Deriving Gridded Hourly Rainfall on O'ahu by Combining Gauge and Radar Rainfall

YU-FEN HUANG,[a] YINPHAN TSANG,[a] AND ALISON D. NUGENT[b]

[a] *Natural Resources and Environmental Management, University of Hawai'i at Mānoa, Honolulu, Hawaii*
[b] *Department of Atmospheric Sciences, University of Hawai'i at Mānoa, Honolulu, Hawaii*

(Manuscript received 28 October 2022, in final form 9 August 2023, accepted 11 August 2023)

ABSTRACT: High temporal and spatial resolution precipitation datasets are essential for hydrological and flood modeling to assist water resource management and emergency responses, particularly for small watersheds, such as those in Hawai'i in the United States. Unfortunately, fine temporal (subdaily) and spatial (<1 km) resolutions of rainfall datasets are not always readily available for applications. Radar provides indirect measurements of the rain rate over a large spatial extent with a reasonable temporal resolution, while rain gauges provide "ground truth." There are potential advantages to combining the two, which have not been fully explored in tropical islands. In this study, we applied kriging with external drift (KED) to integrate hourly gauge and radar rainfall into a 250 m × 250 m gridded dataset for the tropical island of O'ahu. The results were validated with leave-one-out cross validation for 18 severe storm events, including five different storm types (e.g., tropical cyclone, cold front, upper-level trough, kona low, and a mix of upper-level trough and kona low), and different rainfall structures (e.g., stratiform and convective). KED-merged rainfall estimates outperformed both the radar-only and gauge-only datasets by 1) reducing the error from radar rainfall and 2) improving the underestimation issues from gauge rainfall, especially during convective rainfall. We confirmed the KED method can be used to merge radar with gauge data to generate reliable rainfall estimates, particularly for storm events, on mountainous tropical islands. In addition, KED rainfall estimates were consistently more accurate in depicting spatial distribution and maximum rainfall value within various storm types and rainfall structures.

SIGNIFICANCE STATEMENT: The results of this study show the effectiveness of utilizing kriging with external drift (KED) in merging gauge and radar rainfall data to produce highly accurate, reliable rainfall estimates in mountainous tropical regions, such as O'ahu. The validated KED dataset, with its high temporal and spatial resolutions, offers a valuable resource for various types of rainfall-related research, particularly for extreme weather response and rainfall intensity analyses in Hawai'i. Our findings improve the accuracy of rainfall estimates and contribute to a deeper understanding of the performance of various rainfall estimation methods under different storm types and rainfall structures in a mountainous tropical setting.

KEYWORDS: Tropics; Rainfall; Hydrometeorology; Gauges; Radars/radar observations; Stochastic models

1. Introduction

The use of gridded rainfall data for hydrological modeling has exponentially increased in the past two decades, which in part shows the critical role of high temporal and spatial resolution rainfall in hydrological studies (e.g., Delrieu et al. 2014; Fares et al. 2014). Reliable results from hydrological models depend on the accuracy and reliability of the model input (Zhu et al. 2013). Inaccurate rainfall data compromise the model output and seriously impact the associated decisions made using model results (McMillan et al. 2011; Zhu et al. 2013). Furthermore, regional hydrological studies benefit from regional datasets at high spatial resolution, which include detailed observations to resolve physical process at fine scales (subdaily and <1 km) and account for terrain characteristics. For small watersheds, the distribution of rain gauges and the temporal resolution of rainfall data are critical in reproducing streamflow hydrographs, as the time of concentration (i.e., time required for runoff to travel from the hydraulically farthest point in the watershed to the outlet) is short. There is a critical need to acquire reliable rainfall estimates at high spatial (e.g., a kilometer or less) and temporal (e.g., hourly or subhourly) resolutions (Berne and Krajewski 2013). Essou et al. (2016) showed that high spatial resolution regional atmospheric reanalysis datasets improved streamflow simulations in humid and subtropical watersheds, with their sizes between 460 and 127 635 km^2, when rain gauges are sparse (i.e., one station per 1000 km^2). Because weather systems could bring orographically induced heavy rainfall with high spatial gradients and various spatial patterns and directions, hydrological modeling for flooding events in small watersheds (e.g., 0.13–635.79 km^2 with mean and median of 18.27 and 5.78 km^2, respectively, based on monitored gauge locations) in Hawai'i is hindered by a lack of appropriate resolutions of rainfall estimates (El-Kadi and Yamashita 2007; Fares et al. 2014).

Rain gauge and weather radar are the most widely used ground-based instruments for acquiring rainfall estimates, and each has its own strengths and weaknesses. Rain gauges provide highly localized rainfall, with regional representation and resolution. Hourly gauge rainfall in Hawai'i was compiled by Huang et al. (2022), and multiple accumulation and spatial interpolation efforts were also made at annual (Giambelluca et al. 2013), monthly (Frazier and Giambelluca 2017), and daily (Longman et al. 2019; Newman et al. 2019) time scales. However, interpolating hourly gauge rainfall is challenging

Corresponding author: Yu-Fen Huang, yfhuang@hawaii.edu

DOI: 10.1175/JHM-D-22-0196.1

because of its inherent characteristics at a fine temporal scale (e.g., heterogeneous rain gauge distribution, zero values, missing values). While radar supplies high temporal and spatial resolution rain rate and rainfall accumulation, the accuracy is usually lower than that of rain gauges due to increasing scanning angles at distances farther away from the radar station, particularly when radar stations are on different islands; potential beam blockage by terrain; attenuation; and dynamic relationships between rainfall reflectivity and rain rate. Giambelluca et al. (2013) attempted to incorporate the National Oceanic and Atmospheric Administration (NOAA) Next Generation Weather Radar (NEXRAD) level III radar estimates into mean monthly and annual rainfall estimates in Hawai'i. However, they did not include radar data on final annual rainfall interpolations, because they provided little improvement over other predictors. Instead, they included Parameter-Elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 1994), fifth-generation Pennsylvania State University–NCAR Mesoscale Model (MM5; Dudhia 1993) with localized adjustment (Yang and Chen 2008), and vegetation information (see Giambelluca et al. 2013) as predictors with rain gauge data. Since 2020, the NOAA National Weather Service in Hawai'i has used the Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation (QPE) to support their forecasts. The MRMS system is implemented at the National Centers for Environmental Prediction (NCEP) and commonly used in the continental United States. It generates gridded rainfall at resolutions of 1 km and subdaily radar QPE incorporating gauge rainfall by bias adjustment and precipitation climatology (Zhang et al. 2016). Prior to MRMS, there was no other subdaily integrated radar and gauge QPE in Hawai'i. Therefore, radar rainfall is underutilized by hydrologists because of the aforementioned limitations and poor historical accessibility (Fares et al. 2014). Recently, Huang et al. (2022) improved data accessibility of hourly gauge and radar rainfall across the Hawaiian Islands. In this study, we will extend these efforts by 1) producing a spatially continuous, high temporal (e.g., hourly) and high spatial (e.g., 250-m grid) resolution rainfall dataset and 2) examining whether a merged dataset (2015–20) of gauge and radar rainfall improves the accuracy of rainfall estimates for O'ahu.

In the past two decades, many approaches have been used to merge different types of rainfall measurements, for example, gauge, radar, and satellite measurements. Merging gauge and radar rainfall is a common exercise to obtain high temporal and spatial resolution rainfall (Goudenhoofdt and Delobbe 2009; Haberlandt 2007). By combining radar and gauge measurements, we expect to keep the strength and minimize the weakness of each measurement. Approaches for merging gauge and radar rainfall can be categorized into three groups: radar bias reduction, rain gauge interpolation with radar spatial pattern, and gauge–radar integration (e.g., Ochoa-Rodriguez et al. 2019; McKee and Binns 2016; Sideris et al. 2014; Goudenhoofdt and Delobbe 2009). The radar bias reduction method includes the mean field bias (MFB) reduction (Hitschfeld and Bordan 1954) and its modifications (Brandes 1975; Wilson 1970; Wilson and Brandes 1979; Borga 2002; Michelson and Koistinen 2000). Radar bias reduction methods improve radar rainfall based on the gauge rainfall; however, the adjustments are applied on the radar coverage and do not provide full coverage of the spatial extent. This is due to the possibility of missing values in radar data because of beam blockage by the complex terrain. The interpolations of rain gauge with radar spatial patterns include conditional merging, that is, ordinary kriging with radar-based error correction (KRE; Sinclair and Pegram 2005) and kriging with external drift (KED). Conditional merging uses the radar field to estimate the error associated with ordinary kriging (OK) based on gauge rainfall and corrects it. KED uses radar as secondary information to gauge rainfall and assumes the gauge rainfall is linearly correlated with radar rainfall. The gauge–radar integration includes cokriging (CoK) (Krajewski 1987) and Bayesian data combination (BDC; Todini 2001) and aims to minimize the estimation uncertainty. The interpolations with radar spatial patterns assume gauge rainfall is the true rainfall, whereas the gauge–radar integration assumes that the two measurements are both equally true rainfall. In this study, we adopt the rain gauge interpolation with radar spatial pattern and aim to improve the spatial coverage to extend over a mountainous tropical island. To facilitate the later hydrological application, we consider the rain gauge measurements as the true rainfall received on the ground.

McKee and Binns (2016) pointed out a few considerations when selecting the gauge–radar merging methods, including 1) density of rain gauge networks, 2) climate and storm characteristics, 3) proximity of the radar stations, 4) basin response time, and 5) time step of adjustment. With the above considerations, we decided to adopt KED for merging gauge and radar rainfall on O'ahu for the following reasons:

- Gauge rainfall in Hawai'i has relatively high confidence compared to available radar rainfall (see the reasons we listed in the second paragraph of the introduction). With that in mind, rain gauge interpolation with radar spatial pattern is superior to gauge–radar integration, because it considers gauge rainfall with more accuracy as true rainfall. KED is superior to KRE, because KRE still relies on the high quality of radar rainfall (Rabiei and Haberlandt 2015).
- The density of rain gauges on O'ahu (≥70 hourly rain gauges over an area of 1546 km^2 with the average nearest-neighbor distance of 4.1 km) is enough to overcome the sensitivity issues of the KED method to rain gauges (McKee and Binns 2016; Berndt et al. 2014; Goudenhoofdt and Delobbe 2009).
- Previous studies suggest that KED outperformed other methods for merging gauge and radar rainfall (e.g., Erdin 2009; Velasco-Forero et al. 2009; Rabiei and Haberlandt 2015; Delrieu et al. 2014).

Although KED is widely used for merging gauge and radar rainfall and has been used to generate high temporal resolution (subhourly) rainfall data, most prior studies were conducted on continents (e.g., Erdin 2009; Velasco-Forero et al. 2009; Haberlandt 2007; Rabiei and Haberlandt 2015; Ly et al. 2013; Delrieu et al. 2014; Berndt et al. 2014) or bigger islands in the midlatitudes (e.g., Great Britain; Nanding et al. 2015; Cecinati et al. 2017; Jewell and Gaussiat 2015). These studies rarely focused on mountainous areas, or if they did, the performance of KED in mountainous areas was worse than in flatter areas, and rainfall was poorly validated (e.g., Erdin 2009; Jewell and Gaussiat 2015). The performance of KED on mountainous

tropical islands with different storm types and rainfall structures (convective vs stratiform) remains unknown.

The aim of this study is to deliver gridded hourly rainfall suitable for studying severe floods in mountainous tropical islands. We plan to integrate the available radar and rain gauge data to deliver gridded hourly rainfall on O'ahu as a pilot study to identify methods to generate high temporal and spatial rainfall data for mountainous tropical islands. We evaluated and validated the approaches by examining the performance of radar-only, gauge-only, and merged radar–gauge (i.e., KED) rainfall estimates during 18 severe storm events to answer the following questions: 1) Does merging gauge and radar rainfall with the KED method perform better than using gauge-only or radar-only rainfall estimates? 2) If the KED rainfall estimate is better than gauge or radar rainfall, is it uniformly superior in all weather systems? The organization of this paper is as follows: in section 2, we provide information on the data sources and a description of event selection, gauge rainfall, and radar rainfall; in section 3, we outline the methods used for identifying rainfall types, merging different rainfall products, and conducting validations; the results are presented based on events, storm types, and rainfall structures in section 4; in section 5, we discuss the advantages of incorporating radar data into rainfall estimates, address potential radar bias, highlight improvements in estimating convective rainfall, discuss applications, and outline the limitations of our study; finally, we conclude the paper with key findings and recommendations for applying merging radar–gauge on mountainous tropical islands in section 6. The supplementary tables and figures supporting this study can be found in appendix A. A glossary of abbreviations and terms used in this paper can be found in appendix B.

