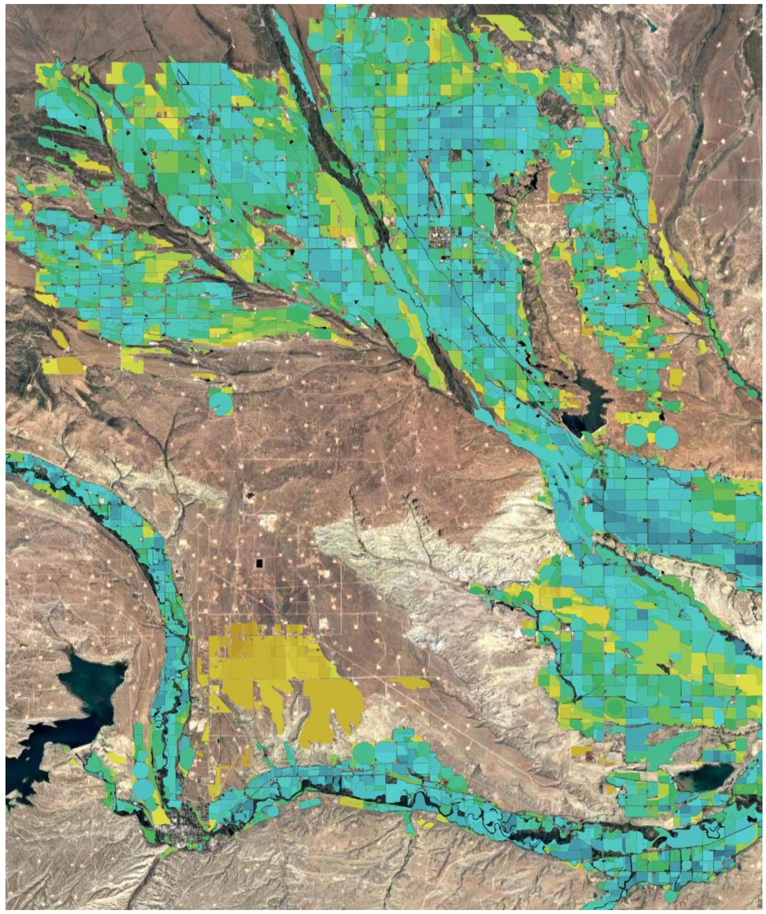
Appendix G

OpenET Comparison

Appendix G: Upper Colorado River Basin OpenET Intercomparison Summary



Justin L. Huntington[1,2]
Christopher Pearson[1]
Blake Minor[1]
John Volk[1]
Charles Morton[3]
Forrest Melton[4,5]
Richard Allen[6]

with contributions from OpenET team members

January 2022

Prepared by
[1]Desert Research Institute, Division of Hydrologic Sciences
[2]Desert Research Institute, Western Regional Climate Center
[3]Desert Research Institute, Division of Earth and Ecosystem Sciences
[4]NASA Ames Research Center, Cooperative for Research in Earth Science and Technology
[5]California State University, Monterey Bay
[6]University of Idaho, College of Agricultural and Life Sciences

Prepared for
U.S. Bureau of Reclamation

THIS PAGE INTENTIONALLY LEFT BLANK

(cover image taken from OpenET Data Explore, near Duchesne, UT)

INTRODUCTION

Water resources managers and stakeholders across the western United States are reliant on evapotranspiration (ET) and consumptive use (ET minus effective precipitation) information from irrigated lands for water management, water rights, hydrologic modeling and prediction, and water planning. To date, some state and federal agencies use outdated or simplified methods to estimate ET that rely on crop coefficients and outdated irrigated area maps to estimate consumptive use volumes, while others rely on manual processing of satellite data over limited areas, making processing and coverage for all irrigated lands challenging and costly, and based on available staff, expertise, and agency funding. Based on our team's experience and interactions with western state water resource agencies, there is a great need for accurate, defensible, and timely maps of ET, consumptive use, and irrigated area summaries, that are further summarized to watershed and administrative boundaries - the spatial scales in which water rights are managed. Additionally, field-by-field, watershed, and administrative boundary estimates of ET, consumptive use, and irrigated lands are needed at the end of each irrigation season, and retrospective analyses on the order of years-to-decades are needed to inform surface and groundwater water budgets, naturalized flow studies, and assessment of water conservation programs. Satellite-based field-scale ET data based on best available science already exist for some regions but are often not readily available at the watershed scale. Consistency, transparency, and reproducibility of ET data are important factors to consider when estimating ET and consumptive use across political boundaries. The OpenET consortium and software platform was established to address these needs.

OpenET is a satellite-based ET cloud computing and data services platform (Melton et al., 2021; https://openetdata.org/) that builds upon decades of research by NASA, USGS and USDA, and involves more than 45 scientists and software engineers from four NASA Centers (ARC, MSFC, JPL, GSFC), USGS, USDA, eight universities including Desert Research Institute (DRI), and non-governmental organization (NGO) and private sector partners. OpenET provides daily, monthly, and annual ET data at a spatial resolution of 30 m x 30 m (0.22 acres) using Landsat, gridded weather data, and an ensemble of well-established satellite-driven ET models implemented on the Google Earth Engine cloud computing platform (Gorelick et al., 2017). Google Earth Engine allows for efficient interoperability across different datasets, massive parallel processing, storage, scalability in time and space, and automated operational updates with new observations that would otherwise be costly and burdensome for State and federal agencies to develop, operate, and maintain.

OBJECTIVE

The objectives of this study were to 1) summarize OpenET estimates of growing season (April through October) ET and consumptive use for irrigated lands within the Upper Colorado River Basin (UCRB), 2) compare OpenET estimates to satellite-based consumptive use estimates developed as part of Reclamation's Consumptive Use Feasibility Study, 3) compare OpenET and Reclamation study ET estimates to in-situ ET data derived from eddy covariance stations, 4) highlight OpenET intercomparison and accuracy assessment results for growing seasons that were developed using over 30 cropland eddy covariance stations across the U.S., and 5) discuss limitations and considerations of OpenET and satellite-based ET estimates.

