WATER RESOURCES RESEARCH, VOL. 49, 7144–7149, doi:10.1002/wrcr.20498, 2013

Extended contingency table: Performance metrics for satellite observations and climate model simulations

A. AghaKouchak[1] and A. Mehran[1]

Received 21 May 2013; revised 29 July 2013; accepted 19 August 2013; published 7 October 2013.

[1] Validation of gridded satellite observations and climate model simulations are fundamental to future improvements in retrieval algorithms and model developments. Among the metrics, the *contingency table*, which includes a number of categorical indices, is extensively used in evaluation studies. While the categorical indices offer invaluable information, they do not provide any insight into the volume of the variable detected correctly/incorrectly. In this study, the contingency table categorical metrics are extended to volumetric indices for evaluation of gridded data. The suggested indices include (a) Volumetric Hit Index (*VHI*): volume of correctly detected simulations relative to the volume of the correctly detected simulations and missed observations; (b) Volumetric False Alarm Ratio (*VFAR*): volume of false simulations relative to the sum of simulations; (c) Volumetric Miss Index (*VMI*): volume of missed observations relative to the sum of missed observations and correctly detected simulations; and (d) the Volumetric Critical Success Index (*VCSI*). The latter provides an overall measure of volumetric performance including volumetric hits, false alarms, and misses. First, using two synthetic time series, the volumetric indices are evaluated against the contingency table categorical metrics. Then, the volumetric indices are used to evaluate a gridded data set at the continental scale. The results show that the volumetric indices provide additional information beyond the commonly used categorical metrics that can be useful in evaluating gridded data sets.

Citation: AghaKouchak, A., and A. Mehran (2013), Extended contingency table: Performance metrics for satellite observations and climate model simulations, *Water Resour. Res.*, *49*, 7144–7149, doi:10.1002/wrcr.20498.

1. Introduction

[2] Remotely sensed radar and satellite observations and climate model simulations are subject to uncertainties and biases arising from physical and algorithmic aspects. Evaluation and uncertainty quantification of remotely sensed data and climate model simulations are fundamental to scientific advancements, algorithm/model developments, and integration of data into applications. For this reason, numerous studies are devoted to evaluation of remote sensing data [e.g., *Anagnostou et al*., 1998; *AghaKouchak et al*., 2010a; *Turk et al*., 2008; *Norouzi et al*., 2011; *Jackson et al*., 2005; *Pinker et al*., 2009; *Dorigo et al*., 2010; *AghaKouchak et al*., 2010b; *Mehran and AghaKouchak*, 2013], and climate model simulations [e.g., *Phillips and Gleckler*, 2006; *Jiang et al*., 2012; *Feddema et al*., 2005; *Liepert and Previdi*, 2012] versus gridded ground-based observations.

[3] Recently, several efforts are devoted to development of metrics and tools for validation of weather and climate models as well as satellite observations. *Gleckler et al*. [2008] suggested several performance metrics for validation of historical climate model simulations. *Gilleland* [2013] proposed the spatial prediction comparison test for evaluation of precipitation forecasts. *AghaKouchak et al*. [2011b] developed several indices for evaluation of high quantiles of satellite precipitation observations. *Entekhabi et al*. [2010] introduced a number of metrics for evaluation of remotely sensed soil moisture observations. *Hossain and Huffman* [2008] recommended a set of spatial, retrieval, and temporal error metrics for satellite data sets that can advance hydrologic applications. *Gebremichael* [2010] outlined a framework for validating satellite data sets using ground-based observations. A number of geometrical and object-oriented metrics are also proposed for spatial validation and verification (e.g., *Davis et al*., 2009; *AghaKouchak et al*., 2011a; *Brown et al*., 2004].