2. Data

a. Event selection

We incorporated rain gauge and radar data to generate the hybrid QPE for 18 severe rainfall events between 2015 and 2020 on O'ahu (Table 1). We selected the rainfall events reported in the NCEI Storm Events Database (https://www.ncdc.noaa.gov/stormevents/). Based on the descriptions of each event, the events can be summarized into five storm types: tropical cyclone (TC), cold front, upper-level trough, kona low, and mix of upper-level trough and kona low. Tropical cyclones usually only occur during the dry season (May–October), with maximum sustained winds > 39 mph. Cold fronts can produce heavy rainfall, along with southerly prefrontal winds and northerly postfrontal winds (Kodama and Barnes 1997; Longman et al. 2021). Upper-level troughs occur near or over the Hawaiian Islands any time of the year and are identified by the absence of a low-pressure center at the surface (Kodama and Barnes 1997; Longman et al. 2021). Finally, kona lows occur when a cold core system gets cut off from the upper-level westerly winds and develops a closed surface low (Simpson 1952; Kodama and Barnes 1997; Longman et al. 2021). In this study, we have three TCs, two cold fronts (CFs), seven upper-level troughs (ULs), two kona lows (KLs), and four mixed upper-level trough and kona lows (Mix).

b. Rain gauge data

We separated the quality-controlled hourly gauge rainfall from 73 rain gauges on O'ahu from Huang et al. (2022) into subsets (Fig. 1). For each event, a rain gauge was included in the analysis if the rain gauge had hourly rainfall data covering more than 80% of the validation duration, one day before/after the begin/end dates. The number of rain gauges used for each event validation ranges from 42 to 66. The correlations between pairs of gauges and their distance are in negative relationships (Fig. A1 in appendix A).

c. Radar data

We used NEXRAD level II radar reflectivity (National Weather Service Radar Operations Center 1991) and sounding data from the atmospheric sounding website hosted by the Department of Atmospheric Science, University of Wyoming (http://weather.uwyo.edu/upperair/sounding.html), to identify rainfall type and the hourly radar rainfall data with a grid size of $0.005° \times 0.005°$. The NEXRAD system is a network of high-resolution S band dual-polarimetric Doppler radars, with maximum coverage of a 230-km radius for Doppler velocity and a 460-km radius for reflectivity, with the typical lowest sweep angle at 0.5°.

The radar rainfall was obtained by applying the Lidar Radar Open Software Environment (LROSE; Dixon and Javornik 2016) on NEXRAD level II radar reflectivity with 5–6-min temporal resolution. There were no missing radar data in selected storm events. The application of LROSE to radar data provides data quality control algorithms, including clutter detection and mitigation (Hubbert et al. 2009), quality metrics and error assessment (Bell et al. 2013), attenuation correction in precipitation (Gu et al. 2011), and the vertical profile of reflectivity adjustment (Kirstetter et al. 2013). For the hourly rainfall, the rain rate accumulates within the radar volume at each grid and is converted to precipitation depth at hourly intervals. We applied the modified NCAR hybrid method (Dixon et al. 2015) with hydrometeor information and localized the reflectivity–rain rate relationship, $Z = 250R^{1.2}$, where Z is the radar reflectivity and R is the rain rate of each radar scan. The detailed processes are described in Huang et al. (2022).

The radar rainfall product in Huang et al. (2022) is a product of four NEXRAD radar stations across the state and was masked to exclude areas with high uncertainty. Here, we used the radar rainfall dataset without the mask because we expected to incorporate gauge rainfall to reduce the bias/errors. The radar rainfall on O'ahu mainly came from the radar station on Moloka'i (PHMO; Fig. A2). The radar rainfall for the northeast of O'ahu, where the lower angle beams of PHMO were blocked by terrain, came from Kaua'i (PHKI; Fig. A2). The distances between PHMO and the rain gauges on O'ahu range from 53 to 123 km, and the distances between PHKI and the gauges range from 138 to 206 km. All rain gauges are in both radars' scanning range; however, the accuracy of radar rainfall usually decreased when the target locations were farther away, when the gauges could only be scanned at a higher angle.

3. Methodology

a. Rainfall structure

In this study, the rainfall structure (i.e., convective vs stratiform) of each event was assessed and determined to facilitate

TABLE 1. The results from point-to-grid cross validation of 18 selected severe storm events between 2016 and 2020 for validation (three TCs, two CFs, seven ULs, two KLs, and four Mix). The validation period starts and ends at 0000 UTC. Radar indicates radar-only rainfall. Bold font represents the best metric for the event between OK, radar, and KED. Refer to sections 3b and 3c for details on methods of rainfall estimates and validation indices.

Event ID	Storm types	Validation period	No. of rain gauges	Mean BIAS (mm)			Mean RMSE (mm)			R^2			MaxRR		
				OK	Radar	KED	OK	Radar	KED	OK	Radar	KED	OK	Radar	KED
TC-01	TC	22 Jul 2016–26 Jul 2016	59	−0.35	−0.48	**−0.27**	0.73	0.75	**0.71**	0.58	0.58	**0.67**	**0.92**	0.79	**1.08**
CF-01	CF	9 Feb 2017– 13 Feb 2017	60	**−0.21**	0.17	−0.22	0.52	0.53	**0.46**	0.71	**0.76**	**0.76**	0.31	**0.51**	0.41
Mix-01	Mix	27 Feb 2017– 2 Mar 2017	50	−0.39	**0.27**	−0.29	0.75	0.83	**0.68**	0.48	**0.52**	**0.52**	0.29	**0.88**	0.45
Mix-02	Mix	22 Oct 2017–25 Oct 2017	66	**−0.11**	0.37	−0.12	0.41	0.65	**0.38**	0.71	0.64	**0.72**	0.64	**0.90**	0.75
UL-01	UL	10 Nov 2017–13 Nov 2017	64	−0.24	−0.30	**−0.16**	0.52	0.50	**0.47**	0.46	**0.55**	0.52	0.50	**0.88**	0.69
Mix-03	Mix	16 Feb 2018–20 Feb 2018	56	−0.23	**0.06**	−0.10	0.48	0.58	**0.42**	0.66	0.67	**0.71**	0.49	1.16	**0.96**
UL-02	UL	23 Feb 2018–27 Feb 2018	62	−0.23	−0.19	**−0.13**	0.39	0.34	**0.32**	0.42	0.47	**0.49**	0.45	0.80	**0.98**
UL-03	UL	30 Mar 2018–2 Apr 2018	63	**0.00**	0.26	−0.05	0.17	0.30	**0.10**	0.72	**0.74**	**0.74**	0.18	**1.09**	0.24
UL-04	UL	11 Apr 2018–16 Apr 2018	64	−0.21	**−0.13**	−0.17	0.41	0.46	**0.38**	0.42	0.14	**0.48**	0.82	1.71	**1.03**
TC-02	TC	24 Aug 2018–29 Aug 2018	63	−0.39	−0.35	**−0.23**	0.66	0.62	**0.59**	0.41	**0.49**	**0.49**	0.30	**0.95**	0.84
TC-03	TC	11 Sep 2018–15 Sep 2018	62	−0.30	**−0.14**	−0.23	0.67	0.73	**0.63**	**0.66**	0.64	**0.66**	0.65	**1.07**	0.88
UL-05	UL	10 Oct 2018–13 Oct 2018	63	−0.29	**−0.12**	−0.17	0.46	0.44	**0.39**	0.28	**0.41**	0.40	0.45	1.33	**0.81**
KL-01	KL	28 Oct 2018–1 Nov 2018	63	−0.20	**0.15**	−0.18	0.54	0.76	**0.51**	0.67	0.66	**0.69**	**0.95**	1.30	0.87
CF-02	CF	8 Nov 2018–12 Nov 2018	60	−0.24	−0.19	**−0.11**	0.47	0.42	**0.39**	0.34	**0.52**	0.49	0.69	1.26	**1.06**
UL-06	UL	12 Feb 2019–16 Feb 2019	66	−0.31	−0.23	**−0.22**	0.63	0.63	**0.58**	0.44	0.45	**0.50**	0.57	**0.94**	0.73
KL-02	KL	23 Jun 2019–27 Jun 2019	64	−0.23	0.48	**−0.19**	0.58	1.11	**0.55**	0.63	0.55	**0.64**	0.60	2.13	**0.88**
Mix-04	Mix	12 Sep 2019–16 Sep 2019	42	−0.30	0.58	**−0.18**	0.49	0.80	**0.38**	0.16	**0.41**	0.36	0.43	2.52	**0.53**
UL-07	UL	9 Oct 2019–13 Oct 2019	40	−0.20	0.42	**−0.16**	0.40	0.62	**0.35**	0.59	**0.62**	0.61	0.65	1.49	**0.81**

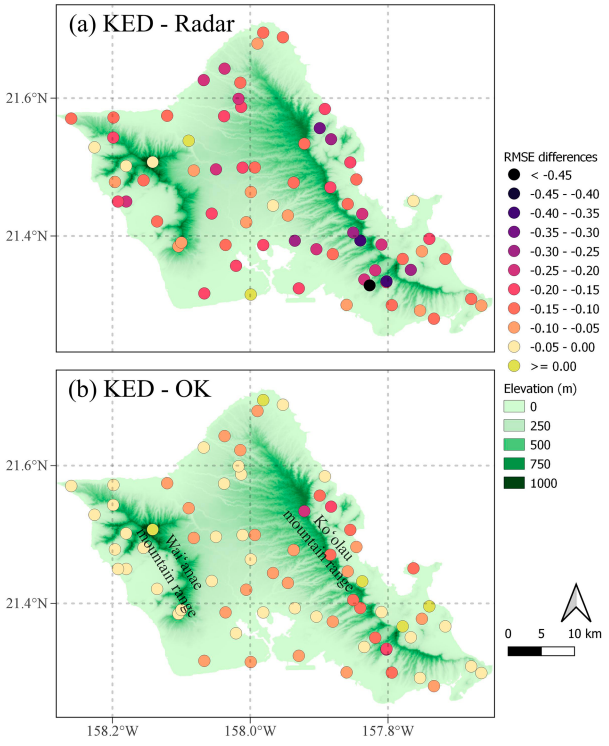

FIG. 1. Locations of the 73 rain gauges on Oʻahu. The colors denote the differences in RMSEs of 18 events between (a) radar and KED rainfall estimates and (b) OK and KED. Purple indicates KED performed better than the other option. The deeper the color, the better the rainfall estimate. Refer to sections 3c and 3d for details on methods of rainfall estimates and validation indices.

the evaluation of the QPE methods' accuracy during different structures. We identified the rainfall structure at every radar grid on every radar scan using open-source software LROSE (Dixon and Javornik 2016; Dixon and Romatschke 2022). The convective and stratiform partitioning algorithm was based on Steiner et al. (1995) (Dixon and Romatschke 2022). The algorithm first excludes grids with reflectivity lower than 10 dBZ and keeps the rest as valid convective grids, because convective rainfall usually corresponds to stronger radar reflectivity. Next, convective grids are defined as having a vertical column (between 0 and 10 km) maximum of radar reflectivity larger than 40 dBZ. In addition, convective grids are identified by their "texture," which is calculated by the square root of the standard deviation of the squared reflectivity within 7 km of each valid grid based on the radar resolution (Dixon and Romatschke 2022). A convective grid is identified when more than one-third of the surrounding grids have a texture value larger than 15 dBZ within a 7-km radius. The texture value is then translated into a value between 0 and 1 (see Dixon and Romatschke 2022, their Fig. 3). A convective grid has a value of 1, while stratiform has a value of 0 in the LROSE product for every radar scan. The convectiveness, which stands for the hourly proportion of convective precipitation at each radar grid (0%–100%), is averaged hourly by grid spacing of the radar rainfall grid. To simplify the experiments, we identified binary rainfall structures based on convectiveness. At each hour, we considered convectiveness $< 50\%$ as stratiform rainfall, whereas convectiveness $\geq 50\%$ was considered convective rainfall.

b. Merging methods

We used OK to interpolate "gauge-only" rainfall to compare possible benefits from incorporating radar rainfall when applying KED. For both OK and KED, we included all available gauges for each event. Before all the kriging processes, we added a small value, 0.001, and applied a log transformation to gauge and radar rainfall to give a more Gaussian-like distribution and prevent the issue caused by zero values. The interpolated log rainfall was then transformed back to rainfall estimates and subtracted by the added value of 0.001. OK is a popular interpolation method and has the lowest error in the applications of daily to yearly rainfall in Hawai'i (Frazier et al. 2016; Mair and Fares 2011). OK interpolates data with a weighted average, where the weights are a function of the distance between the rain gauge measurements and estimated point (e.g., variogram or semivariogram; Goovaerts 2000, 1997; Kitanidis 1997). In this study, OK interpolated gauge rainfall into the configuration of $0.005° \times 0.005°$, the same as the radar rainfall grid, with the following:

$$\mathrm{RF}_{\mathrm{OK}}(x_0) = \sum_{i=1}^{n} \lambda_i^{\mathrm{OK}} \mathrm{RF}_G(x_i),$$

where RF in this study refers to hourly rainfall, with the method in its subscript in this study; here, $\mathrm{RF}_{\mathrm{OK}}(x_0)$ is the OK-estimated rainfall at the unknown location x_0; $\mathrm{RF}_G(x_i)$ are the known rainfall values at locations $x_1, x_2, \ldots, x_n$; and λ_i^{OK} are OK weights, which minimize the variance under unbiased conditions (the sum of weights is one). The weights are determined by the variogram model using a covariance function C at the distance between each rain gauge point.

$$\begin{cases} \sum_{j=1}^{n} \lambda_i^{\mathrm{OK}} C(x_i, x_j) + \mu = C(x_i, x_0) \\ \sum_{j=1}^{n} \lambda_i^{\mathrm{OK}} = 1 \end{cases}, \tag{1}$$

where μ is the value of the constant mean. The assumption of $\mu = 0$ is commonly made in OK when the underlying trend or drift is not considered significant. If μ is assumed to have a nonzero value, it would introduce a spatially varying trend and lead to incorrect rainfall in areas where the drift is not adequately modeled. Given the spatial complexity and diverse storm types and rainfall structure in our study, we opted for a more conservative approach by assuming μ to be zero.

