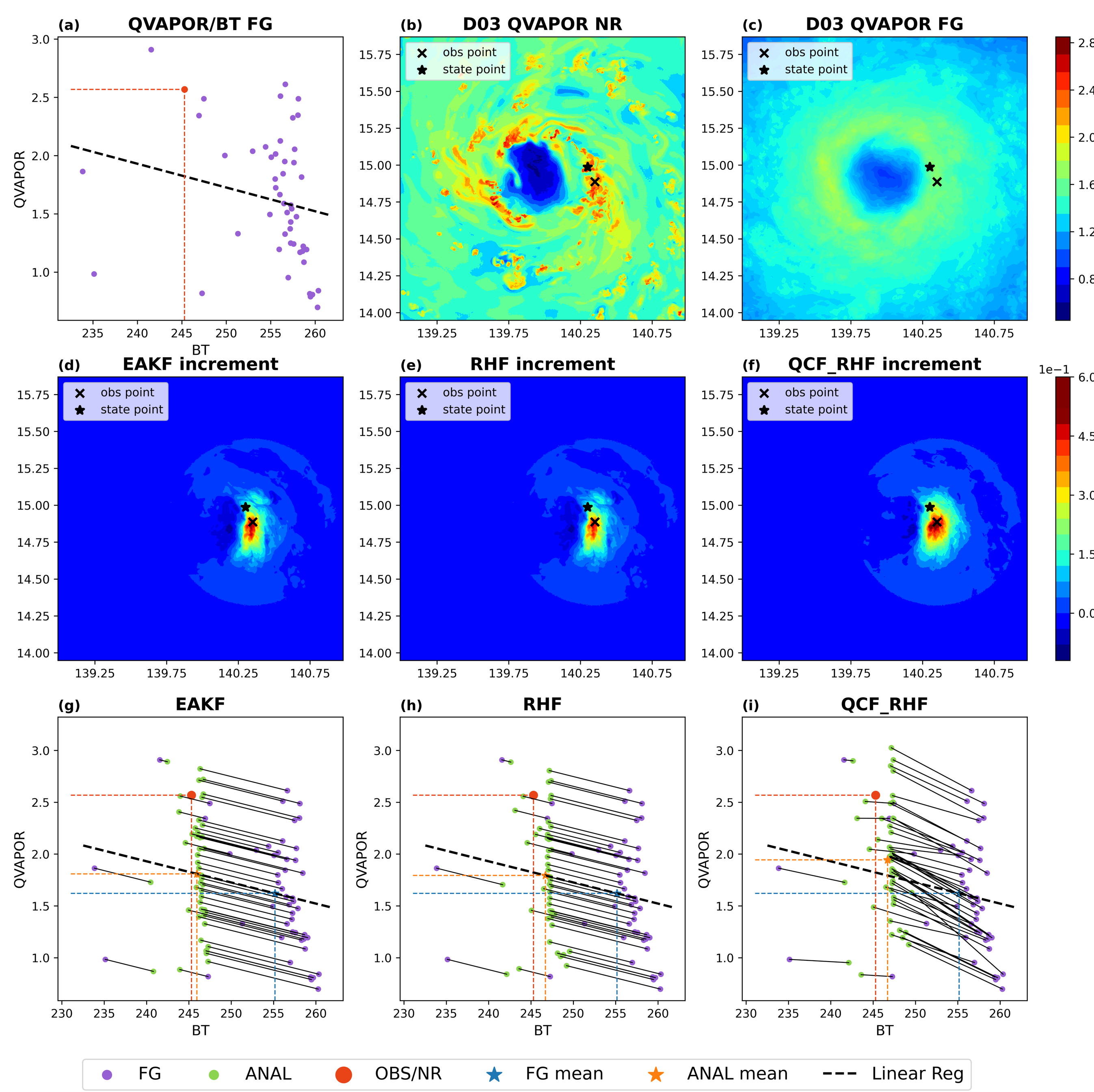

Figure 2. Single observation assimilation results of point A. (a) Firstguess QVAPOR-BT scatter at given model point. Red dot represent BT observation and QVAPOR NR, blue dots are QVAPOR/BT of firstguess members. Dashed line represents linear regression of firstguess QVAPOR-BT. (b-c) Nature run and firstguess field of QVAPOR at vertical level 60. "x" marks the single observation location and "*" marks selected model variable location for scatter figures. (d-e) Increment of assimilating single observation using EAKF, RHF, QCF_RHF, respectively. (g-i) Firstguess and analysis QVAPOR-BT scatter at given model point. Green dots are QVAPOR/BT of analysis members. blue star represents firstguess mean; yellow star for analysis mean.