[4] Among the metrics, the *contingency table* [*Wilks*, 2006] which includes a number of categorical indices is extensively used in evaluation studies [e.g., *Behrangi et al*., 2011; *Haile et al*., 2012; *Hao et al.*, 2013; *Gourley et al*., 2012; *Hirpa et al*., 2010; *Anagnostou et al*., 2010]. The contingency table is used to analyze or validate the relationship between two categorical variables and is the categorical equivalent of the scatterplot. The contingency table metrics describe whether simulations or remote sensing observations (hereafter, *SIM*) *hit* or *miss* the reference observations (hereafter, *OBS*) and/or lead to *false* estimates

[1]Department of Civil and Environmental Engineering, University of California Irvine, Irvine, California, USA.

Corresponding author: A. AghaKouchak, University of California Irvine, E4130 Engineering Gateway, Irvine, CA 92697-2175, USA. (amir.a@uci.edu)

0043-1397/13/10.1002/wrcr.20498

relative to *OBS*. While the contingency table metrics offer invaluable information, they do not provide any insight into biases and errors in the magnitude of *SIM* relative to *OBS*. Hence, errors and biases should be evaluated using additional metrics such as the unbiased root mean square error [*Entekhabi et al*., 2010], quantile bias [*AghaKouchak et al*., 2011b], and relative error [*Gleckler et al*., 2008]. In this study, the commonly used categorical metrics are extended to volumetric measure such that one can investigate both the categorical *hit*, *miss*, *false*, and their corresponding volumetric errors. The main purpose of the suggested indices is to decompose the total bias into volumetric errors terms associated with *hit*, *miss*, and *false* components.

[5] This technical note is organized into three sections. After the introduction, the commonly used categorical indices are reviewed briefly in section 2. The proposed volumetric indices are then described followed by an example application. The last section summarizes the conclusions and final remarks.

2. Methodology and Results

[6] The most common form of the contingency table is 2 × 2, which is used to evaluate dichotomous variables (see Figure 1). In this table, hit (*H*) indicates that both reference observation and simulation detect the event, whereas miss (*M*) refers to events identified by reference observation but missed by the simulation. False (*F*), also known as false alarm, represents events identified by the simulation but not confirmed by observations. Based on the contingency table, several metrics are defined as follows [*Wilks*, 2006]:

[7] 1. The Probability of Detection (*POD*) describes the fraction of the reference observations detected correctly by the simulation: $POD = H/(H+M)$. The *POD* ranges from 0 to 1; 0 indicates no skill and 1 indicates perfect score.

[8] 2. The False Alarm Ratio (FAR) corresponds to the fraction of events identified by simulation but not confirmed by reference observations: $FAR = F/(H+F)$. The *FAR* ranges from 0 to 1; 0 indicates perfect score.

[9] 3. The Critical Success Index (*CSI*), also known as the Threat Score, combines different aspects of the *POD* and *FAR*, describing the overall skill of the simulation relative to reference observation: $CSI = H/(H+M+F)$. The *CSI* ranges from 0 to 1; 0 indicates no skill and 1 indicates perfect skill.

[10] The original contingency table metrics provide categorical measures of performance. For example, a *POD* of 0.8 indicates that the simulation detects 80% of events (e.g., precipitation events). However, it does not provide any information as to what fraction of the volume of precipitation is detected. For most climate variables one may need to go beyond the *POD* and estimate the volume of the variable of interest detected correctly. For this reason, the Volumetric Hit Index (*VHI*) can be defined as follows:

$$VHI = \frac{\Sigma_{i=1}^{n}(SIM_i|(SIM_i > t \& OBS_i > t))}{\Sigma_{i=1}^{n}(SIM_i|(SIM_i > t \& OBS_i > t)) + \Sigma_{i=1}^{n}(OBS_i|(SIM_i \le t \& OBS_i > t))} \quad (1)$$

[11] *SIM* refers to satellite observations or climate model simulations being evaluated, whereas *OBS* represents reference observations. In equation (1), *n* is the sample size and *t* is the threshold above which the *VHI* is computed. A $t = 0$ indicates evaluation of the entire distribution of simulated versus observed variables. A higher threshold can be used to evaluate solely the higher quantiles of simulations relative to observations (e.g., *VHI* of values above 50th percentile of observations). By computing the *VHI* above different thresholds, one can plot the performance of *SIM* relative to the magnitude of *OBS*. The *VHI* ranges from 0 to 1 with 1 being the perfect score. A similar threshold concept can be used to derive the Quantile Probability of Detection (*QPOD*) [*AghaKouchak et al*., 2011b] to describe correct detection and identification above a certain threshold (see equation (A1) in Appendix A).