METHODS

OpenET

Most of the models that make up the OpenET ensemble are based on full or simplified implementations of the surface energy balance (SEB) approach. The SEB approach accounts for the energy used to transform liquid water in plants and soil into vapor that is released to the atmosphere. The SEB approach relies on satellite measurements of surface temperature and surface reflectance combined with other key land surface and weather variables to estimate components of the energy balance—net radiation, sensible heat flux, ground heat flux, and latent heat flux, which is the energy consumed through ET. METRIC, geeSEBAL, and DisALEXI estimate each component of the energy balance using optical (i.e., short-wave) and thermal (i.e., long-wave) data, whereas SSEBop and PT-JPL are simplified approaches in which certain components of the energy balance are not estimated or are calculated using a set of simplifying assumptions. SIMS relies on surface reflectance data and crop type information to compute ET as a function of canopy density using a crop coefficient approach for agricultural lands.

Landsat is currently the primary satellite dataset used within the OpenET platform. The Landsat program, a joint program of the U.S. Geological Survey and NASA, provides the longest continuous space-based record of Earth's land in existence, dating back to 1972 for optical data and to 1982 for thermal data. Landsat is the only operational satellite that combines thermal and optical data at the spatial resolution needed to assess field-scale water use. Multiple models implemented within the OpenET framework also integrate data from GOES, Sentinel-2, Suomi NPP, Terra, Aqua and other satellites to produce ET data at a range of spatial and temporal scales.

One of the primary variables derived from gridded weather data is the reference ET, which is the amount of ET from a reference surface, typically a well-watered grass or alfalfa crop. OpenET uses reference ET data calculated using the American Society of Civil Engineers (ASCE) Standardized Penman-Monteith equation (ASCE 2005) for a grass reference surface, usually notated as ETo, and is a function of solar radiation, air temperature, humidity, and wind speed. ETo data are used to support the calculation of actual ET between Landsat satellite overpasses, which occur every eight days (excepting cloud cover) with Landsat 7 and 8 satellites in orbit, and every five days with Landsat 7, 8, and 9 in orbit. First, the fraction of ETo for each satellite overpass date is calculated by dividing the satellite ET on the overpass date by the ETo. Fraction of ETo values are then linearly interpolated on a daily timestep for all days between clear satellite overpass dates, one image pixel at a time, and are then multiplied by the daily ETo values to calculate a daily time series of actual ET for every pixel. These per-pixel daily time series of actual ET are then aggregated to monthly and annual time periods. The fraction of ETo is interpolated in time because it tends to change in proportion to changes in vegetation cover, like its equivalent, the widely used crop coefficient. Daily ETo is then used to reintroduce the impacts of day-to-day changes in weather on actual ET rates. It should be noted that eeMETRIC and SSEBop rely on alfalfa reference ET (ETr) internally within OpenET, and ratios of ETr to ETo are computed and used to allow for time integration using ETo. Differences in ETo and ETr are not consistent, and are more pronounced during winter months, and during high wind and low humidity conditions due to the more sensitive wind function within the ETr equation (Irmak et al., 2008).

Similar to the approach used by the USBR Consumptive Use Working Group, reference ET surfaces for OpenET are based on a bias corrected gridMET reference ET product developed using agricultural weather stations located within and surrounding the UCRB (see Appendix D). Mean monthly bias correction factors for each station location were developed by comparing weather station calculated reference ET to gridMET reference ET throughout the Western U.S. OpenET bias correction factors are identical to those used by the UCRB CU Working Group for stations within and surrounding the UCRB, but include other stations located greater than 31 miles (50 km) outside the UCRB. However, the influence of these additional stations on spatially interpolated bias correction factors within the UCRB basin is minimal. Station-based mean monthly bias correction factors were spatially interpolated throughout the study area using a smoothed kriging approach for OpenET, whereas the USBR interpolation was based on an Inverse Distance Weighting approach. Comparison between the two approaches indicate only small differences, however, larger differences are evident in areas far from agricultural weather stations. See Appendix D, section *Weather Station Data Quality Assurance and Quality Control, Comparison, and Bias Correction of Gridded Reference ET***,** for a detailed description of the development of USBR bias correction layers.

OpenET Multi-Model Ensemble

OpenET provides ET estimates for each of the individual models shown in Table 1, as well as a multi-model mean value. Differences in model physics, assumptions, and input data result in a range of ET estimates from the ensemble of models included in OpenET. The use of multi-model ensembles is a common practice within the climate science, hydrology, and decision-making communities. For many climate and hydrology applications, it has been shown that when estimates from an ensemble of models are combined, they yield estimates that are, on average, equally or more accurate than any individual model (Thompson, 1977; Branzei et al., 2001; Kirtman et al., 2014; Arsenault et al., 2015). In addition to improved accuracy, the use of a single estimate calculated from an ensemble of ET models may reduce confusion about which ET model to use, provide a path toward acceptance and consistency, and is useful for identifying both model outliers and potential errors in ground-based ET datasets. In cases where ET estimates vary substantially, legitimate questions around model accuracy and which model is "the best" can present significant barriers to the operational use and adoption of satellite-based ET data.

Table 1. Ensemble of ET models included in the OpenET platform.

Model Acronym	**Model Name**	**Primary References**
ALEXI/DisALEXI	**A**tmosphere-**L**and **Ex**change **I**nverse / **Dis**aggregation of the **A**tmosphere-**L**and **Ex**change **I**nverse	Anderson et al. (2007); Anderson et al.(2018);
eeMETRIC	Google **E**arth **E**ngine implementation of the **M**apping **E**vapo**t**ranspiration at high **R**esolution with **I**nternalized **C**alibration model	Allen et al. (2005); Allen et al. (2007); Allen et al. (2011)
geeSEBAL	**G**oogle **E**arth **E**ngine implementation of the **S**urface **E**nergy **B**alance **A**lgorithm for **L**and	Bastiaanssen et al. (1998); Laipelt et al. (2021)
PT-JPL	**P**riestley-**T**aylor **J**et **P**ropulsion **L**aboratory	Fisher et al. (2008)
SIMS	**S**atellite **I**rrigation **M**anagement **S**upport	Melton et al. (2012); Pereira et al. (2020)
SSEBop	**Op**erational **S**implified **S**urface **E**nergy **B**alance	Senay et al. (2013); Senay et al. (2018)