Offline Assimilation Results: horizontal

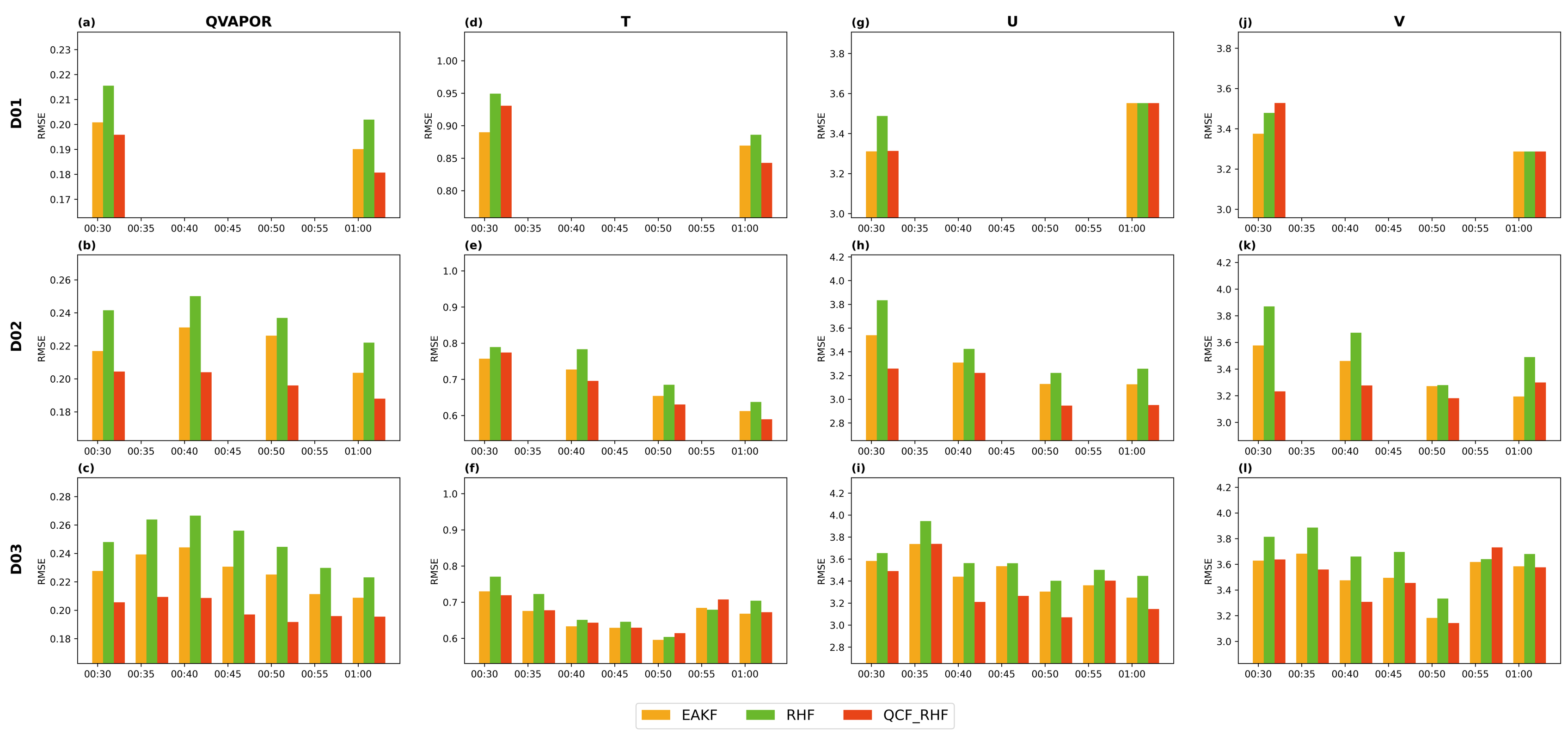

Figure 3. Domain-averaged RMSE of model variable (a-c) QVAPOR, (d-f) T, (g-i) U, (j-l) V at 353hpa. Results at different domain is devided as (a,d,g,j) D01 with 7.5km resolution, (b,e,h,k) D02 with 1.5km resolution and (c,f,i,l) D03 with 300m resolution. Yellow bar for EAKF analysis, green bar for RHF analysis, and red bar for QCF_RHF analysis. Time points with no data is left with blank.

figure/QVAPOR_rmse_profile_forparper.png

Figure 4. Vertical profile of horizontal RMSE for QVAPOR at (a) D01 with 7.5km resolution. (b) D02 with 1.5km resolution. (c) D03 with 300m resolution. Black dashed line for FG, and yellow, green, red solid lines are EAKF/RHF/QCF_RHF analysis, respectively.

- In horizontal, QCF_RHF analysis RMSE of QVAPOR is always the lowest in the model layer closest to synthetic observation height (353hpa).
- In vertical, an evidently lower QVAPOR RMSE could be found for QCF_RHF ranging from 600hpa to 300 hpa.
- The advantage of QCF_RHF in RMSE would be more pronounced if resolution goes up.

There could be critiques that using RMSE as the only evaluation metric is inadequate. Other evaluation like sphere harmonic analysis and image histogram is available by emailing me if you have interests.