		Event Occurred? **Reference Observations**	
		Yes	No
Event Occurred? **Simulations**	Yes	Hit (*H*)	False (*F*)
	No	Miss (*M*)	True Null Event (*Q*)

Figure 1. The Contingency Table.

[12] It should be noted that *VHI* is an extension of *POD*; however, it is defined slightly differently. In *POD*, the number of $(SIM_i|(SIM_i > t \ \& OBS_i > t))$ is the same as the number of $(OBS_i|(SIM_i > t \ \& OBS_i > t))$. However, the $\Sigma_{i=1}^{n}(SIM_i|(SIM_i > t \ \& OBS_i > t))$ is not identical to $\Sigma_{i=1}^{n}(OBS_i|(SIM_i > t \ \& OBS_i > t))$ as simulated, and observed data sets are often biased against each other. For this reason, the *VHI* is defined as the volume of correctly detected simulations relative to the volume of the correctly detected simulations and missed observations. The *VHI* should be complemented by information on bias, hit bias (defined in *Tian et al*. [2009]), and mean quantile bias $\left(MQB = \Sigma_{i=1}^{n}(SIM_i|SIM_i \ge t)/\Sigma_{i=1}^{n}(OBS_i|OBS_i \ge t)\right)$ for a better understanding of the performance of simulations against observations.

[13] Figure 2 illustrates the differences between *VHI* and *POD* using synthetic precipitation data. In this example, the solid blue line shows the reference observation (*OBS*), whereas the dashed red (SIM_1) and green (SIM_2) lines represent two sets of model simulations (or satellite

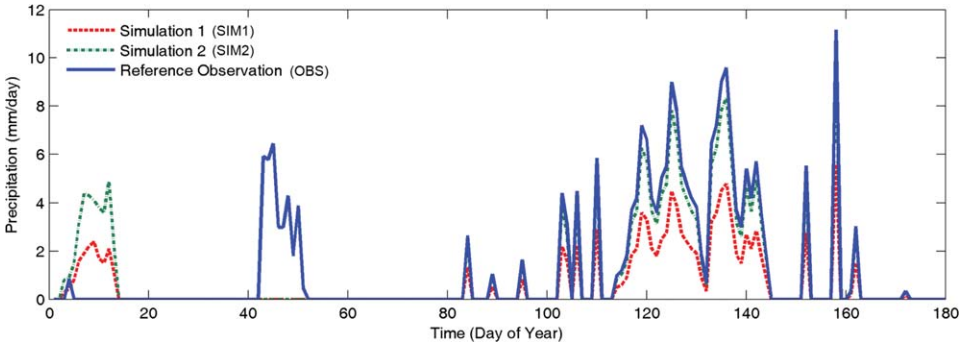

Figure 2. Two synthetic precipitation simulations (SIM_1 and SIM_2) and reference observation (OBS).

observations). The bias values, defined as SIM/OBS show that SIM_1 underestimates by over 50%, while SIM_2 underestimates by 11% (Table 1). Also, a visual comparison indicates that SIM_2 is in better agreement with OBS relative to SIM_1. However, both SIM_1 and SIM_2 lead to the same POD of 0.83 as the number of categorical matches between the two data sets and the observations are the same. The VHI, on the other hand, shows 0.74 and 0.84 for SIM_1 and SIM_2, respectively, indicating the SIM_2 is in better agreement with OBS compared to SIM_1.