A key objective of OpenET is to provide a single ET estimate for each location and time step, calculated from an ensemble of six models, while making individual model results available to provide transparency and support assessment and increased understanding of uncertainties, and allow for the use of a single or subset of models in areas where model performance and biases are well understood. Many multi-model ensemble averaging approaches exist, ranging from the simple arithmetic average, weighted average, to stochastic Bayesian model averaging. Each approach has strengths and weaknesses related to simplicity, speed, accuracy, and ease of operational implementation. The optimal approach ideally addresses most, if not all, of these factors. Limitations due to small sample size, outliers, and overfitting also need to be considered. For OpenET, a simple yet robust approach was chosen where the single ensemble ET estimate is computed at monthly time steps as the simple arithmetic average after outlier ET estimates are removed. Outlier ET estimates are detected and removed using the Median Absolute Deviation (MAD) method initially developed by Carl Friedrich Gauss in 1816, and more recently rediscovered and popularized by Hampel (1974) and Leys et al. (2013). The MAD is a measure of scale, or spread of the data, based on the median of the absolute deviations from the median of the distribution. Huber (1981) describes the method as "the single most useful ancillary estimate of scale" since it overcomes many limitations of more common standard deviation and interquartile approaches for identifying outliers. The MAD approach is applied at the per-pixel level prior to computing the average ET for the remaining ensemble of models for a given timestep. The use of a single ET value calculated from the ensemble of models may ultimately reduce barriers to use and adoption of remotely sensed ET for a wide range of water management applications.

OpenET Platform

OpenET includes a Data Explorer (Figure 1) that provides a web-based graphical user interface (UI) allowing users to easily explore and query ET data across the western U.S. for any field or location of interest. The web UI is based on Leaflet for mapping functionality, JQuery (Severance, 2015) for front end interactive features, and HighCharts (Kuan, 2012) for interactive generation of graphs. OpenET provides access to both spatially continuous gridded datasets and choropleth maps that summarize data to individual field boundaries.

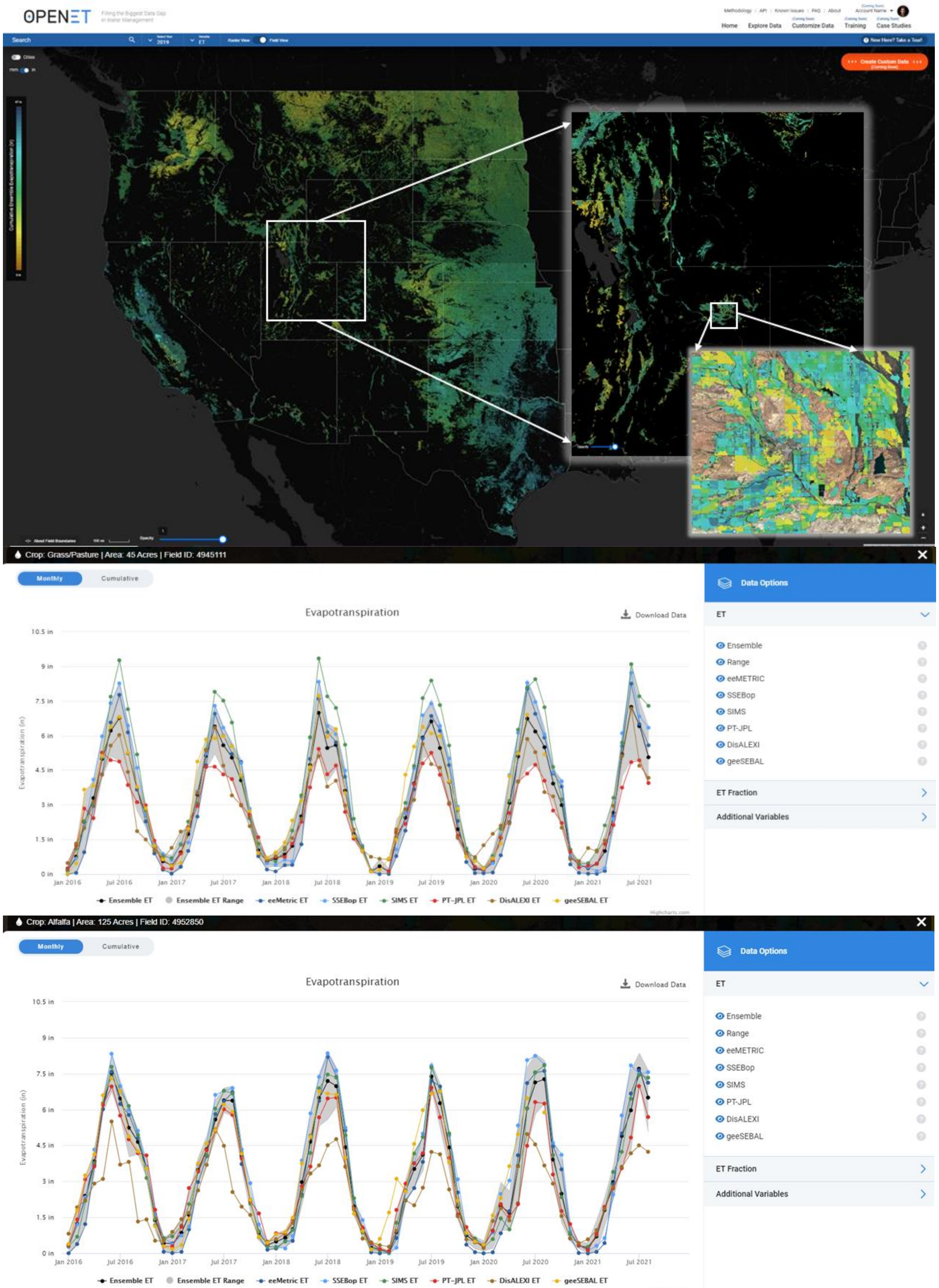


Figure 1. Illustration of the OpenET User Interface and field-level ET summaries, regional inset of a portion of the UCRB, local inset of the Roosevelt, UT area (Top), monthly ET time series of OpenET models for a single pasture grass and alfalfa field, with the range of ET estimates included in the ensemble average (shaded area) (middle and lower, respectively).