On the other hand, KED is an efficient algorithm that allows the incorporation of secondary variables as additional information, which are assumed to be linearly related to the expected value of the primary variable. In this study, KED allows us to derive rainfall estimates with rain gauges and observe spatial features of the radar rainfall. KED with gauge and radar rainfall has been successfully applied in many locations, including Germany, the Netherlands, and the United Kingdom (Goudenhoofdt and Delobbe 2009; Haberlandt 2007; Cecinati et al. 2017;

Nanding et al. 2015; Jewell and Gaussiat 2015; Courty et al. 2018; Rabiei and Haberlandt 2015; Schuurmans et al. 2007). KED follows OK but considers the mean of the study field nonstationary in space. KED models a drift term (mean) plus a residual term, where the drift term is an unknown linear function defined by the radar field, as follows:

$$m_G(x) = a + b \times \mathrm{RF}_R(x), \tag{2}$$

where $m_G(x)$ is the drift term (mean), and $\mathrm{RF}_R(x)$ is the hourly radar rainfall at location x, while a and b are linear coefficients to be determined. KED has a similar expression to OK but with different weights:

$$\mathrm{RF}_{\mathrm{KED}}(x_0) = \sum_{i=1}^{n} \lambda_i^{\mathrm{KED}} \mathrm{RF}_G(x_i), \quad \text{with}$$
$$\begin{cases} \sum_{j=1}^{n} \lambda_i^{\mathrm{KED}} C_R(x_i, x_j) + \mu_0 + \mu_1 \mathrm{RF}_R(x_i) = C_R(x_i, x_0) \\ \sum_{j=1}^{n} \lambda_i^{\mathrm{KED}} = 1 \\ \sum_{j=1}^{n} \lambda_i^{\mathrm{KED}} \mathrm{RF}_G(x_i) = \mathrm{RF}_G(x_0) \end{cases}, \tag{3}$$

where $\mathrm{RF}_{\mathrm{KED}}$ is the rainfall estimate from KED with the weights λ_i^{KED}; $\mathrm{RF}_G(x)$ and $\mathrm{RF}_R(x)$ are gauge and radar rainfall at location x, respectively; C_R is the covariance of the residuals $\mathrm{RF}_G(x) - m_G(x)$; and μ_0 and μ_1 are Lagrangian multipliers. Both OK and KED apply an isotropic spherical model on the gauge rainfall for the variogram, with parameters automatically calculated using the L1-norm minimization scheme (Murphy et al. 2021) for finding the best fit solution by minimizing the sum of the absolute differences between observed and predicted values.

c. Validation and evaluation

We evaluated all gridded products against the rain gauge measurements in the same grid at hourly intervals. Radar rainfall was validated at all rain gauge locations for each event. Both OK and KED rainfall were validated with leave-one-out cross validation (LOOCV), where one rain gauge was left out during the kriging process. This was repeated for each rain gauge and validated with the remaining rain gauges in each of the 18 events.

Six indices were used to quantitatively evaluate radar, OK, and KED rainfall estimates: bias (BIAS); root-mean-square error (RMSE); coefficient of determination (R^2); maximum rainfall ratio (MaxRR); and area under the curve (AUC), where the curve is the receiver operating characteristics (ROCs).

1) BIAS

BIAS shows the amount of rainfall under- or overestimated from the first time step $t = 1$ to the last time step $t = T$:

$$\mathrm{BIAS}(x_i) = \frac{1}{T} \sum_{t=1}^{T} [\mathrm{RF}_{\mathrm{EST},t}(x_i) - \mathrm{RF}_{G,t}(x_i)], \tag{4}$$

where $\mathrm{RF}_{\mathrm{EST},t}(x_i)$ is the rainfall estimate of radar, OK, or KED, and $\mathrm{RF}_{G,t}$ is the original gauge rainfall at time step t at rain gauge location x_i. The best estimation has BIAS = 0. To compare the results between events, we normalized BIAS (nBIAS) into nBIAS by dividing BIAS by the range of the event rainfall values, which represents the difference between the maximum and minimum rainfall values for each event.

2) RMSE AND NORMALIZED RMSE (NRMSE)

The RMSE is a measurement of accuracy and indicates the errors between the estimated value and the gauged value:

$$\mathrm{RMSE}(x_i) = \sqrt{\frac{\sum_{t=1}^{T} [\mathrm{RF}_{\mathrm{EST},t}(x_i) - \mathrm{RF}_{G,t}(x_i)]^2}{T}}. \tag{5}$$

The best estimation has RMSE = 0. To compare the results between events, we calculated the nRMSE by dividing the RMSE by the range of the event rainfall values, which represents the difference between the maximum and minimum rainfall values for each event.

3) R^2

The R^2 equals the square of the Pearson correlation coefficient, which measures the precision of the relationship between the estimated and gauge rainfall:

$$r(x_i) = \frac{\sum_{t=1}^{T} [\mathrm{RF}_{\mathrm{EST},t}(x_i) - \overline{\mathrm{RF}_{\mathrm{EST}}(x_i)}][\mathrm{RF}_{G,t}(x_i) - \overline{\mathrm{RF}_G(x_i)}]}{\sqrt{\sum_{t=1}^{T} [\mathrm{RF}_{\mathrm{EST},t}(x_i) - \overline{\mathrm{RF}_{\mathrm{EST}}(x_i)}]^2 \sum_{t=1}^{T} [\mathrm{RF}_{G,t}(x_i) - \overline{\mathrm{RF}_G(x_i)}]^2}}, \tag{6}$$

where $\overline{\mathrm{RF}_{\mathrm{EST}}(x_i)}$ and $\overline{\mathrm{RF}_G(x_i)}$ are the means of estimated and gauge rainfall at location x_i, respectively. The best estimation has $R^2 = 1$.

4) MAXRR

In addition to the three indices above that show the overall bias and precision in rainfall estimates over the period of an event, we applied MaxRR [see Eq. (7)] to measure how the maximum hourly rainfall of each merging method in an event compares to that of the gauge rainfall. The maximum rainfall ratio is defined as the ratio of maximum hourly estimated rainfall from each gauge location to the maximum hourly gauge rainfall across all left-out rain gauges in the LOOCV iterations:

$$\mathrm{MaxRR} = \frac{\max(\{\max[\mathrm{RF}_{\mathrm{EST},1}(x_1), \ldots, \mathrm{RF}_{\mathrm{EST},T}(x_1)]\}, \ldots, \{\max[\mathrm{RF}_{\mathrm{EST},1}(x_n), \ldots, \mathrm{RF}_{\mathrm{EST},T}(x_n)]\})}{\max(\{\max[\mathrm{RF}_{G,1}(x_1), \ldots, \mathrm{RF}_{G,T}(x_1)]\}, \ldots, \{\max[\mathrm{RF}_{G,1}(x_n), \ldots, \mathrm{RF}_{G,T}(x_n)]\})}, \tag{7}$$

where n is the total number of rain gauges during the validation period. The best estimation has MaxRR = 1. Note that the maximum gauge rainfall from each event ranged between 47 and 103 mm, which exceeded the definition of extreme rainfall (≥20 mm h^{-1}) for tropical rainfall in Tokay and Short (1996) and heavy rainfall (≥7.6 mm) in general according to the American Meteorological Society (2023).

5) ROC CURVE AND AUC

The ROC curve is a curve of the true positive rate (TPR) and false positive rate (FPR):

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}; \quad \mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}, \tag{8}$$

where TP is the true positive (the number of positive cases predicted that match actual positive cases), FN is the false negative (i.e., type II error; the number of negative cases predicted that were actually negative), FP is the false positive (i.e., type I error; the number of positive cases predicted that were actually negative), and TN is the true negative (the number of negative cases predicted that match actual negative cases). Here, we used the ROC curve to evaluate each rainfall estimate at all rain gauge locations with 15 threshold percentiles, including 0th, 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th, 95th, 97.5th, 98th, 99th, 99.5th, and 99.9th percentiles. The count of TP, FN, FP, and TN is based on the estimated rainfall that is greater than or equal to the threshold. We plot the ROCs starting from the 80th to 99.9th percentile (Fig. 2), with corresponding values 0.254, 1.524, 4.064, 7.620, 9.144, 14.478, 21.844, and 42.164 mm, respectively. The overall performances of accuracy are summarized to the AUC, which is an aggregated evaluation of accuracy across all the thresholds and ranges from zero to one. When the AUC scores 1.0, this indicates the best overall accuracy in rainfall estimate. When the AUC is 0.5, this means that the rainfall estimates are no different from guesses, and when the AUC scores 0.0, this means the rainfall estimates are the opposite of the truth (e.g., when the observed rainfall is larger than or equal to the threshold, the estimated rainfall will be smaller than the threshold).

Among all the indices, BIAS allows us to determine whether the rainfall estimates are under- or overestimated overall, but the value can be offset by different time steps. The RMSE is the most recommended and used index to compare model performances (Willmott 1982) and represents the differences between estimated and actual values. The evaluation of different QPEs will be mainly based on the RMSE. The R^2 value provides an intuitive view of how well the estimations fit the observations. Although previous studies (e.g., Willmott 1982; Taylor 1990) indicated that the Pearson correlation coefficient could be misleading when used to compare model-estimated and observed variables, R^2 is commonly used in the statistical evaluation of model performances because it shows how much of the variance in the dependent variable that is explained by the independent variables in the model. Therefore, we still include R^2 to understand whether the estimated and observed time series covary over time. The MaxRR was developed because maximum rainfall critically drives stormflow responses in flood-related research, which is one of the reasons to deliver hourly gridded rainfall on O'ahu. It evaluates the estimated maxima against the observed maximum rainfall at rain gauges; however, this evaluation can be imperfect when the locations of maxima between the estimated and observed rainfall differ. For example, if a storm event had multiple peak rainfall within an event, the low density of gauges could not capture the peak at the maximum locations to be evaluated against radar or interpolated estimates (i.e., KED). The ROC curve and AUC are widely used to effectively evaluate the overall prediction accuracy of a classification model (e.g., Ma and Huang 2007; Fawcett 2006). The ROC curve represented the proportion of correct estimations to various thresholds, while the AUC represented a summary of how the model performs at all thresholds. In addition, we applied Tukey's test (Tukey 1949) to determine the statistical significance ($p \leq 0.05$) of the indices of RMSE, R^2, and MaxRR. Instead of taking a mean over all rain gauge locations for each index, we keep the indices at each rain gauge location for each event duration.

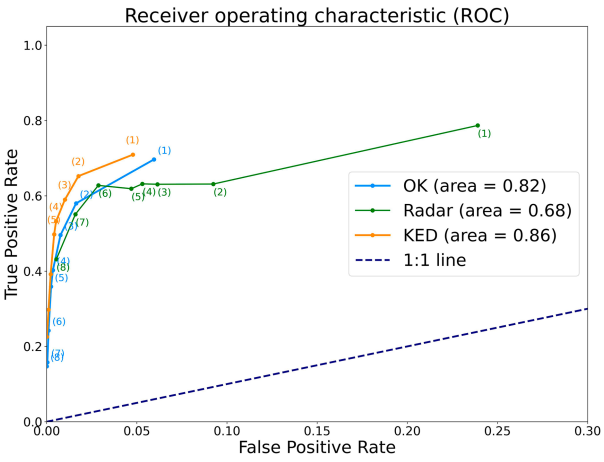

FIG. 2. ROCs at different rainfall thresholds of hourly rainfall (annotations 1–8 are 0.254, 1.524, 4.064, 7.620, 9.144, 14.478, 21.844, and 42.164 mm, respectively) with AUC values between different methods of rainfall estimates. OK is denoted in blue, radar in green, and KED in orange. In the ROC curve, the closer the point is to the top left (i.e., TPR = 1 and FPR = 0), the better. Note that the x and y axes are on different scales.