[14] Similarly, the Volumetric False Alarm Ratio ($VFAR$) can be expressed as the volume of false SIM above the threshold t relative to the sum of simulations:

$$VFAR = \frac{\Sigma_{i=1}^{n}(SIM_i|(SIM_i > t \& OBS_i \le t))}{\Sigma_{i=1}^{n}(SIM_i|(SIM_i > t \& OBS_i > t)) + \Sigma_{i=1}^{n}(SIM_i|(SIM_i > t \& OBS_i \le t))} \quad (2)$$

[15] The denominator of equation (2) can be summarized as total volume of simulations $\left(\Sigma_{i=1}^{n}(SIM_i|SIM_i > t)\right)$. The $VFAR$ ranges from 0 to 1, with 0 being the perfect score. It should be noted that similar to $QPOD$, one can define the Quantile False Alarm Ratio (QFAR) [*AghaKouchak et al.*, 2011b], which describes the categorical ratio of the number of false identifications of SIM relative to the number of exceedances above a certain threshold (e.g., 90% and 95% quantiles), see equation (A2) in Appendix A. In the earlier example, the FAR values of both SIM_1 and SIM_2 are 0.19, while a visual comparison shows that SIM_2 exhibits more false precipitation (see Figure 2). As shown in Table 1, the $VFAR$ value of SIM_2 is higher than SIM_1 confirming the visual comparison.

[16] The fraction of the volume of missed SIM relative to OBS can be expressed using the Volumetric Miss Index (VMI):

$$VMI = \frac{\Sigma_{i=1}^{n}(OBS_i|(SIM_i \le t \& OBS_i > t))}{\Sigma_{i=1}^{n}(SIM_i|(SIM_i > t \& OBS_i > t)) + \Sigma_{i=1}^{n}(OBS_i|(SIM_i \le t \& OBS_i > t))} \quad (3)$$

[17] The VMI ranges from 0 to 1, with 0 being the perfect score. Based on the definition of $MISS$ ($1 - POD$), the categorical Quantile Miss Index (QMISS) can be expressed as $1 - QPOD$. In the provided example, the $MISS$ index for both SIM_1 and SIM_2 are 0.17 (indicating 17% of categorical miss). However, Figure 2 clearly shows that SIM_2 is in better agreement with OBS. The VMI values confirm that SIM_2 (0.16) is superior to SIM_1 (0.26) with respect to the volume of missed precipitation (see Table 1).

[18] Finally, following the original CSI concept, the Volumetric Critical Success Index ($VCSI$) is defined as an overall measure of volumetric performance:

Table 1. Summary Statistics for Synthetic Data SIM_1 and SIM_2 Presented in Figure 2

Simulation	*POD*	*VHI*	*FAR*	*VFAR*	*MISS*	*VMI*	*CSI*	*VCSI*	Bias (*SIM*/*OBS*)
SIM_1	0.83	0.74	0.19	0.14	0.17	0.26	0.70	0.66	0.49
SIM_2	0.83	0.84	0.19	0.16	0.17	0.16	0.70	0.71	0.89

$$VCSI = \frac{\Sigma_{i=1}^{n}(SIM_i|(SIM_i > t \& OBS_i > t))}{\Sigma_{i=1}^{n}((SIM_i|(SIM_i > t \& OBS_i > t)) + (OBS_i|(SIM_i \le t \& OBS_i > t)) + (SIM_i|(SIM_i > t \& OBS_i \le t)))} \quad (4)$$

[19] The *VCSI* ranges from 0 to 1, with 1 being the perfect score. While the *CSI* values of SIM_1 and SIM_2 are the same (0.70), the *VCSI* values indicate that the SIM_2 is in better agreement with SIM_1 (Table 1). The *QCSI* (equation (A3) in Appendix A) can be used as the categorical equivalent of *VCSI*.