Google Earth Engine is used to generate individual model and ensemble average ET estimates, and spatial statistics of ET, ETo, fraction of ETo, irrigation status, USDA crop type classifications, vegetation indices, precipitation estimates, and other spatial data mapped to predefined polygons (e.g., states, counties, watersheds, irrigation districts, agricultural field boundaries). Spatial statistics are stored in a large PostGIS database for polygon summaries and time series analyses. Raster and polygon databases are connected to an API (application programming interface) and open source raster and vector tiling software. This framework supports rapid response to data queries, as well as user-friendly spatial and temporal visualizations of ET and associated variables (e.g., NDVI, reference ET, fraction of reference ET, irrigation status) via the UI (user interface). An overview of the OpenET system architecture is illustrated in Figure 2.

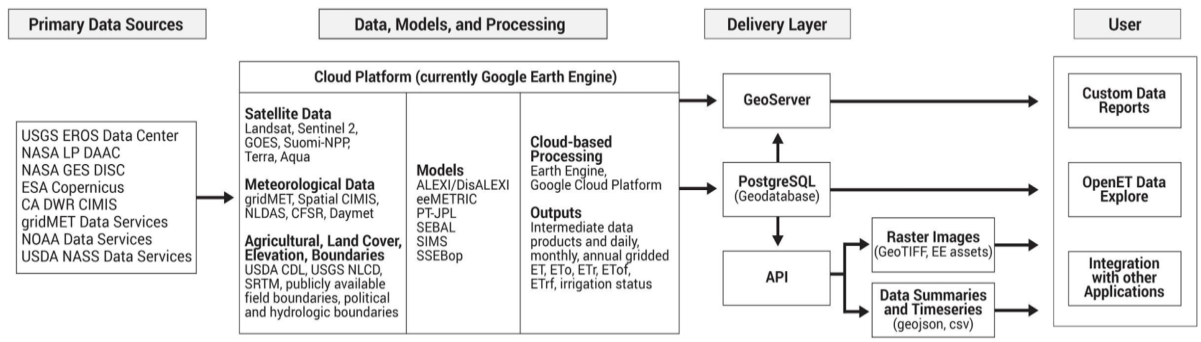


Figure 2. Overview of the OpenET system architecture.

Effective Precipitation and Consumptive Use

Currently, OpenET does not compute or provide consumptive use estimates of irrigation water. Rather, it computes and provides total ET. Therefore, for this analysis, effective precipitation was estimated and subtracted from OpenET estimates of ET so that consumptive use of irrigation water could be estimated, summarized, and compared with Reclamation estimates. Consumptive use of irrigation water is defined in this work as the ET minus effective precipitation, and is synonymous with the terms 'depletion' and 'net ET.' Because total actual ET includes ET derived from precipitation, a portion of precipitation that is 'effective' or contributes to ET must be subtracted from the total ET estimate to estimate the amount of net ET or consumptive use of irrigation water, or of groundwater in the case of groundwater dependent ecosystems. Consumptive use estimates ultimately serve as the foundational data for estimating irrigation application rates, irrigation withdrawals, and irrigation water requirements (Jensen and Allen, 2016; Allen et al., 1998).

Effective precipitation is the amount of precipitation that is available for ET (Bos et al., 2009). It is commonly estimated as precipitation minus losses to runoff and deep percolation beyond the crop root zone (Allen et al., 1998). There are many approaches for estimating effective precipitation - most are simple (e.g., as a ratio of ETo and precipitation) and some are complex (e.g., detailed soil water balance and crop modeling) (Dastane, 1974; Kumar et al., 2017). Many empirical approaches have been developed for a specific set of conditions, and unless they consider the primary factors that influence infiltration, runoff, and deep percolation, their accuracy and ability for general application and transferability are limited. An older and commonly used method developed by the USDA Soil Conservation Service (SCS) (USDA, 1967) to estimate monthly effective precipitation is dependent on the estimated net depth of irrigation application, total monthly precipitation, and average monthly Blaney-Criddle estimated consumptive use. The Reclamation method (Stamm, 1967) is based on the mean seasonal precipitation of the five consecutive driest years, and percentages that are applied to increments of monthly precipitation for the time period of interest. Most of the traditional effective precipitation methods, including the USDA and Reclamation methods, are typically only applied during the growing season. More complex soil water balance and crop modeling approaches consider runoff, deep percolation, and crop specific factors such as root depth and plant available water capacity, winter cover, and winter and spring soil water accumulation or "carry over" that is available for ET during spring and that offsets early season irrigation water requirements (Feddes, 1988; Huntington et al., 2015).

For this work, consumptive use of irrigation water was estimated by subtracting November through October (water year shifted forward by one month) total effective precipitation estimates derived from the daily soil water balance model within the ET Demands model from growing season total ET. Components of the ET Demands model are described in Huntington et al., (2015) and Allen et al., (2020). Spatial distributions of crop-weighted effective precipitation rasters were developed from each year's ET Demands model output (Figure 3), and respective effective precipitation estimates were subtracted from satellite-based estimates of ET. Effective precipitation is calculated within ET Demands as a function of daily precipitation (from gridMET; Abatzoglou, 2013), antecedent soil moisture prior to a precipitation event, deep percolation from precipitation, and runoff from precipitation. The daily soil water balance method was demonstrated to be the most accurate means for estimating effective precipitation by Patwardhan et al. (1990). Soil moisture and plant available water are a function of the water holding capacity and root depths for crop areas within each model cell. Runoff from precipitation is calculated based on daily precipitation using the NRCS curve number (CN) method (USDA-SCS 1972) while scaling CN values in between dry and wet conditions according to antecedent soil water content using expressions of Hawkins et al. (1985).