4. Results

a. Evaluation by event

The validation results for the hourly rainfall from each event are shown in Table 1 and Fig. 3. The event-mean BIAS of OK, radar, and KED ranges from −0.39 to 0, from −0.48 to 0.58, and from −0.29 to −0.05 mm, respectively. Although the event-mean BIAS showed that OK and KED always underestimated hourly rainfall, KED had a smaller bias than OK for most events (n = 15). While radar had the event-mean BIAS (0.01 mm) closest to zero, the large range of its mean BIAS (from −0.48 to 0.58 mm) indicates the uncertainty and

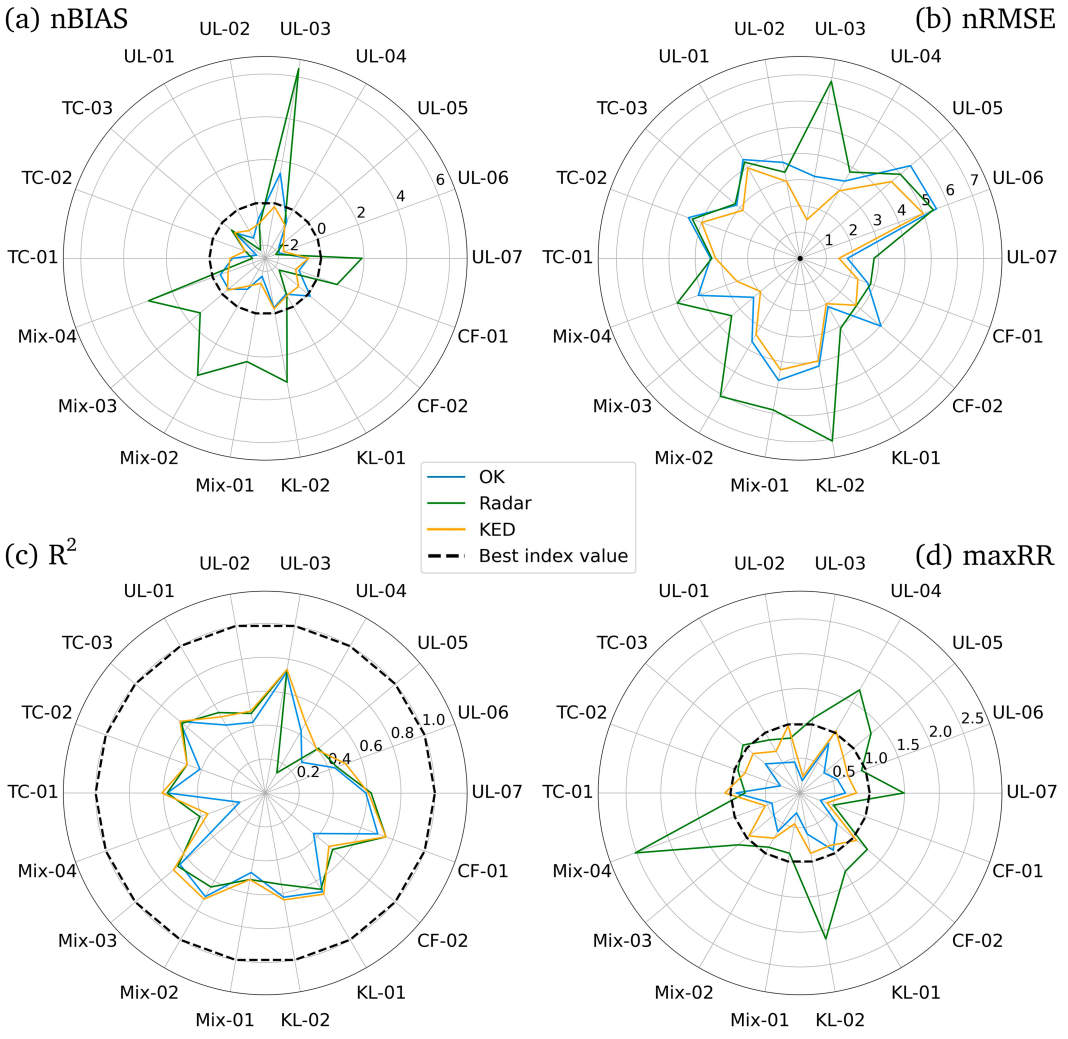

FIG. 3. Results from the cross validation of 18 selected severe storm events between 2016 and 2020 for evaluation (two CFs, seven ULs, two KLs, three TCs, and four Mix). Events are identified by two letters, which indicate the storm type and event ID (see Table 1), shown next to the perimeter. (a) nBIAS, (b) nRMSE, (c) R^2, and (d) MaxRR. Validation indices of each method for each event are shown (i.e., blue for OK, green for radar, orange for KED). Bold black dashed lines or the single point [for nRMSE in (b)] indicate the best possible value for that validation index. For the validation indices of each method, the closer to the black dashed lines or the single points, the better the method performed.

inconsistency of the rainfall estimates in different events. The mean RMSE of OK, radar, and KED ranges from 0.17 to 0.75, from 0.3 to 1.11, and from 0.1 to 0.71 mm, respectively. Based on the mean RMSE, KED outperformed OK and radar rainfall estimates for all events, while OK had better results than radar during 13 events. Figure 3b shows that KED especially reduced errors in two events. One is an upper-level trough (event ID: UL-03), and the other is a mix of an upper-level trough and a kona low (event ID: Mix-04). The event-mean R^2 of OK, radar, and KED ranges between 0.16 and 0.72, between 0.14 and 0.76, and between 0.36 and 0.76, respectively. The hourly rainfall estimates of OK, radar, and KED had the best mean R^2 at 1, 9, and 13 events, respectively. In addition, KED had the highest values in both low and high ranges of R^2, which indicates the consistency in estimates of hourly rainfall. The event MaxRR of OK, radar, and KED ranges from 0.18 to 0.95, from 0.24 to 2.52, and from 0.51 to 1.08, respectively, which indicates KED is the most consistent method at accurately estimating high rainfall values.

b. *Evaluation by storm type*

The validation matrix of different storm types is shown in Table 2. Whereas KED rainfall estimates were not always the best in every storm event, they were the best in every storm type based on mean RMSE with the support of R^2 and AUC (Table 2). The accuracy of each rainfall estimate varied with storm type, but gauge-included interpolation methods (i.e., OK and KED) had similar orders of value for each validation

TABLE 2. The mean validation matrices (i.e., BIAS, RMSE, R^2, and AUC) for different storm types (n is the count of each storm type). Radar indicates radar-only rainfall. Bold font denotes the best performance between OK, radar, and KED for each storm type and validation matrix.

	Mean BIAS (mm)			Mean RMSE (mm) (mean nRMSE; %)			R^2			AUC		
Storm type (n)	OK	Radar	KED	OK	Radar	KED	OK	Radar	KED	OK	Radar	KED
All types (18)	−0.25	**0.01**	−0.18	0.52 (3.75)	0.62 (4.44)	**0.47 (3.17)**	0.52	0.49	**0.59**	0.82	0.68	**0.86**
TC (3)	−0.34	−0.31	**−0.23**	0.67 (3.65)	0.69 (3.64)	**0.63 (3.32)**	0.56	0.50	**0.58**	0.80	0.46	**0.83**
CF (2)	−0.24	**−0.02**	−0.18	0.51 (3.38)	0.49 (2.88)	**0.44 (2.58)**	0.59	0.53	**0.67**	**0.88**	0.65	**0.88**
UL (7)	−0.23	**−0.09**	−0.17	0.44 (4.05)	0.48 (4.57)	**0.39 (3.32)**	0.45	0.49	**0.55**	0.80	0.78	**0.83**
KL (2)	−0.21	0.33	**−0.18**	0.57 (3.18)	0.94 (5.05)	**0.54 (2.99)**	**0.64**	0.48	**0.64**	0.85	0.48	**0.87**
Mix (4)	−0.24	0.30	**−0.17**	0.52 (3.56)	0.69 (5.00)	**0.46 (3.02)**	0.49	0.52	**0.58**	**0.86**	**0.86**	**0.86**

index. This means that the differences in index values likely were driven by the dataset(s) used, not by the kriging method. According to BIAS, most storm events were underestimated, except for radar rainfall in the kona low and mixed upper-level trough and kona low. Based on the nRMSE with R, OK was most accurate with rainfall estimates for kona lows and cold fronts, radar estimates were the most accurate for cold fronts and TCs, and KED had the best rainfall estimates for cold fronts and kona lows. In terms of AUC, both OK and KED had the most accurate results for cold fronts, and radar showed the best performance in the mix of upper-level troughs and kona lows. The gauge-included interpolation methods had more consistent accuracy across all validation indices than radar rainfall. KED had the least accuracy in both TCs and upper-level troughs; however, the validation indices had larger variations in the seven upper-level trough events compared to indices among three TC events (Fig. 3b). Although KED outperformed OK and radar rainfall in all storm types, the differences in the KED validation indices in each event within the same storm type are inconsistent, except for TC events (Fig. 3b).

c. Evaluation by rainfall structure

When the rainfall estimates were separated by convectiveness based on radar reflectivity, KED showed improved performance over OK and radar rainfall estimates in different ways (Table 3). KED had overall better performances in both convective and stratiform rainfall structures, particularly under convective conditions. Under stratiform conditions, radar rainfall performed worst, whereas OK and KED had very similar RMSEs. All methods showed better correlation with convective than stratiform rainfall, indicating that rainfall estimates corresponded more closely with gauge rainfall under convective conditions. KED addressed the issue of lower rainfall estimates by OK and higher rainfall estimates by radar rainfall under convective conditions (Figs. 4a,b). Under stratiform conditions, KED improved the accuracy of both OK and radar rainfall estimates (Figs. 4c,d). Note that only Fig. 4 has R^2 based on the logarithmic hourly rainfall, to be consistent with the scale in the figure; the rest of the indices throughout the paper, including all tables, were based on the hourly rainfall.

TABLE 3. The mean validation matrices (i.e., BIAS, RMSE, and R^2) for different rainfall structures over all hourly rainfall at all locations. Radar indicates radar-only rainfall. Bold font denotes the best performance between OK, radar, and KED for each storm type and each validation matrix.

	BIAS (mm)	RMSE (mm)	R^2
	Convective		
OK	−1.19	2.25	0.45
Radar	**0.62**	3.49	0.49
KED	−0.7	**2.00**	**0.55**
	Stratiform		
OK	−0.23	0.42	0.34
Radar	−0.47	0.95	0.19
KED	**−0.18**	**0.40**	**0.35**

d. Validation summary

Overall, KED outperformed OK and radar rainfall estimates. Radar rainfall estimates exhibited the largest error, and OK underestimated the heaviest rainfall. According to the mean BIAS across all events, the hourly rainfall estimates of OK and KED tended to be underestimated (−0.25 and −0.18 mm), whereas the radar rainfall was overestimated (0.01 mm). In terms of RMSE, KED had the best results based on mean RMSE (0.47 mm) compared to OK (nonsignificant; $p = 0.1762$) and radar (significant; $p = 0.0188$). KED also had the best R^2 over all events. All AUCs were larger than 0.5, which means all interpolated methods are better than random guessing. This also showed that the rainfall estimates of KED (AUC = 0.86) were overall more accurate than the rainfall estimates of OK (AUC = 0.82) and radar (AUC = 0.68) while raining. Regarding the ROC curve, the OK rainfall had a parallel ROC curve below the KED rainfall ROC curve, which indicated that OK rainfall estimates were farther away from the actual gauge rainfall at every threshold than the KED rainfall estimates (Fig. 2). Radar rainfall estimates had a larger FPR compared to OK and KED rainfall estimates, particularly at the thresholds below 14.478 mm, whereas radar rainfall had a larger TPR when the thresholds were equal or above 14.478 mm (Fig. 2). These indicate that radar rainfall estimates are more accurate when the rain is heavier. The radar-only method gave the closest estimated values of maximum rainfall during events, but it often overestimated, with a MaxRR of 1.21. KED also had a close estimation of the maximum, but it often underestimated, with a MaxRR of