[20] For two daily precipitation data sets (*OBS*: Stage IV radar-based gauge adjusted data; *SIM*: PERSIANN [*Sorooshian et al.*, 2000; *Hsu et al.*, 1997] satellite data; spatial resolution 0.25°), Figure 3 displays sample *POD*, *VHI*, *FAR*, *VFAR*, *MISS*, *VMI*, *CSI*, and *VCSI* values. One can see that the volumetric indices provide additional information beyond the contingency table categorical metrics. For example, the *POD* values range primarily between 0.4 and 0.6, while *VHI* values indicate that *SIM* detects more than 80% of the volume of observed precipitation. While the *FAR* values are relatively high (0.5 in the eastern United States indicating around 50% false precipitation), the *VFAR* values over the eastern United States show that the false precipitation with respect to volume of precipitation is mainly below 10%. Similarly, the *MISS* index shows that *SIM* does not detect a large fraction of precipitation. Based on *VMI*, however, the fraction of precipitation *SIM* does not detect is relatively small (compare *MISS* and *VMI* in Figure 3). This implies that most of the missed events in *SIM* are light rainfall events. The *CSI* values show that for example, in the eastern United States the overall performance score of *SIM* is between 0.3 and 0.5, whereas the *VCSI* indicates a higher performance score (between 0.5 and 0.8) with respect to the volume of precipitation.

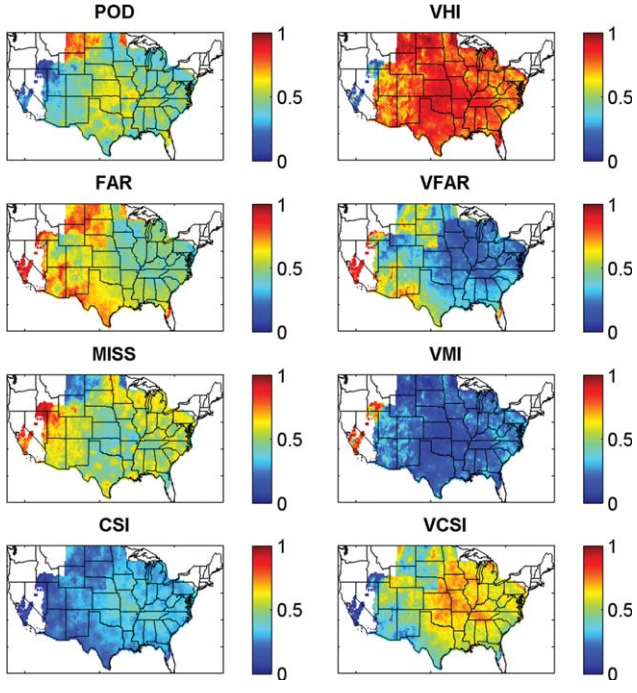

Figure 3. *POD*, *VHI*, *FAR*, *VFAR*, *MISS*, *VMI*, *CSI*, and *VCSI* values for two daily precipitation data sets (*OBS*: Stage IV radar-based gauge adjusted data; *SIM*: PERSIANN [*Sorooshian et al.*, 2000] satellite data; spatial resolution 0.25°).

[21] This example shows that the volumetric measures provide additional information that cannot be achieved from the original categorical metrics. These indices can also be used to decompose biases at high quantiles of a data set by computing them for different thresholds (e.g., $t = 25$th, 50th, 75th percentiles). It is worth pointing out that the volumetric indices should be computed along with the categorical metrics for a comprehensive assessment of simulations against observations.