Ensemble forecast results

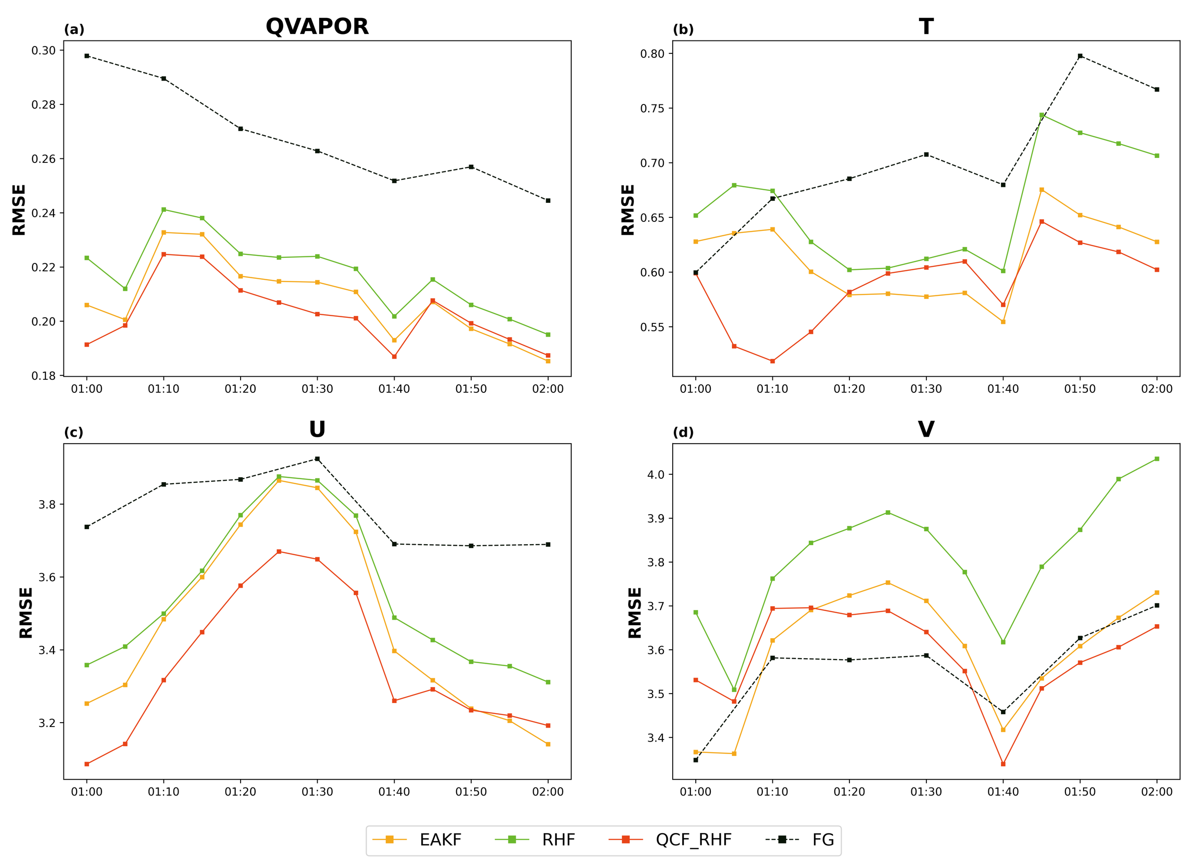

Figure 5. Forecast ensemble mean RMSE at 353hpa. X-axis for time and y-axis for RMSE value. Subplots are (a) QVAPOR, (b) T, (c) U, and (d) V. Black dashed line for FG, and yellow, green, red solid lines are EAKF/RHF/QCF_RHF analysis, respectively.

Starting from 040100, an ensemble forecast with 50 members is launched in D02, resolution 1.5km. Model settings keep unchanged as firstguess construction, with initial field variables replaced by filter analysis

- QCF_RHF outperform than other two filters by having the lowest QVAPOR RMSE during most of forecast range.
- RHF has a poor performance with the highest RMSE value for almost all variables at all available forecast time.

References

- Anderson, J. L. (2023). "A Quantile-Conserving Ensemble Filter Framework. Part II: Regression of Observation Increments in a Probit and Probability Integral Transformed Space". In: *Monthly Weather Review* 151.10, pp. 2759–2777. ISSN: 0027-0644. DOI: 10.1175/mwr-d-23-0065.1. URL: https://dx.doi.org/10.1175/mwr-d-23-0065.1.
- Anderson, J. L. (2022). "A Quantile-Conserving Ensemble Filter Framework. Part I: Updating an Observed Variable". In: *Monthly Weather Review* 150.5. Anderson, Jeffrey L. 1520-0493, pp. 1061–1074. ISSN: 0027-0644. DOI: 10.1175/mwr-d-21-0229.1. URL: https://doi.org/10.1175/MWR-D-21-0229.1.