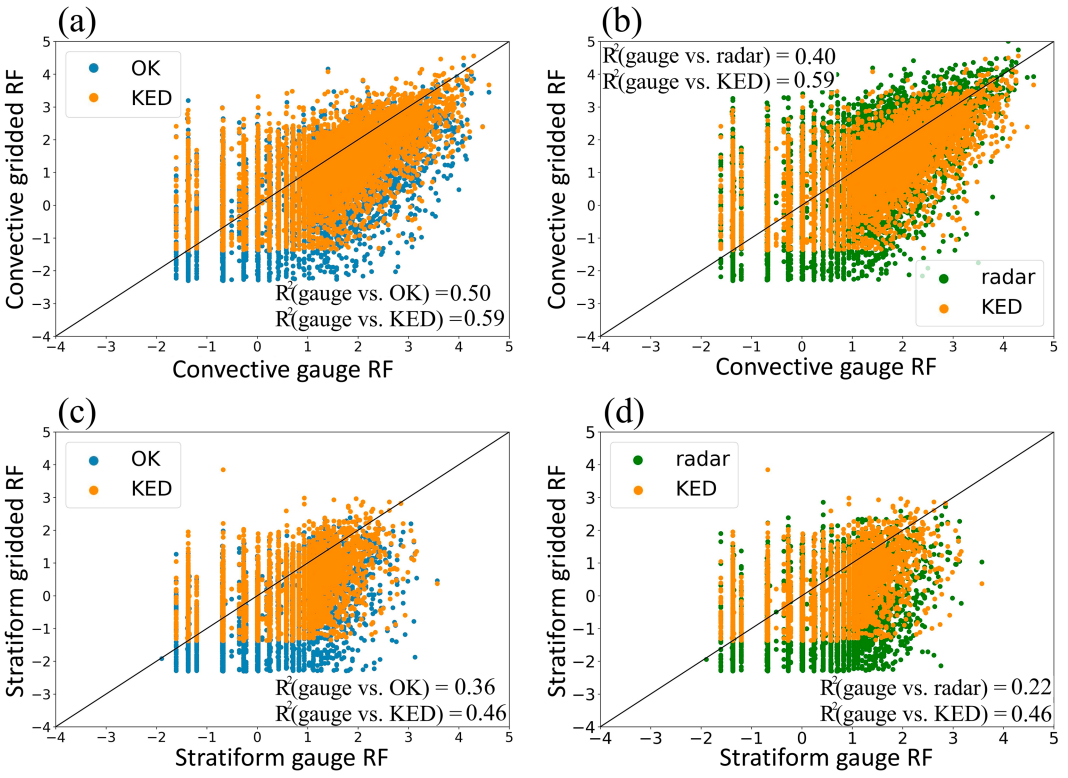

FIG. 4. Scatterplots (x axis: gauge value; y axis: gridded rainfall estimates; both axes in natural log scale) of the results from the leave-one-out cross validation to show R^2 of the (a),(b) convective and (c),(d) stratiform rainfall (RF). Orange, KED; blue, OK; green, radar-only (radar) rainfall estimates; and black lines, 1:1 relationship.

0.78. OK had a MaxRR of 0.55, which indicated OK only captured half of the heaviest rainfall in an event. KED significantly ($p = 0.0013$) outperformed OK in estimating maximum hourly rainfall.

In addition, KED showed improvement in the spatial representation of rainfall on O'ahu (Fig. 1). Compared to radar rainfall, KED reduced the error along the Ko'olau mountain range, particularly south of the ridge. Compared to OK, KED decreased the error along the Ko'olau mountain range on the east side of O'ahu, particularly north of the ridge, where rain gauges are sparse (Fig. 1). This indicates that KED improved the rainfall estimates by providing gauge rainfall (south of Ko'olau mountain range) and filling the ungauged areas (north of Ko'olau mountain range) with radar rainfall. When we plotted an example of the validation results (Fig. 5), it showed that OK could have inaccurate rainfall estimates (~11 mm underestimated) at the omitted rain gauge site (arrows in Figs. 5a,c). OK rainfall also has smoother contours, while radar and KED rainfall inherited more spatial details. Although the uncertainty maps (Fig. A3) showed no pronounced or consistent improvement in uncertainties, the uncertainties decreased in the less-gauged areas by adding radar rainfall [Fig. A3(i)]. KED also adjusted the radar rainfall FPR (Fig. 2) that radar tended to detect rainfall in areas where there is no actual rainfall. For example, radar's heavy rainfall areas extended to one of the rain gauges in the Ko'olau mountain range, which has very low rainfall (red circle in Fig. 5e). This feature was not captured in KED rainfall estimates (red circle in Fig. 5f). Because radar rainfall is not a direct measurement, the rainfall pattern could be shifted. Figure 5h shows spatially inconsistent maximum rainfall to the north of the Wai'anae mountain range between radar and gauge rainfall, and KED (red boxes in Figs. 5h,i) adjusted it. In addition, the radar rainfall in the red box of Fig. 5h would be less accurate, because it is partially based on PHKI radar, which is farther than PHMO (see Fig. A2).

5. Discussion

a. *Advantages of incorporating radar data into rainfall estimates*

The KED-merged rainfall estimates include the features of gauges (OK) and radar. Although both the KED and OK products underestimated rainfall, overall results indicate that KED provides the least error in hourly rainfall estimates and reasonable maximum rainfall values on O'ahu with minimal excessive under- or overestimation (Table 2). While OK underestimated rainfall in this study, radar generally either over- or underestimated rainfall (Table 1). The underestimation of OK may result from the gauge distribution–dependent smoothing effect of geostatistical interpolations in which large values are usually underestimated and small values are usually overestimated

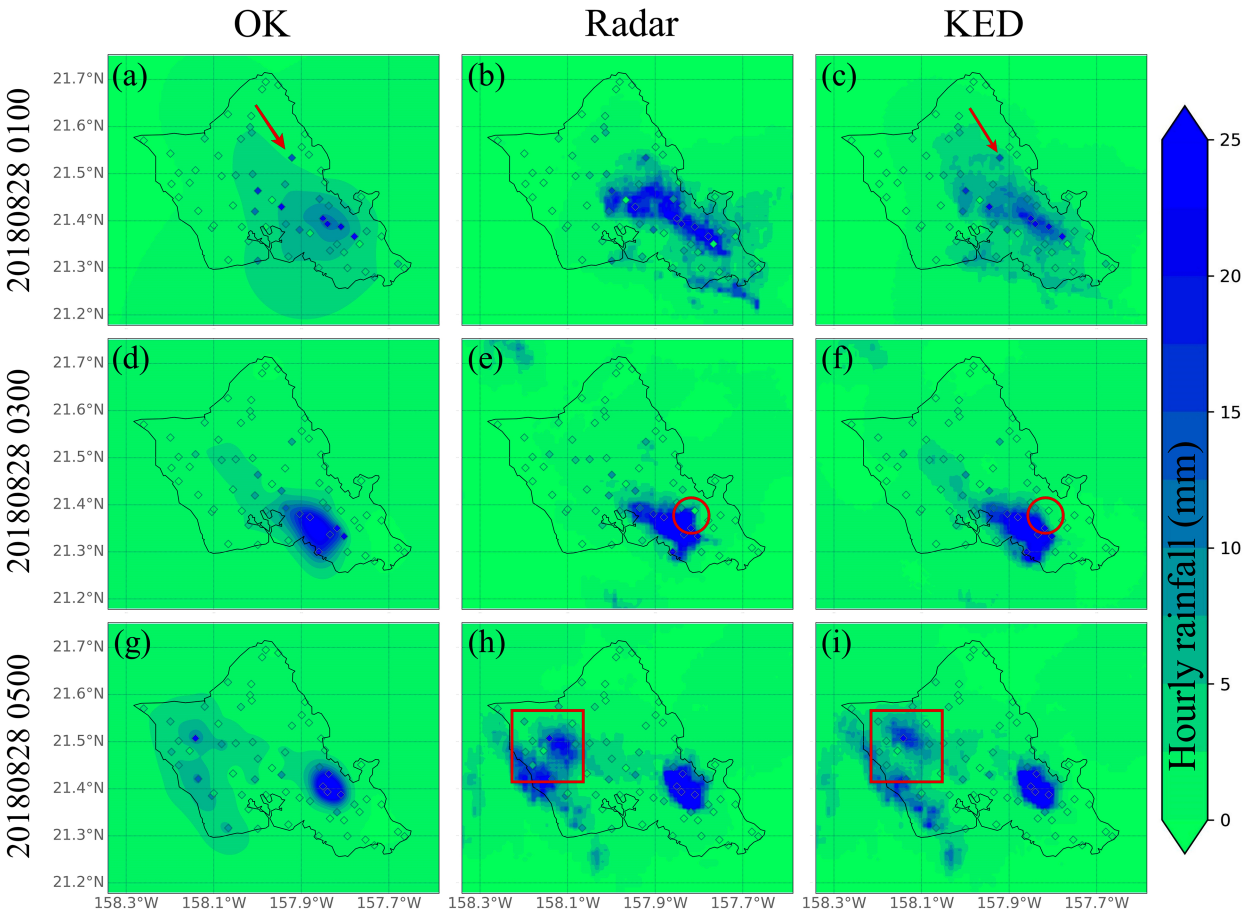

FIG. 5. One example of the leave-one-out hourly rainfall validation, showing each rain gauge (diamond) and different rainfall estimates (shaded; both rain gauge and gridded rainfall estimates use the same color ramp), including (a),(d),(g) OK, (b),(c),(h) radar, and (c),(f),(i) KED, in a 2-h interval for (top) 0100, (middle) 0300, and (bottom) 0500 UTC. In this example, the rain gauge, USGS_uv213215157552800 (red arrows), was left out for OK and KED rainfall estimation for three different hourly periods on 28 Aug 2018, within the TC event TC-02 (see Table 1). The chosen snapshots here were subjected to high hourly rainfall. Red arrows, circles, and boxes mark the areas discussed in the text. The additional figure of uncertainty maps is in Fig. A3.

(Yamamoto 2005; Nikolopoulos et al. 2015). Comparing OK to radar rainfall estimates, OK usually has lower RMSEs; however, the overall underestimation of OK is a major disadvantage when applying the rainfall estimates to flooding studies, which focus on flooding generated by heavy rainfall. In addition, the R^2 of OK is lower than that of radar for some events (Table 1). This shows that OK lacks the capability to capture the dynamics (i.e., rainfall spatial variation in time) of hourly rainfall in ungauged areas (Figs. 5a,b). Radar rainfall estimates usually have larger errors, but they provide spatial estimates in ungauged areas without interpolation smoothing effects and capture dynamic changes well (Table 1). Because the KED-merged rainfall product includes the spatial patterns corrected by the gauge rainfall, the KED product inherits the underestimation from OK. It retains lower errors, better extreme values, and the capacity to capture the rainfall dynamics in ungauged areas from the radar. Furthermore, the KED estimates showed improvements in convective rainfall, which is characterized as heavy rainfall with short duration that often triggers flooding. The validation indices of stratiform rainfall indicate that there are few differences in the hourly rainfall estimates between KED and OK. Thus, when studying time periods with likely stratiform types of rainfall, OK (using gauge data only) could be used to interpolate hourly rainfall estimates to save computational time and resources.

b. *Hidden bias in radar data*

The bias of radar rainfall estimates is likely related to the storm type (Table 2) or rainfall structure (convection vs stratiform rainfall; Fig. 4) because radar tends to detect convective rainfall better than stratiform rainfall (e.g., Fabry et al. 1992). The radar is restricted by low-level rain that is at or below the level of the terrain, or behind the terrain, so radar rainfall estimates may miss the rainfall from the seeder–feeder effect (i.e., orographic precipitation-enhancement mechanism) that may change the rain-rate conversion process [i.e., reflectivity–rain rate (Z–R) relationship]. Therefore, the employed Z–R relationship could be one of the inherited errors of radar rainfall due to the inconsistent Z–R relationship among different

storm types. Additionally, radar can only detect a fraction of the rainfall column beyond a certain range. For example, at a 60-km range, NEXRAD radar can still observe precipitation at altitudes approximately 750 m above the radar (Fig. 3 in Feng et al. 2009). Because PHMO is located about 400 m above sea level, radar rainfall can miss the precipitation below 1250 m, where cloud depths in Hawai'i range from 500 to 2500 m (refer to Figs. 5 and 6 in Rogers et al. 1993) or may even be higher in some of our selected events from deep convections.

c. *Improvement in convective rainfall estimates*

The overall accuracy of the KED hourly rainfall estimates among different storm types (Table 2) indicates that KED can be used to merge radar with gauge data to generate reliable rainfall estimates in various weather systems. However, both OK and KED show distinct differences within each storm type, because there is large variability between events in the same storm type (Fig. 3). On the other hand, large errors, mostly overestimation, in radar rainfall estimates for kona low–related events may be due to the seeder–feeder effect or different raindrop distributions (Blanchard 1953), which may have misled the hydrometeor classification algorithm. More research is needed to understand why radar rainfall underperformed in kona low storms. As a result of both under- and overestimated radar rainfall, the averaged BIAS of overall radar rainfall estimates is low, particularly in front and upper-level trough events. In summary, although there is little consistency between OK and KED validation indices in different storm types and rainfall structures, KED can improve the convective radar rainfall estimates (Fig. 4). Our results suggest that the next step to improve rainfall estimates is to refine radar rainfall estimates in kona low–related events.

d. *Advancement and applications of KED*

The incorporation of radar rainfall has been shown to improve the accuracy of hourly rainfall interpolation in this study. Historically, kriging was applied to daily to yearly rainfall datasets in Hawai'i without incorporating radar information (i.e., OK; Frazier et al. 2016; Mair and Fares 2011). However, kriging alone may be less suitable for interpolating hourly rainfall due to its tendency to be more intermittent, skewed, and anisotropic (Cernesson et al. 1996; Shah et al. 1996). This study demonstrated the feasibility of using KED with both gauge and radar rainfall in Hawai'i. The results of this study show that this approach effectively improved the accuracy of hourly rainfall interpolation, even in the complex and variable weather patterns found in the region. Previous research found that radar rainfall provided little improvement for the annual rainfall interpolation in Hawai'i (Giambelluca et al. 2013). In this study, the radar data were used to address the distribution and structure complexities at the hourly scale. Also, the use of LROSE, which includes clutter detection and mitigation, quality metrics and error assessment, and attenuation correction in precipitation (Hubbert et al. 2009; Gu et al. 2011; Bell et al. 2013), potentially reduced the uncertainty in hourly radar data. Overall, this study highlights the importance of incorporating radar information in the interpolation of hourly rainfall in Hawai'i and the effectiveness of KED in this context, despite the theoretical challenges of interpolating hourly data.