3. Conclusions

[22] Satellite observations are projected to increase enormously in future. On the other hand, weather and climate models have been widely used to simulate historical and future climate over various spatial and temporal scales. Validation and uncertainty quantification of gridded satellite observations and climate model simulations are fundamental to future improvements in retrieval algorithms and model developments. In this study, the contingency table categorical metrics are extended to volumetric indices for evaluation of gridded data relative to a reference data set. Several indices are introduced including (a) the Volumetric Hit Index (*VHI*) which describes the volume of correctly detected simulations relative to the volume of the correctly detected simulations and missed observations; (b) the Volumetric False Alarm Ratio (*VFAR*) which identifies the volume of false simulations relative to the volume of simulations; (c) Volumetric Miss Index (*VMI*) which expresses the fraction of the volume of missed observations relative to the volume of the correctly detected simulations and missed observations; and (d) the Volumetric Critical Success Index (*VCSI*), defined as an overall measure of volumetric performance including the volumetric hits, false alarms, and misses. The suggested indices decompose the total volumetric error (bias) into volumetric errors terms associated with hit, false, and miss components in simulations.

[23] Using two synthetic time series of simulated precipitation, the volumetric indices are evaluated against the contingency table categorical indices. The synthetic example highlights the difference between the commonly used categorical and the volumetric metrics. The volumetric indices are then applied for validation of a gridded satellite data set relative to reference observations. The results show that the volumetric indices provide additional information beyond the commonly used categorical metrics that can be useful in evaluating gridded data sets.

[24] This study contributes to ongoing metrics development efforts for validation and verification of gridded data sets. It is noted that the introduced volumetric indices are not meant to replace the commonly used categorical metrics. Rather, they should be viewed as metrics that can provide additional information and complement the contingency table categorical metrics. Furthermore, we do not claim that these indices are sufficient for a thorough evaluation of gridded data sets. Additional metrics such as quantile bias, hit bias, relative error, and unbiased root mean square error should also be used for validation and verification studies.

[25] The authors stress that the volumetric indices, similar to other categorical measures, offer tools to identify potential discrepancies in gridded estimates or simulations relative to reference observations. However, interpretation of the results and methods proposed to improve estimates or simulations depends largely on the choice of data and the specific problem at hand. The source code for the suggested volumetric indices is available to public, and interested readers can request a copy from the authors.

Appendix A

[26] Quantile Probability of Detection (*QPOD*) [*AghaKouchak et al.*, 2011] is defined as the *POD* above the threshold t:

$$QPOD = \frac{\Sigma_{i=1}^{n}\mathbf{I}(SIM_i|(SIM_i > t \& OBS_i > t))}{\Sigma_{i=1}^{n}\mathbf{I}(SIM_i|(SIM_i > t \& OBS_i > t)) + \Sigma_{i=1}^{n}\mathbf{I}(OBS_i|(SIM_i \le t \& OBS_i > t))} \quad \text{(A1)}$$

where t is the threshold (e.g., 90% and 95% quantiles); $\mathbf{I}$ is the indicator function; and n is the number of exceedances. *QPOD* represents the ratio of the number of correct identifications above a certain threshold (t) relative to the total number of exceedances (n). The *QPOD* ranges from 0 (no detection skill) to 1 (perfect detection). Similarly, the Quantile False Alarm Ratio (QFAR) can be expressed as

$$QFAR = \frac{\Sigma_{i=1}^{n}\mathbf{I}(SIM_i|(SIM_i > t \& OBS_i \le t))}{\Sigma_{i=1}^{n}\mathbf{I}(SIM_i|(SIM_i > t \& OBS_i > t)) + \Sigma_{i=1}^{n}\mathbf{I}(SIM_i|(SIM_i > t \& OBS_i \le t))} \quad \text{(A2)}$$

[27] The *QFAR* ranges from 0 (perfect score) to 1. Quantile Critical Success Index (QCSI) is defined as the *CSI* above the threshold t:

$$QCSI = \frac{\Sigma_{i=1}^{n}\mathbf{I}(SIM_i|(SIM_i > t \& OBS_i > t))}{\Sigma_{i=1}^{n}\mathbf{I}((SIM_i|(SIM_i > t \& OBS_i > t)) + (OBS_i|(SIM_i \le t \& OBS_i > t)) + (SIM_i|(SIM_i > t \& OBS_i \le t)))} \quad \text{(A3)}$$

[28] **Acknowledgments.** We thank the Editor and reviewers for their thoughtful suggestions and comments upon an early draft of this technical note. The financial support for this study is made available from the National Science Foundation (NSF) awards EAR-1316536 and OISE-1243543 and the United States Bureau of Reclamation (USBR) award R11AP8145 1.