Comparison with UCRB Eddy Covariance Stations

The following section compares satellite-based USBR and OpenET estimates to in-situ ET estimates from the four UCRB eddy covariance (EC) flux stations. Satellite-based USBR ET data and EC station ET data, including those that had been adjusted for energy balance closure, were provided by Wilson Water Group. USBR and OpenET results were extracted for pixels surrounding each EC station location using dynamic EC flux source areas or "flux footprints" (Kljun et al., 2015) that were weighted by hourly ETo to take into account the area contributing to evaporative flux measured at the station. Satellite-based USBR ET data at daily and monthly time steps were sampled at the EC stations using daily and monthly EC flux source areas (Figures 10-13), both weighted by hourly ETo. The daily sampling method is consistent with past USBR reports and the monthly method is used by OpenET.

For this work, USBR and OpenET estimates were compared to closed energy balance estimates of ET. Closed energy balance estimates of ET were estimated by Utah State University partners using the Bowen ratio technique (Twine et al., 2000). Energy imbalance has been a significant topic of research and discussion among the micrometeorological community (Foken, 2008, Leuning et al., 2012). Energy imbalance, where available energy (net radiation less soil heat flux) is larger than the sum of turbulent fluxes, commonly ranges between 10 to 30 percent (Foken, 2008). For the UCRB EC stations, average daily available energy was larger than the average daily turbulent flux by 27 to 33 percent. There are many reasons for energy imbalance, but the most obvious is the disparity in measurement height and horizontal scales (i.e., EC source area) between available energy and turbulent flux sensors as well as impacts of regional heterogeneities on large turbulent structures (Foken, 2008). While measurement errors clearly contribute to energy imbalance, they alone cannot solve the closure problem, nor can the many processing steps and corrections required to solve for closure. Other sources of imbalance are flux divergence due to terrain and vegetation heterogeneities, interaction of scales, and low-frequency mesoscale eddies not being captured by the instrumentation (Foken, 2008). Energy imbalance corrections are commonly applied by researchers and practitioners even though sources of errors or the energy balance terms in question are not fully understood and quantifiable. Typically, turbulent fluxes are targeted for adjustment through increasing latent heat flux (LE) and sensible heat flux (H) according to the energy balance ratio (EBR; the ratio of turbulent fluxes to available energy) (Pastorello et al., 2020) and by maintaining a consistent Bowen ratio (the ratio of H to LE) (Twine et al., 2000). For this analysis, satellite-based USBR and OpenET estimates were compared to EBR closed energy balance ET estimates, however, for illustration purposes closed and unclosed ET estimates are shown in time series comparison plots to illustrate the full range of uncertainty in the EC derived ET data.

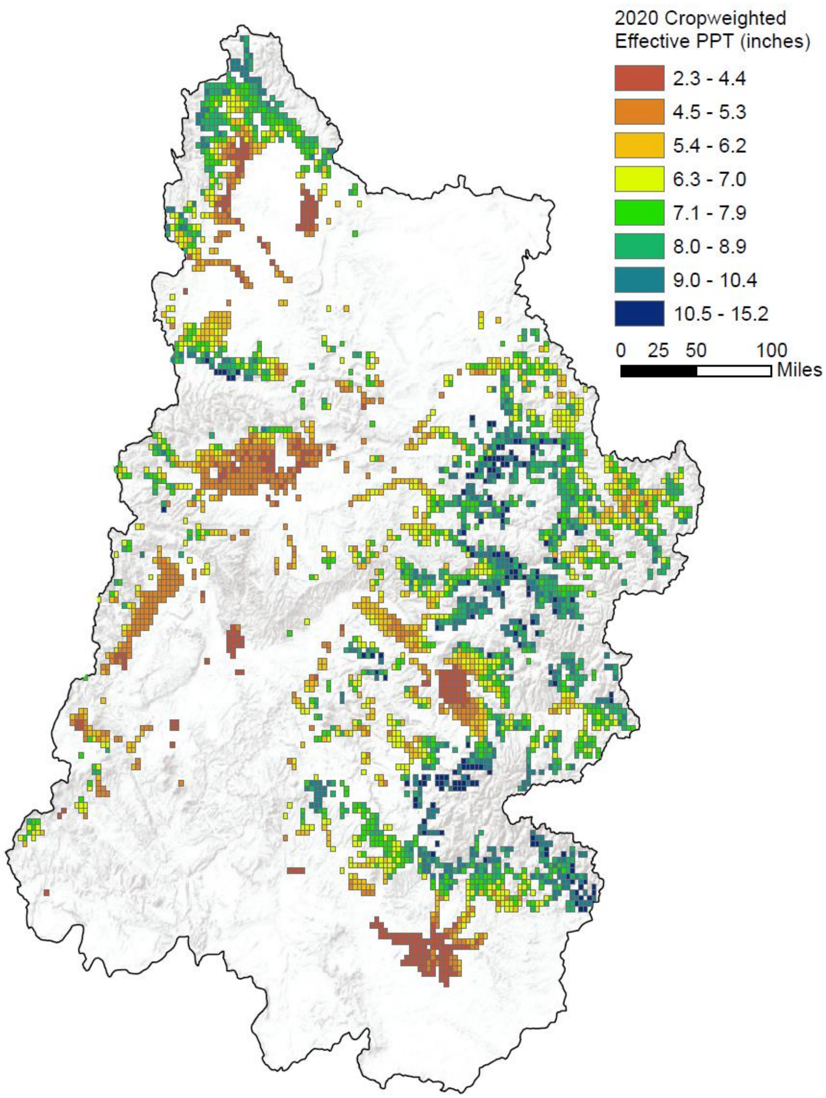


Figure 3. Distribution of 2020 effective precipitation for gridMET grids containing irrigated parcels in the UCRB from the ET Demands model. Total effective precipitation was calculated using daily estimates from the 12-month period ending in October of the growing season for the year of interest.

The dual crop coefficient method used within ET Demands allows for separate accounting of transpiration and evaporation to better quantify evaporation from variable precipitation and simulated irrigation events, and in turn allows for accounting of wintertime soil moisture gains. Accounting for winter soil moisture gains and losses is important for accurately estimating consumptive use of irrigation water in areas that receive the majority of precipitation during the non-growing season, such as in the UCRB. While soil water balance modeling approaches, such as ET Demands, are generally considered to be the most accurate for estimating effective precipitation (Patwardhan et al., 1990; Kumar et al., 2017; Ali and Mubarak, 2017), they are not without limitations. Further analyses, intercomparisons, and discussions with UCRB working group partners are needed to better understand differences between effective precipitation methods, as well as working group partner philosophies on what is considered "effective" (e.g., should the E in ET be considered effective, since it may not offset irrigation water requirements? (see Allen and Robison, 2007 and Jensen and Allen 2016)), time periods for totaling effective precipitation (calendar year vs. water year), and potential impacts on growing season and annual consumptive use estimates.