Comparing KED applications to hourly rainfall globally, our KED rainfall estimates for O'ahu work very well. The RMSE of KED on O'ahu, ranges from 0.1 to 0.71 mm h^{-1} based on the LOOCV results over 18 rainfall events. In similar studies, the KED RMSE ranged from 0.3 to 3.5 mm h^{-1} in the United Kingdom (Jewell and Gaussiat 2015; Nanding et al. 2015; Courty et al. 2018) and from 0.5 to 2 mm h^{-1} in Germany (Haberlandt 2007; Rabiei and Haberlandt 2015). Interestingly, in other regions, the radar rainfall was often underestimated, and KED rainfall was often overestimated, including in the United Kingdom (Jewell and Gaussiat 2015; Nanding et al. 2015; Courty et al. 2018), Germany (Haberlandt 2007; Rabiei and Haberlandt 2015), and the contiguous United States (Yu 2009; Skinner et al. 2009). Our results for O'ahu found the opposite; the hourly rainfall was usually overestimated by the radars and always underestimated by KED. Also, in this study, there was no strong relationship between radar rainfall performance and radar ranges in the preliminary investigation (Fig. A4b). Additional work is needed to understand why the results for O'ahu are the opposite of under- and overestimation, respectively, of radar and KED rainfall estimates.

e. *Limitations and future directions*

We showed that KED can be used on O'ahu to estimate rainfall. While this KED application shows encouraging results, this pilot application is not a guarantee of success for other islands because 1) this KED application on O'ahu is less subject to possible beam blockage, while radar data on other islands likely have larger gaps in coverage due to beam blockage (e.g., Kaua'i and Hawai'i Island), or no radar data are available at all (e.g., Sri Lanka, Madagascar, and some Pacific islands; Saltikoff et al. 2019); 2) rain gauges on other islands typically have a coarser density (e.g., Maui and Hawai'i Island) that could limit the application of KED, testing other geostatistical methods and adding additional information (e.g., elevations or anisotropy angles; Verworn and Haberlandt 2011) that could be helpful; 3) while not shown in our study, the effective distances to the radar stations are likely variable, and this distance likely affects the KED estimation and the over- or underestimation of radar rainfall; 4) more advanced and dynamic radar rainfall conversion may be needed for different interactions between the terrain and weather systems on each island; and 5) we only applied two-dimensional KED in this study, and incorporating a third dimension (e.g., the time variable) could potentially improve the rainfall estimation, because in theory it adjusts the spatial mean values by the time dimension (Nowak and Litvinenko 2013). In addition to the constraints of geostatistical areas and methods due to the data availability within the given period, there are only 18 storm events in this study. We could only provide pilot results and discussions on the differences of QPEs performed during different storm types. A much stronger statistical conclusion could be made with a larger sample size.

6. Conclusions

We successfully merged hourly gauge and radar rainfall on O‘ahu, Hawai‘i, using KED. In addition, we confirmed that the hourly rainfall estimates of KED outperformed gauge-only or radar rainfall using 18 studied storm events and that KED can be used to produce gridded hourly rainfall on O‘ahu. The KED hourly rainfall estimates provided a better rainfall dataset for flood study and simulations than OK and radar rainfall estimates, because it remained reasonable and consistent in estimating maximum rainfall values regardless of storm type. The best spatial improvement from KED rainfall estimates occurred along the Ko‘olau mountain range (Fig. 1). Among five storm types, KED had the most accurate rainfall estimate for cold fronts and kona lows. After studying convectiveness in these events, we concluded that KED surpassed OK and radar rainfall estimates under convective rainfall. The KED rainfall product is better than the OK of rain gauge data for studying floods, particularly flash floods caused by local short-duration and intense rainfall events. Under stratiform rainfall conditions, KED notably improved the underestimation from radar rainfall but was only a slight improvement over the estimates of OK rainfall. This study showed the applicable KED-merging method for hourly rainfall on a mountainous tropical island with a dense rain gauge network. Our next steps are to apply the KED-merged radar–gauge rainfall approach to other tropical islands, improve radar rainfall estimates, and explore the use of the KED method or other merging methods that incorporate a third dimension (time) into rainfall estimations.

Acknowledgments. We thank Yu-Fen Huang’s Ph.D. committee members, including Dr. Thomas Giambelluca (Water Resources Research Center), Dr. David Gochis (NCAR), Dr. Tomoaki Miura (University of Hawai‘i at Mānoa), and Dr. Ayron Strauch (Commission on Water Resource Management) for feedback and suggestions. The authors acknowledge support by the USDA National Institute of Food and Agriculture, McIntire Stennis Project 1140 M, managed by the College of Tropical Agriculture and Human Resources, NSF Office of IIA, RII Track- 4: Building the Next Generation Meteo-Hydrological Model for Hawai‘i (OIA-1929155), and U.S. Geological Survey, Pacific Island Climate Adaptation Science Center (G21AC10392).

Data availability statement. The gauge and radar rainfall datasets we used in this study were from Huang et al. (2022). The hourly rainfall dataset in this study, which was created by merging gauge and radar data, will be made available upon request by contacting the corresponding author.

APPENDIX A

Additional Table and Figures

Figure A1 shows an example of the correlation between pairs of gauges as a function of their distance during TC-02 (see Table 1). Figure A2 is the map of radar locations that supply radar rainfall data on O‘ahu and the specific radar used at 1.5-km height at the gauge location. Figure A3 presents uncertainty maps of OK and KED and uncertainty differences between KED and OK of Fig. 5. Figure A4 shows the mean nRMSE of radar, OK, and KED rainfall. Table A1 lists the results from point-to-grid cross validation of 18 selected severe storm events between 2016 and 2020 for validation.

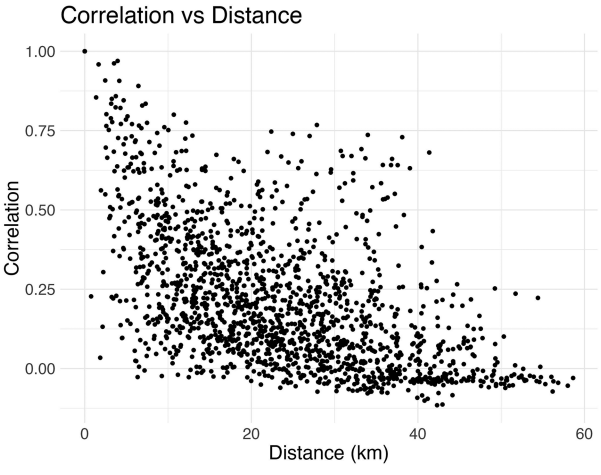

FIG. A1. Example of the correlation between pairs of gauges as a function of their distance during TC-02 (see Table 1).

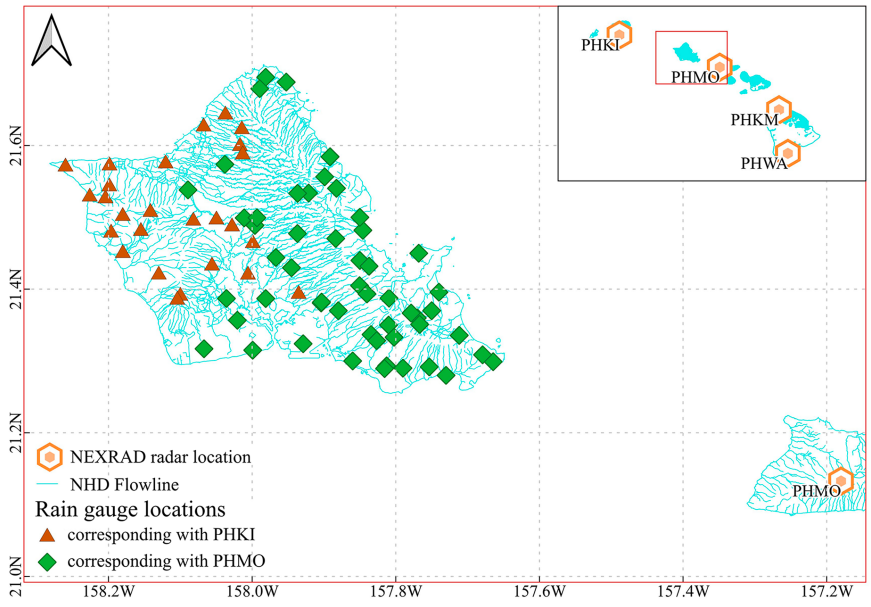

FIG. A2. The map of radar locations that supply radar rainfall data on O'ahu and the specific radar used at 1.5-km height at the gauge location.

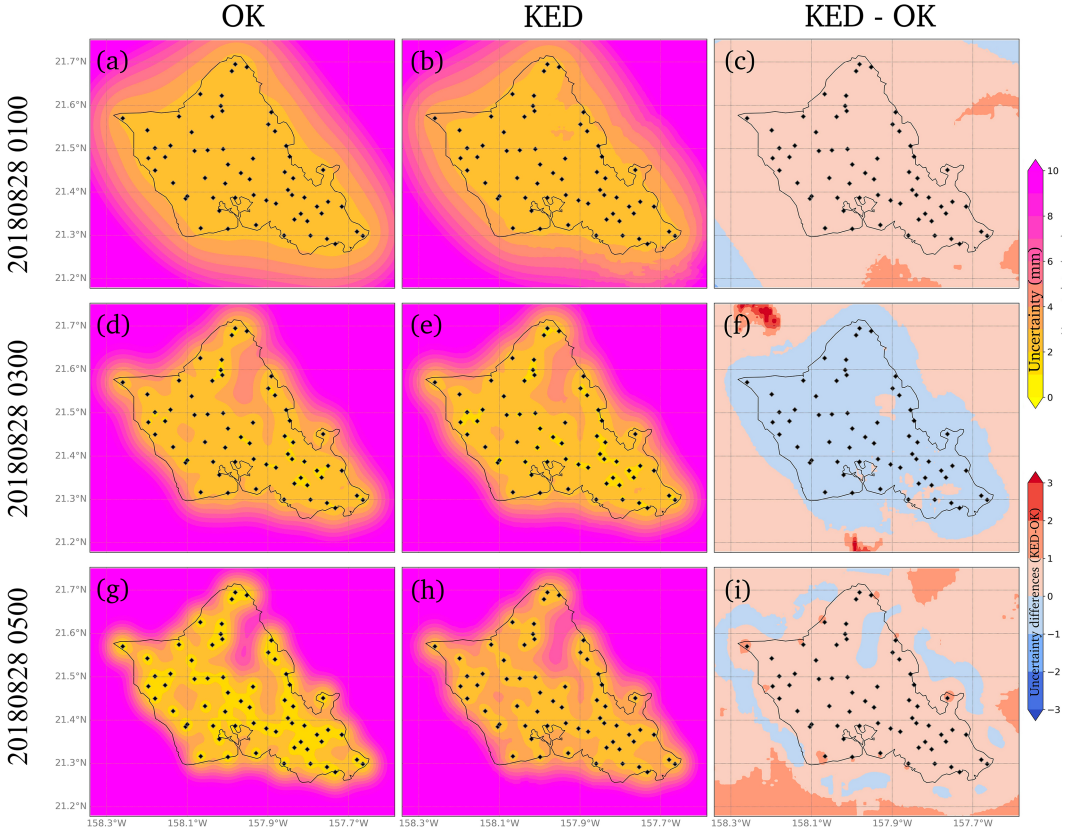

FIG. A3. Uncertainty maps of (a),(d),(g) OK and (b),(e),(h) KED and (c),(f),(i) uncertainty differences between KED and OK of Fig. 5. Black dots denote locations of rain gauges.