References

AghaKouchak, A., A. Bárdossy, and E. Habib (2010a), Copula-based uncertainty modeling: Application to multi-sensor precipitation estimates, *Hydrol. Processes*, *24*(15), 2111–2124.

AghaKouchak, A., E. Habib, and A. Bárdossy (2010b), Modeling radar rainfall estimation uncertainties: Random error model, *J. Hydrol. Eng.*, *15*(4), 265–274.

AghaKouchak, A., N. Nasrollahi, J. Li, B. Imam, and S. Sorooshian (2011a), Geometrical characterization of precipitation patterns, *J. Hydrometeorol.*, *12*(2), 274–285.

AghaKouchak, A., A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai (2011b), Evaluation of satellite-retrieved extreme precipitation rates across the central United States, *J. Geophys. Res.*, *116*, D02115, doi:10.1029/2010JD014741.

Anagnostou, E. N., W. F. Krajewski, D.-J. Seo, and E. R. Johnson (1998), Mean-field rainfall bias studies for WSR-88D, *J. Hydrol. Eng.*, *3*(3), 149–159.

Anagnostou, E. N., V. Maggioni, E. I. Nikolopoulos, T. Meskele, F. Hossain, and A. Papadopoulos (2010), Benchmarking high-resolution global satellite rainfall products to radar and rain-gauge rainfall estimates, *IEEE Trans. Geosci. Remote Sens.*, *48*(4), 1667–1683.

Behrangi, A., B. Khakbaz, T. Jaw, A. AghaKouchak, K. Hsu, and S. Sorooshian (2011), Hydrologic evaluation of satellite precipitation products at basin scale, *J. Hydrol.*, *397*, 225–237.

Brown, B., R. Bullock, C. David, J. Gotway, M. Chapman, A. Takacs, E. Gilleland, K. Manning, and J. Mahoney (2004), New verification approaches for convective weather forecasts, in 11th Conference on Aviation, Range, and Aerospace Meteorology, Hyannis, Mass., 4–8 Oct.

Davis, C. A., B. Brown, R. Bullock, and J. Gotway (2009), The Method for Object-based Diagnostic Evaluation (MODE) applied to WRF forecasts from the 2005 5 NSSL/SPC Spring program, *Weather Forecasting*, *24*, 1327–1342.

Dorigo, W., K. Scipal, R. Parinussa, Y. Liu, W. Wagner, R. De Jeu, and V. Naeimi (2010), Error characterisation of global active and passive microwave soil moisture datasets, *Hydrol. Earth Syst. Sci.*, *14*(12), 2605–2616.

Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. Crow (2010), Performance metrics for soil moisture retrievals and application requirements, *J. Hydrometeorol.*, *11*(3), 832–840.

Feddema, J., K. Oleson, G. Bonan, L. Mearns, W. Washington, G. Meehl, and D. Nychka (2005), A comparison of a GCM response to historical anthropogenic land cover change and model sensitivity to uncertainty in present-day land cover representations, *Clim. Dyn.*, *25*(6), 581–609, doi:10.1007/s00382-005-0038-z.

Gebremichael, M. (2010), Framework for satellite rainfall product evaluation, *Geophys. Monogr. Ser.*, *191*, 265–275.

Gilleland, E. (2013), Testing competing precipitation forecasts accurately and efficiently: The spatial prediction comparison test, *Mon. Weather Rev.*, *141*(1), 340–355.

Gleckler, P., K. Taylor, and C. Doutriaux (2008), Performance metrics for climate models, *J. Geophys. Res.*, *113*, D06104, doi:10.1029/2007JD008972.