RESULTS AND DISCUSSION

Comparison of USBR and OpenET Satellite-based ET Estimates

The USBR Consumptive Use Study produced two satellite-based ET datasets generated using "manual" applications (i.e., expert supervised, semi-automated model runs) of METRIC and a mix of manual and automated applications of SSEBop. The core principles and algorithms of the model versions used in the USBR study and the OpenET model versions remain the same; however, differences related to processing platforms, satellite image collections (e.g., Landsat Collection 1 versus Landsat Collection 2), reference ET source and type (alfalfa vs. grass references), time integration, cloud and pixel screening, and calibration (i.e., automated versus manual) lead to differences in final ET estimates. Please refer to the USBR METRIC and USBR SSEBop Appendices for detailed descriptions and applications of each model version.

Comparison of the 2020 USBR manual applications with OpenET eeMETRIC and SSEBop results were summarized for the growing season (April-October) for irrigated lands over the entire UCRB. The comparisons show consumptive use differences of 197,400 acre-ft (7.1%) and -132,546 acre-ft (-4.2%), respectively (Figure 4). These differences are within expected ranges considering differences between manual and automated workflows, reference ET sources and types, and cloud and pixel screening thresholds. Results from previous years show relatively consistent relationships between USBR METRIC and eeMETRIC, however, early year applications of SSEBop show larger differences than current year comparisons (Table 2). Model improvements and refinements to SSEBop throughout the course of the study make direct comparison of OpenET and early year USBR SSEBop inconsistent. SSEBop developers recommend use of the latest model version applied during the 2020 UCRB assessment, which is the latest SSEBop version applied within OpenET for all years (ver 0.1.5). While the model version is identical for both 2020 OpenET and USBR SSEBop applications, OpenET uses grass reference ET (ETo) while USBR applications use alfalfa reference ET (ETr) for time integration. In addition, a 70% versus 40% or greater cloud cover threshold was used to filter and remove images from time integration within OpenET and USBR SSEBop processing, respectively. The following section highlights 2020 comparisons, however, previous year summaries are provided in the Supplemental Figures section of this Appendix.

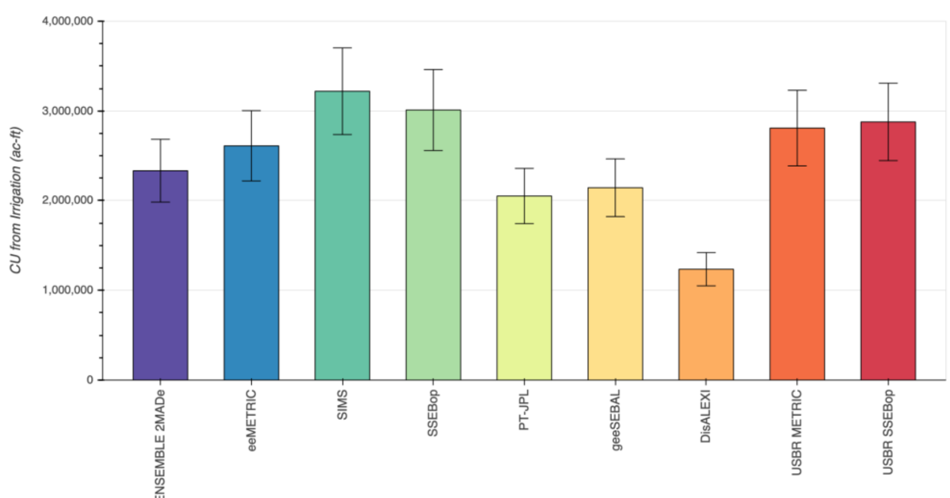


Figure 4. UCRB 2020 growing season consumptive use from irrigation totals for each model. Theoretical potential error bars of +/- 15% are shown for all models for illustration and comparison purposes only. Irrigation consumptive use was calculated as the difference between growing season ET and annual effective precipitation from ET Demands.

Table 2. Summary of 2017-2020 basin-wide growing season irrigation consumptive use from the OpenET and USBR METRIC and SSEBop model applications.

	2017	2018	2019	2020
USBR METRIC (acre-ft)	2,354,471	2,260,233	2,331,068	2,808,759
OpenET eeMETRIC (acre-ft)	2,131,156	2,336,939	2,379,487	2,611,377
Difference (acre-ft)	223,315	-76,706	-48,419	197,382
% Difference	10.5	-3.3	-2.0	7.6
USBR SSEBop (acre-ft)	1,812,464	1,852,339	2,003,163	2,877,750
OpenET SSEBop (acre-ft)	2,324,074	2,674,529	2,672,766	3,010,296
Difference (acre-ft)	-511,610	-822,190	-669,603	-132,546
% Difference	-22.0	-30.7	-25.1	-4.4

The 2020 basin total OpenET ensemble estimate of 2,504,000 acre-ft falls within 11% to 27% of the USBR and OpenET eeMETRIC and SSEBop results (Table 3). The OpenET ensemble estimate is lower than both USBR and OpenET eeMETRIC and SSEBop estimates due to the influence of lower ET estimates by PT-JPL, geeSEBAL, and DisALEXI (Figures 1 and 4). While the MAD outlier detection approach aims to remove estimates that are substantially different from the median of all models, it is challenged to identify outliers when a relatively large range among all estimates occurs. This was common for a number of areas and time periods, where differences among the models was large. Therefore, due to large ranges and the small sample sizes (six models), the outlier detection method cannot identify outliers in all cases. As a result, the ensemble average often includes all model estimates of ET when the ensemble has a moderate to relatively large range in ET, as a dependable estimate of the true median value cannot be determined (e.g., Figure 1 middle and lower time series graphs). Similar model differences and patterns occur at the state level, with USBR and OpenET models of eeMETRIC and SSEBop generally showing similar estimates and totals (Figure 5; Table 4). Differences among model estimates of consumptive use at the state level are generally consistent with basin-wide results.