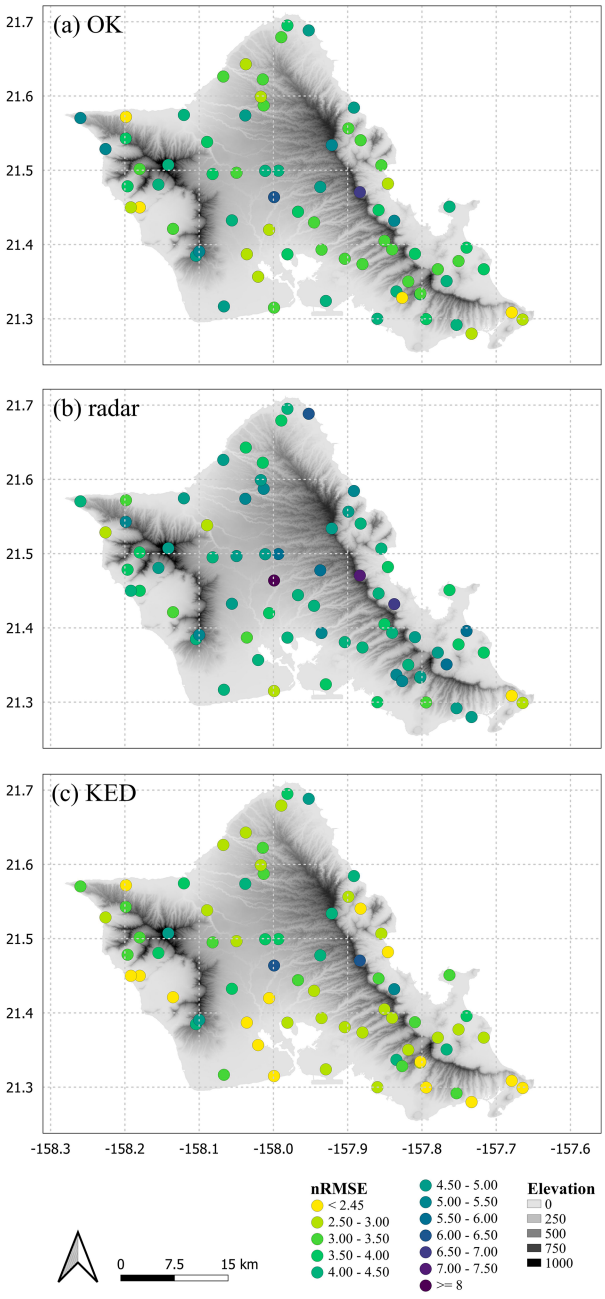

FIG. A4. Mean nRMSE of (a) radar, (b) OK, and (c) KED rainfall. Purple indicates better performance at the gauge locations. To learn more details about each method of rainfall estimates and nRMSE, please refer to sections 3b and 3c.

TABLE A1. Rainfall characteristics for each event across all gauge locations, including mean hourly rainfall over all gauges and the entire event period, standard deviation of hourly rainfall, maximum hourly rainfall, accumulated rainfall divided by the number of gauges, mean convectiveness, standard deviation of convectiveness, dominant anisotropy angle (i.e., the median angle of the best-fit hourly directional variogram; clockwise from the positive y axis) during an event, and the number of angles (i.e., the number of different anisotropic angles shown during an event based on the best-fit hourly directional variogram with 10° increase; in total, there are 18 different angles). Refer to Table 1 for more event information.

Event ID	Mean rainfall (mm h^{-1})	Rainfall std dev (mm h^{-1})	Maximum rainfall (mm h^{-1})	Gauge mean accumulated rainfall (mm h^{-1})	Mean convectiveness	Std dev of convectiveness	Dominant anisotropy angle (°)	No. of angles
TC-01	1.108	5.004	91.694	107.385	0.469	0.417	30	12
CF-01	0.848	3.291	59.182	81.818	0.408	0.393	60	7
Mix-01	0.956	3.353	62.2	81.001	0.463	0.412	50	10
Mix-02	0.718	2.225	35.56	52.371	0.401	0.386	80	6
UL-01	0.709	2.425	36.322	51.735	0.456	0.434	160	10
Mix-03	0.827	4.350	99.06	80.022	0.493	0.445	0	10
UL-02	0.497	2.507	47.244	47.949	0.223	0.376	20	13
UL-03	0.165	1.445	54.61	11.931	0.200	0.359	0	4
UL-04	0.541	2.952	87.376	65.232	0.336	0.416	150	11
TC-02	0.840	2.749	63.5	101.172	0.328	0.409	150	15
TC-03	1.149	4.010	60.706	111.252	0.260	0.387	150	12
UL-05	0.511	2.222	58.674	37.135	0.292	0.392	50	12
KL-01	0.933	3.837	67.564	90.384	0.287	0.376	10	8
CF-02	0.646	3.130	49.784	62.425	0.430	0.454	130	13
UL-06	0.875	2.468	39.624	84.263	0.485	0.414	10	16
KL-02	1.046	3.342	56.6	100.725	0.273	0.389	30	12
Mix-04	0.486	2.773	57.912	46.271	0.557	0.444	0	10
UL-07	0.509	3.251	60.452	48.971	0.306	0.407	50	5

APPENDIX B

Glossary

a. Glossary of acronyms and abbreviations in alphabetical order

AMS	American Meteorological Society
AUC	Area under the ROC curve
BDC	Bayesian data combination
CoK	Cokriging
KED	Kriging with external drift
KRE	Ordinary kriging with radar-based error correction
LOOCV	Leave-one-out cross validation
LROSE	Lidar Radar Open Software Environment
NCEI	National Centers for Environmental Information
NEXRAD	Next Generation Weather Radar
MFE	Mean field bias
MM5	Fifth-generation Pennsylvania State University–NCAR Mesoscale Model
NCAR	National Center for Atmospheric Research
NHD	National Hydrography Dataset
NOAA	National Oceanic and Atmospheric Administration
OK	Ordinary kriging
PHKI	The NEXRAD radar at Kaua'i
PHMO	The NEXRAD radar at Moloka'i
PRISM	Parameter-Elevation Regressions on Independent Slopes Model
QPE	Quantitative precipitation estimation
ROC	Receiver operating characteristic
TC	Tropical cyclone
USGS	U.S. Geological Survey

b. Glossary of mathematical symbols in the order of appearance

BIAS	Bias
RMSE	Root-mean-square error
R^2	Coefficient of determination
MaxRR	Maximum rainfall ratio
RF_{method}	Rainfall from the method. The method can be gauge rainfall (G), OK, radar, KED, or, in general, any rainfall estimate (EST)
C	Covariance
λ^{method}	Weights for kriging processes. The method can be OK or KED.
TPR	True positive rate
TP	Count of true positive
FN	Count of false negative
FPR	False positive rate
FP	Count of false positive
TN	Count of true negative

REFERENCES

American Meteorological Society, 2023: "Rain." Glossary of Meteorology, https://glossary.ametsoc.org/wiki/Rain.

Bell, M. M., W.-C. Lee, C. A. Wolff, and H. Cai, 2013: A solo-based automated quality control algorithm for airborne tail Doppler radar data. *J. Appl. Meteor. Climatol.*, **52**, 2509–2528, https://doi.org/10.1175/JAMC-D-12-0283.1.

Berndt, C., E. Rabiei, and U. Haberlandt, 2014: Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios. *J. Hydrol.*, **508**, 88–101, https://doi.org/10.1016/j.jhydrol.2013.10.028.

Berne, A., and W. F. Krajewski, 2013: Radar for hydrology: Unfulfilled promise or unrecognized potential? *Adv. Water Resour.*, **51**, 357–366, https://doi.org/10.1016/j.advwatres.2012.05.005.

Blanchard, D. C., 1953: Raindrop size-distribution in Hawaiian rains. *J. Meteor.*, **10**, 457–473, https://doi.org/10.1175/1520-0469(1953)010<0457:RSDIHR>2.0.CO;2.

Borga, M., 2002: Accuracy of radar rainfall estimates for stream-flow simulation. *J. Hydrol.*, **267**, 26–39, https://doi.org/10.1016/S0022-1694(02)00137-3.

Brandes, E. A., 1975: Optimizing rainfall estimates with the aid of radar. *J. Appl. Meteor.*, **14**, 1339–1345, https://doi.org/10.1175/1520-0450(1975)014<1339:OREWTA>2.0.CO;2.

Cecinati, F., O. Wani, and M. A. Rico-Ramirez, 2017: Comparing approaches to deal with non-Gaussianity of rainfall data in kriging-based radar–gauge rainfall merging. *Water Resour. Res.*, **53**, 8999–9018, https://doi.org/10.1002/2016WR020330.

Cernesson, F., J. Lavabre, and J.-M. Masson, 1996: Stochastic model for generating hourly hyetographs. *Atmos. Res.*, **42**, 149–161, https://doi.org/10.1016/0169-8095(95)00060-7.

Courty, L. G., M. Á. Rico-Ramirez, and A. Pedrozo-Acuña, 2018: The significance of the spatial variability of rainfall on the numerical simulation of urban floods. *Water*, **10**, 207, https://doi.org/10.3390/w10020207.

Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical–topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140–158, https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2.

Delrieu, G., A. Wijbrans, B. Boudevillain, D. Faure, L. Bonnifait, and P. E. Kirstetter, 2014: Geostatistical radar–raingauge merging: A novel method for the quantification of rain estimation accuracy. *Adv. Water Resour.*, **71**, 110–124, https://doi.org/10.1016/j.advwatres.2014.06.005.

Dixon, M. J., and B. Javornik, 2016: LROSE—Lidar radar open software environment. NCAR, https://doi.org/10.5065/60HZ-RY38.

——, and U. Romatschke, 2022: Three-dimensional convective–stratiform echo-type classification and convectivity retrieval from radar reflectivity. *J. Atmos. Oceanic Technol.*, **39**, 1685–1704, https://doi.org/10.1175/JTECH-D-22-0018.1.

——, J. Wilson, T. Weckwerth, D. Albo, and E. Thompson, 2015: A dual-polarization QPE method based on the NCAR particle ID algorithm—Description and preliminary results. *37th Conf. on Radar Meteorology*, Norman, OK, Amer. Meteor. Soc., 9A.1, https://ams.confex.com/ams/37RADAR/webprogram/Paper275705.html.

Dudhia, J., 1993: A nonhydrostatic version of the Penn State–NCAR Mesoscale Model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.*, **121**, 1493–1513, https://doi.org/10.1175/1520-0493(1993)121<1493:ANVOTP>2.0.CO;2.

El-Kadi, A. I., and E. Yamashita, 2007: Modeling streamflows and flood delineation of the 2004 flood disaster, Mānoa, O'ahu, Hawai'i. *Pac. Sci.*, **61**, 235–255, https://doi.org/10.2984/1534-6188(2007)61[235:MSAFDO]2.0.CO;2.

Erdin, R., 2009: Combining rain gauge and radar measurements of a heavy precipitation event over Switzerland: Comparison of geostatistical methods and investigation of important influencing factors. Bundesamt für Meteorologie und Klimatologie, MeteoSwiss Publ. 81, 114 pp., https://www.meteoswiss.admin.ch/dam/jcr:5625315d-407c-48f5-9920-dfaa1704bb31/veroeff81.pdf.

Essou, G. R. C., F. Sabarly, P. Lucas-Picher, F. Brissette, and A. Poulin, 2016: Can precipitation and temperature from meteorological reanalyses be used for hydrological modeling? *J. Hydrometeor.*, **17**, 1929–1950, https://doi.org/10.1175/JHM-D-15-0138.1.

Fabry, F., G. L. Austin, and D. Tees, 1992: The accuracy of rainfall estimates by radar as a function of range. *Quart. J. Roy. Meteor. Soc.*, **118**, 435–453, https://doi.org/10.1002/qj.49711850503.

Fares, A., R. Awal, J. Michaud, P.-S. Chu, S. Fares, K. Kodama, and M. Rosener, 2014: Rainfall-runoff modeling in a flashy tropical watershed using the distributed HL-RDHM model. *J. Hydrol.*, **519**, 3436–3447, https://doi.org/10.1016/j.jhydrol.2014.09.042.

Fawcett, T., 2006: An introduction to ROC analysis. *Pattern Recognit. Lett.*, **27**, 861–874, https://doi.org/10.1016/j.patrec.2005.10.010.

Feng, Z., X. Dong, and B. Xi, 2009: A method to merge WSR-88D data with ARM SGP millimeter cloud radar data by studying deep convective systems. *J. Atmos. Oceanic Technol.*, **26**, 958–971, https://doi.org/10.1175/2008JTECHA1190.1.

Frazier, A. G., and T. W. Giambelluca, 2017: Spatial trend analysis of Hawaiian rainfall from 1920 to 2012. *Int. J. Climatol.*, **37**, 2522–2531, https://doi.org/10.1002/joc.4862.

——, ——, H. F. Diaz, and H. L. Needham, 2016: Comparison of geostatistical approaches to spatially interpolate month–year rainfall for the Hawaiian Islands. *Int. J. Climatol.*, **36**, 1459–1470, https://doi.org/10.1002/joc.4437.