Gourley, J. J., J. M. Erlingis, Y. Hong, and E. B. Wells (2012), Evaluation of tools used for monitoring and forecasting flash floods in the United States, *Weather Forecasting*, *27*(1), 158–173.

Haile, A. T., E. Habib, and T. Rientjes (2012), Evaluation of the Climate Prediction Center (CPC) Morphing technique (CMORPH) rainfall product on hourly time scales over the source of the Blue Nile River, *Hydrol. Processes*, *27*, 1829–1839.

Hao, Z., A. AghaKouchak, and T. J. Phillips (2013), Changes in concurrent monthly precipitation and temperature extremes, *Environ. Res. Lett.*, *8*(3), 034014, doi:10.1088/1748-9326/8/3/034014.

Hirpa, F. A., M. Gebremichael, and T. Hopson (2010), Evaluation of high-resolution satellite precipitation products over very complex terrain in Ethiopia, *J. Appl. Meteorol. Climatol.*, *49*(5), 1044–1051.

Hossain, F., and G. Huffman (2008), Investigating error metrics for satellite rainfall data at hydrologically relevant scales, *J. Hydrometeorol.*, *9*(3), 563–575.

Hsu, K., X. Gao, S. Sorooshian, and H. Gupta (1997), Precipitation estimation from remotely sensed information using artificial neural networks, *J. Appl. Meteorol.*, *36*, 1176–1190.

Jackson, T., D. Entekhabi, and E. Njoku (2005), The Hydros mission and validation of soil moisture retrievals, *Geophys. Res. Abstr.*, *7*, 00434.

Jiang, J. H., et al. (2012), Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA A-Train satellite observations, *J. Geophys. Res.*, *117*, D14105, doi:10.1029/2011JD017237.

Liepert, B. G., and M. Previdi (2012), Inter-model variability and biases of the global water cycle in CMIP3 coupled climate models, *Environ. Res. Lett.*, *7*(1), 014006, doi:10.1088/1748–9326/7/1/014006.

Mehran, A., and A. AghaKouchak (2013), Capabilities of satellite precipitation datasets to estimate heavy precipitation rates at different temporal accumulations, *Hydrol. Processes*, doi:10.1002/hyp.9779.

Norouzi, H., M. Temimi, W. Rossow, C. Pearl, M. Azarderakhsh, and R. Khanbilvardi (2011), The sensitivity of land emissivity estimates from AMSR-E at C and X bands to surface properties, *Hydrol. Earth Syst. Sci.*, *15*(11), 3577–3589.

Phillips, T., and P. Gleckler (2006), Evaluation of continental precipitation in 20th century climate simulations: The utility of multimodel statistics, *Water Resour. Res.*, *42*(3), W03202, doi:10.1029/2005WR004313.

Pinker, R. T., D. Sun, M.-P. Hung, C. Li, and J. B. Basara (2009), Evaluation of satellite estimates of land surface temperature from GOES over the United States, *J. Appl. Meteorol. Climatol.*, *48*(1), 167–180.

Sorooshian, S., K. Hsu, X. Gao, H. Gupta, B. Imam, and D. Braithwaite (2000), Evolution of the PERSIANN system satellite-based estimates of tropical rainfall, *Bull. Am. Meteorol. Soc.*, *81*(9), 2035–2046.

Tian, Y., C. Peters-Lidard, J. Eylander, R. Joyce, G. Huffman, R. Adler, K. Hsu, F. Turk, M. Garcia, and J. Zeng (2009), Component analysis of errors in satellite-based precipitation estimates, *J. Geophys. Res.*, *114*, D24101, doi:10.1029/2009JD011949.

Turk, F. J., P. Arkin, E. E. Ebert, and M. R. P. Sapiano (2008), Evaluating high-resolution precipitation products, *Bull. Am. Meteorol. Soc.*, *89*(12), 1911–1916.

Wilks, D. (2006), *Statistical Methods in the Atmospheric Sciences*, 2nd ed., 627 pp., Academic, Burlington, Mass.