Comparison maps of growing season ET estimates for croplands were summarized for HUC8 watersheds to illustrate spatial distributions of differences between satellite-based USBR and OpenET estimates of ET. USBR and OpenET model implementations of METRIC show differences throughout the UCRB, with the majority of HUC8 irrigated land ET averages falling within 20% of each other (Figure 6). Differences between USBR and OpenET eeMETRIC estimates are both positive and negative throughout the basin, with the largest differences occurring in areas with complex topography and limited crop acreage. More investigation is needed to better understand the causes of these differences.

Table 3. Summary comparison of 2020 growing season basin-wide irrigation consumptive use differences between the OpenET Ensemble and OpenET and USBR METRIC and SSEBop model applications.

	Acre-Ft	**Difference from Ensemble (Acre-Ft)**	**Percent Difference from Ensemble**
OpenET METRIC	2,611,377	277,550	11.9
OpenET SSEBOP	3,010,296	676,469	29.0
OpenET ENSEMBLE	2,333,827		
USBR METRIC	2,808,759	474,932	20.3
USBR SSEBOP	2,877,750	543,923	23.3
OpenET SSEBOP	3,010,296	**Difference (USBR - OpenET)**	**Percent Difference**
USBR SSEBOP	2,877,750	-132,546	-4.4
OpenET METRIC	2,611,377	**Difference (USBR - OpenET)**	**Percent Difference**
USBR METRIC	2,808,759	197,382	7.6

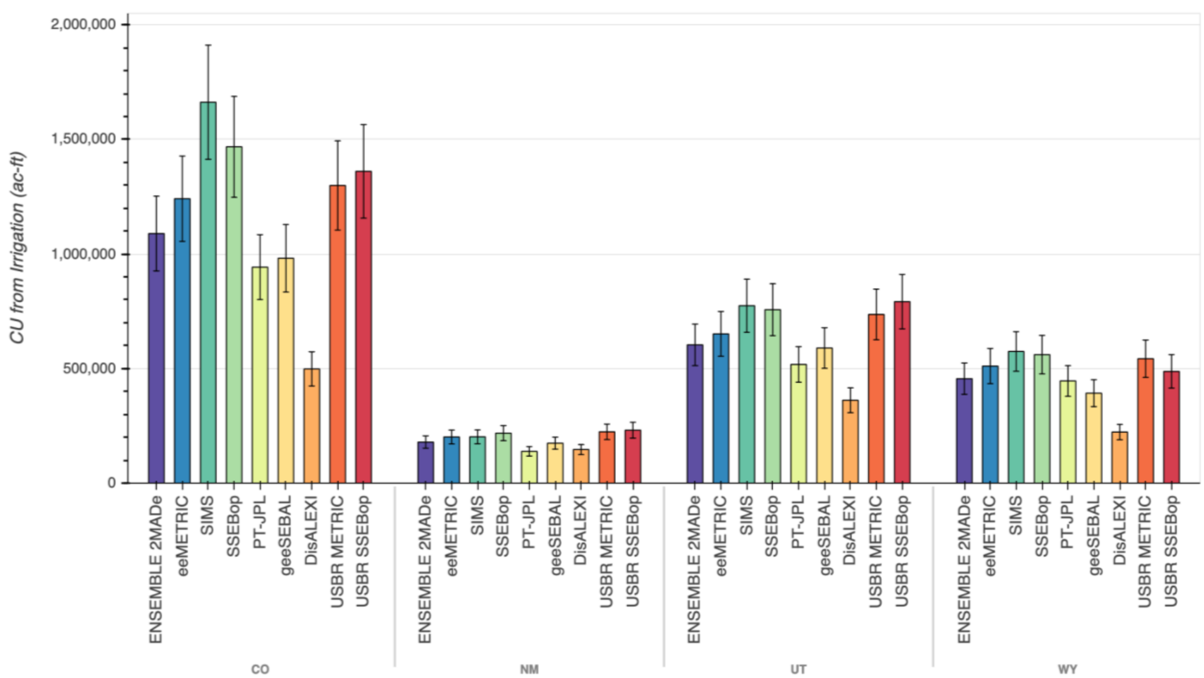


Figure 5. UCRB 2020 growing season consumptive use from irrigation totals for each model summarized by state. Potential error bars of +/- 15% are shown for all models for illustration and comparison purposes only. Irrigation consumptive use was calculated as the difference between growing season ET and annual effective precipitation from ET Demands.

Table 4. Summary of state and basin-wide 2020 growing season consumptive use totals for each model.

	CO (acre-ft)	**NM (acre-ft)**	**UT (acre-ft)**	**WY (acre-ft)**	**UCRB Total (acre-ft)**
ENSEMBLE	1,092,572	179,073	605,788	456,393	2,333,826
eeMETRIC	1,244,724	201,401	653,664	511,588	2,611,377
SIMS	1,665,748	202,044	776,639	575,364	3,219,795
SSEBop	1,471,293	218,143	759,371	561,489	3,010,296
PT-JPL	946,111	138,669	520,393	446,768	2,051,941
geeSEBAL	984,922	174,311	591,961	393,177	2,144,370
DisALEXI	501,689	146,629	363,883	223,637	1,235,837
USBR METRIC	1,302,645	223,822	738,575	543,717	2,808,759
USBR SSEBop	1,364,174	231,010	794,170	488,396	2,877,750