Giambelluca, T. W., Q. Chen, A. G. Frazier, J. P. Price, Y.-L. Chen, P.-S. Chu, J. K. Eischeid, and D. M. Delparte, 2013: Online rainfall atlas of Hawai'i. *Bull. Amer. Meteor. Soc.*, **94**, 313–316, https://doi.org/10.1175/BAMS-D-11-00228.1.

Goovaerts, P., 1997: *Geostatistics for Natural Resources Evaluation*. Oxford University Press, 483 pp.

——, 2000: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.*, **228**, 113–129, https://doi.org/10.1016/S0022-1694(00)00144-X.

Goudenhoofdt, E., and L. Delobbe, 2009: Evaluation of radar–gauge merging methods for quantitative precipitation estimates. *Hydrol. Earth Syst. Sci.*, **13**, 195–203, https://doi.org/10.5194/hess-13-195-2009.

Gu, J.-Y., A. Ryzhkov, P. Zhang, P. Neilley, M. Knight, B. Wolf, and D.-I. Lee, 2011: Polarimetric attenuation correction in heavy rain at C band. *J. Appl. Meteor. Climatol.*, **50**, 39–58, https://doi.org/10.1175/2010JAMC2258.1.

Haberlandt, U., 2007: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *J. Hydrol.*, **332**, 144–157, https://doi.org/10.1016/j.jhydrol.2006.06.028.

Hitschfeld, W., and J. Bordan, 1954: Errors inherent in the radar measurement of rainfall at attenuating wavelengths. *J. Meteor.*, **11**, 58–67, https://doi.org/10.1175/1520-0469(1954)011<0058:EIITRM>2.0.CO;2.

Huang, Y.-F., M. Gayte, Y. Tsang, R. J. Longman, A. D. Nugent, K. Kodama, M. P. Lucas, and T. W. Giambelluca, 2022:

Hourly rainfall data from rain gauge networks and weather radar up to 2020 across the Hawaiian Islands. *Sci. Data*, **9**, 334, https://doi.org/10.1038/s41597-022-01430-2.

Hubbert, J. C., M. Dixon, and S. M. Ellis, 2009: Weather radar ground clutter. Part II: Real-time identification and filtering. *J. Atmos. Oceanic Technol.*, **26**, 1181–1197, https://doi.org/10.1175/2009JTECHA1160.1.

Jewell, S. A., and N. Gaussiat, 2015: An assessment of kriging-based rain-gauge–radar merging techniques. *Quart. J. Roy. Meteor. Soc.*, **141**, 2300–2313, https://doi.org/10.1002/qj.2522.

Kirstetter, P.-E., H. Andrieu, B. Boudevillain, and G. Delrieu, 2013: A physically based identification of vertical profiles of reflectivity from volume scan radar data. *J. Appl. Meteor. Climatol.*, **52**, 1645–1663, https://doi.org/10.1175/JAMC-D-12-0228.1.

Kitanidis, P. K., 1997: *Introduction to Geostatistics: Applications in Hydrogeology*. Cambridge University Press, 276 pp.

Kodama, K., and G. M. Barnes, 1997: Heavy rain events over the south-facing slopes of Hawaii: Attendant conditions. *Wea. Forecasting*, **12**, 347–367, https://doi.org/10.1175/1520-0434(1997)012<0347:HREOTS>2.0.CO;2.

Krajewski, W. F., 1987: Cokriging radar-rainfall and rain gage data. *J. Geophys. Res.*, **92**, 9571–9580, https://doi.org/10.1029/JD092iD08p09571.

Longman, R. J., and Coauthors, 2019: High-resolution gridded daily rainfall and temperature for the Hawaiian Islands (1990–2014). *J. Hydrometeor.*, **20**, 489–508, https://doi.org/10.1175/JHM-D-18-0112.1.

——, O. E. Timm, T. W. Giambelluca, and L. Kaiser, 2021: A 20-year analysis of disturbance-driven rainfall on O'ahu, Hawai'i. *Mon. Wea. Rev.*, **149**, 1767–1783, https://doi.org/10.1175/MWR-D-20-0287.1.

Ly, S., C. Charles, and A. Degré, 2013: Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale: A review. *Biotechnol. Agron. Soc. Environ.*, **17**, 392–406, https://doi.org/10.6084/m9.figshare.1225842.v1.

Ma, S., and J. Huang, 2007: Combining multiple markers for classification using ROC. *Biometrics*, **63**, 751–757, https://doi.org/10.1111/j.1541-0420.2006.00731.x.

Mair, A., and A. Fares, 2011: Comparison of rainfall interpolation methods in a mountainous region of a tropical island. *J. Hydrol. Eng.*, **16**, 371–383, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000330.

McKee, J. L., and A. D. Binns, 2016: A review of gauge–radar merging methods for quantitative precipitation estimation in hydrology. *Can. Water Resour. J.*, **41**, 186–203, https://doi.org/10.1080/07011784.2015.1064786.

McMillan, H., B. Jackson, M. Clark, D. Kavetski, and R. Woods, 2011: Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. *J. Hydrol.*, **400**, 83–94, https://doi.org/10.1016/j.jhydrol.2011.01.026.

Michelson, D. B., and J. Koistinen, 2000: Gauge–radar network adjustment for the Baltic Sea experiment. *Phys. Chem. Earth*, **25B**, 915–920, https://doi.org/10.1016/S1464-1909(00)00125-8.

Murphy, B., S. Müller, and R. Yurchak, 2021: GeoStat-Framework/PyKrige: v1.6.1. Zenodo, https://doi.org/10.5281/zenodo.5380342.

Nanding, N., M. A. Rico-Ramirez, and D. Han, 2015: Comparison of different radar–raingauge rainfall merging techniques. *J. Hydroinform.*, **17**, 422–445, https://doi.org/10.2166/hydro.2015.001.

National Weather Service Radar Operations Center, 1991: NOAA Next Generation Radar (NEXRAD) level 2 base data. NOAA, accessed 23 January 2021, https://doi.org/10.7289/V5W9574V.

Newman, A. J., M. P. Clark, R. J. Longman, E. Gilleland, T. W. Giambelluca, and J. R. Arnold, 2019: Use of daily station observations to produce high-resolution gridded probabilistic precipitation and temperature time series for the Hawaiian Islands. *J. Hydrometeor.*, **20**, 509–529, https://doi.org/10.1175/JHM-D-18-0113.1.

Nikolopoulos, E. I., M. Borga, J. D. Creutin, and F. Marra, 2015: Estimation of debris flow triggering rainfall: Influence of rain gauge density and interpolation methods. *Geomorphology*, **243**, 40–50, https://doi.org/10.1016/j.geomorph.2015.04.028.

Nowak, W., and A. Litvinenko, 2013: Kriging and spatial design accelerated by orders of magnitude: Combining low-rank covariance approximations with FFT-techniques. *Math. Geosci.*, **45**, 411–435, https://doi.org/10.1007/s11004-013-9453-6.

Ochoa-Rodriguez, S., L.-P. Wang, P. Willems, and C. Onof, 2019: A review of radar–rain gauge data merging methods and their potential for urban hydrological applications. *Water Resour. Res.*, **55**, 6356–6391, https://doi.org/10.1029/2018WR023332.

Rabiei, E., and U. Haberlandt, 2015: Applying bias correction for merging rain gauge and radar data. *J. Hydrol.*, **522**, 544–557, https://doi.org/10.1016/j.jhydrol.2015.01.020.

Rogers, R. R., W. L. Ecklund, D. A. Carter, K. S. Gage, and S. A. Ethier, 1993: Research applications of a boundary-layer wind profiler. *Bull. Amer. Meteor. Soc.*, **74**, 567–580, https://doi.org/10.1175/1520-0477(1993)074<0567:RAOABL>2.0.CO;2.

Saltikoff, E., and Coauthors, 2019: An overview of using weather radar for climatological studies successes, challenges, and potential. *Bull. Amer. Meteor. Soc.*, **100**, 1739–1752, https://doi.org/10.1175/BAMS-D-18-0166.1.

Schuurmans, J. M., M. F. P. Bierkens, E. J. Pebesma, and R. Uijlenhoet, 2007: Automatic prediction of high-resolution daily rainfall fields for multiple extents: The potential of operational radar. *J. Hydrometeor.*, **8**, 1204–1224, https://doi.org/10.1175/2007JHM792.1.

Shah, S. M. S., P. E. O'Connell, and J. R. M. Hosking, 1996: Modelling the effects of spatial variability in rainfall on catchment response. 1. Formulation and calibration of a stochastic rainfall field model. *J. Hydrol.*, **175**, 67–88, https://doi.org/10.1016/S0022-1694(96)80006-0.

Sideris, I. V., M. Gabella, R. Erdin, and U. Germann, 2014: Real-time radar–rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland. *Quart. J. Roy. Meteor. Soc.*, **140**, 1097–1111, https://doi.org/10.1002/qj.2188.

Simpson, R. H., 1952: Evolution of the Kona storm a subtropical cyclone. *J. Meteor.*, **9**, 24–35, https://doi.org/10.1175/1520-0469(1952)009<0024:EOTKSA>2.0.CO;2.

Sinclair, S., and G. Pegram, 2005: Combining radar and rain gauge rainfall estimates using conditional merging. *Atmos. Sci. Lett.*, **6**, 19–22, https://doi.org/10.1002/asl.85.

Skinner, C., F. Bloetscher, and C. S. Pathak, 2009: Comparison of NEXRAD and rain gauge precipitation measurements in south Florida. *J. Hydrol. Eng.*, **14**, 248–260, https://doi.org/10.1061/(ASCE)1084-0699(2009)14:3(248).

Steiner, M., R. A. Houze Jr., and S. E. Yuter, 1995: Climatological characterization of three-dimensional storm structure from operational radar and rain gauge data. *J. Appl. Meteor.*, **34**, 1978–2007, https://doi.org/10.1175/1520-0450(1995)034<1978:CCOTDS>2.0.CO;2.

Taylor, R., 1990: Interpretation of the correlation coefficient: A basic review. *J. Diagn. Med. Sonogr.*, **6**, 35–39, https://doi.org/10.1177/875647939000600106.

Todini, E., 2001: A Bayesian technique for conditioning radar precipitation estimates to rain-gauge measurements. *Hydrol. Earth Syst. Sci.*, **5**, 187–199, https://doi.org/10.5194/hess-5-187-2001.

Tokay, A., and D. A. Short, 1996: Evidence from tropical raindrop spectra of the origin of rain from stratiform versus convective clouds. *J. Appl. Meteor.*, **35**, 355–371, https://doi.org/10.1175/1520-0450(1996)035<0355:EFTRSO>2.0.CO;2.

Tukey, J. W., 1949: Comparing individual means in the analysis of variance. *Biometrics*, **5**, 99–114, https://doi.org/10.2307/3001913.

Velasco-Forero, C. A., D. Sempere-Torres, E. F. Cassiraga, and J. Jaime Gómez-Hernández, 2009: A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data. *Adv. Water Resour.*, **32**, 986–1002, https://doi.org/10.1016/j.advwatres.2008.10.004.

Verworn, A., and U. Haberlandt, 2011: Spatial interpolation of hourly rainfall—Effect of additional information, variogram inference and storm properties. *Hydrol. Earth Syst. Sci.*, **15**, 569–584, https://doi.org/10.5194/hess-15-569-2011.

Willmott, C. J., 1982: Some comments on the evaluation of model performance. *Bull. Amer. Meteor. Soc.*, **63**, 1309–1313, https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2.

Wilson, J. W., 1970: Integration of radar and raingage data for improved rainfall measurement. *J. Appl. Meteor.*, **9**, 489–497, https://doi.org/10.1175/1520-0450(1970)009<0489:IORARD>2.0.CO;2.

——, and E. A. Brandes, 1979: Radar measurement of rainfall—A summary. *Bull. Amer. Meteor. Soc.*, **60**, 1048–1060, https://doi.org/10.1175/1520-0477(1979)060<1048:RMORS>2.0.CO;2.

Yamamoto, J. K., 2005: Correcting the smoothing effect of ordinary kriging estimates. *Math. Geol.*, **37**, 69–94, https://doi.org/10.1007/s11004-005-8748-7.

Yang, Y., and Y.-L. Chen, 2008: Effects of terrain heights and sizes on island-scale circulations and rainfall for the island of Hawaii during HaRP. *Mon. Wea. Rev.*, **136**, 120–146, https://doi.org/10.1175/2007MWR1984.1.

Yu, B., 2009: Improving the quality of NEXRAD products in terms of resolution and accuracy. M.S. thesis, Dept. of Earth & Environmental Science, The University of Texas at San Antonio, 103 pp., https://www.proquest.com/docview/305159046/abstract/B9AFBBCE4FA240C7PQ/1.

Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–638, https://doi.org/10.1175/BAMS-D-14-00174.1.

Zhu, D., D. Z. Peng, and I. D. Cluckie, 2013: Statistical analysis of error propagation from radar rainfall to hydrological models. *Hydrol. Earth Syst. Sci.*, **17**, 1445–1453, https://doi.org/10.5194/hess-17-1445-2013.