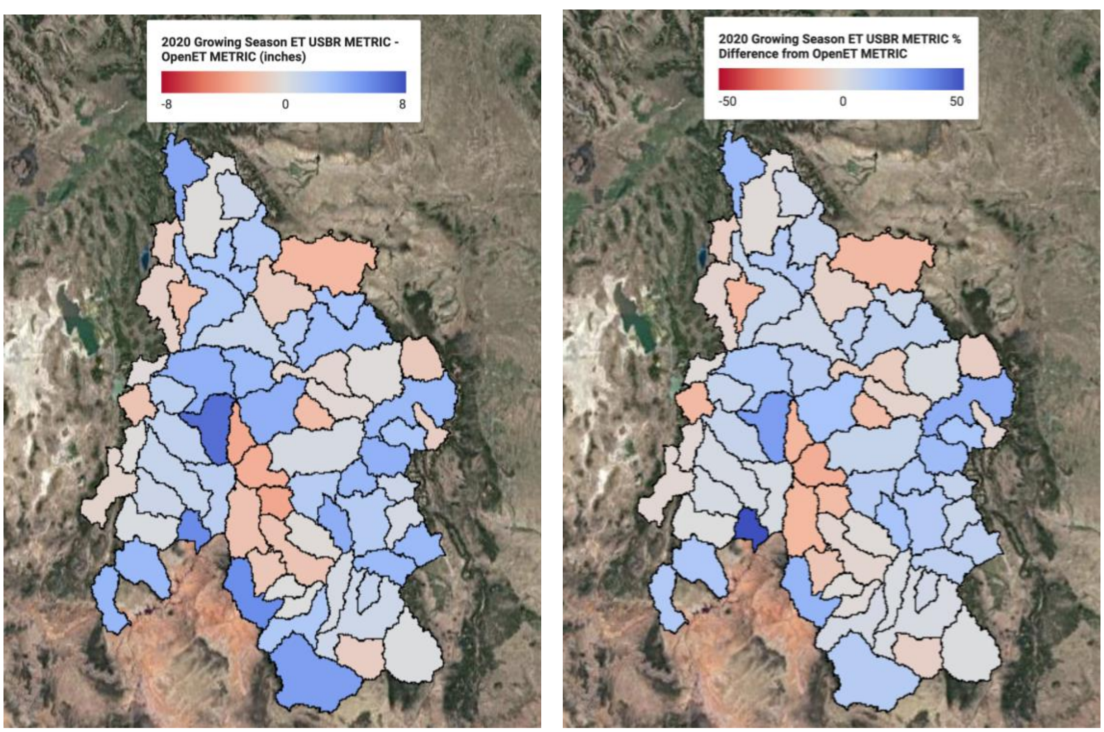


Figure 6. Maps of 2020 growing season ET differences between the USBR and OpenET eeMETRIC model estimates. The left panel shows ET differences while the right panel shows percent differences. Pixel level differences were averaged to the HUC8 level after masking to include only cropland estimates.

Differences between USBR and OpenET SSEBop estimates for 2020 tended to be more spatially consistent for the majority of HUC8s through the UCRB (Figure 7). These differences are likely due to differences in cloud screening and reference ET used for time integration between USBR and OpenET SSEBop implementations. A higher scene average cloud cover threshold used in the OpenET implementation of SSEBop (OpenET: 70%; USBR: 40%) may be causing a low bias in surface temperature within high elevation areas where clouds are prevalent, thus causing estimated ET to be biased high in these areas. However, further investigation is needed to test this hypothesis and better understand the potential causes of the differences observed.

Comparisons between USBR METRIC and SSEBop versus OpenET implementations of METRIC and SSEBop are shown in Figures 8 and 9. USBR METRIC estimates show both positive and negative differences throughout the basin relative to the OpenET ensemble average, while USBR SSEBop estimates are generally higher than the OpenET ensemble average, with the exception of high elevation HUCs in Wyoming and Colorado (Figure 9).

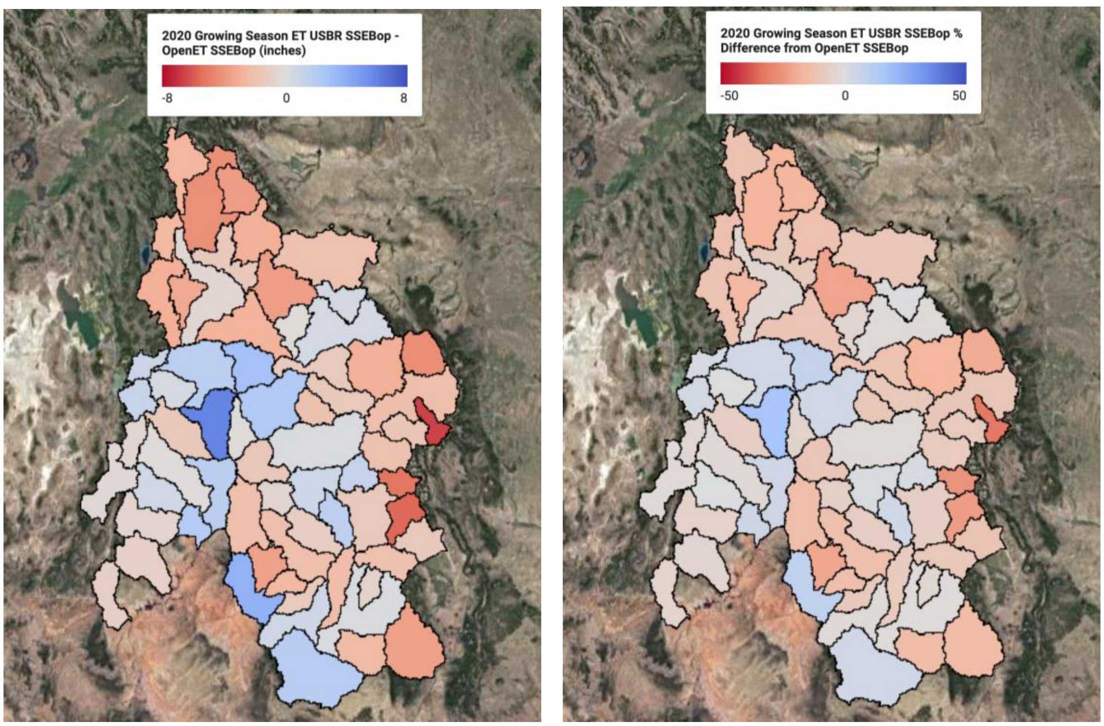


Figure 7. Maps of 2020 growing season ET differences between the USBR and OpenET SSEBop model estimates. The left panel shows absolute ET differences while the right panel shows percent differences. Pixel level differences were averaged to the HUC8 level after masking to include only cropland estimates.

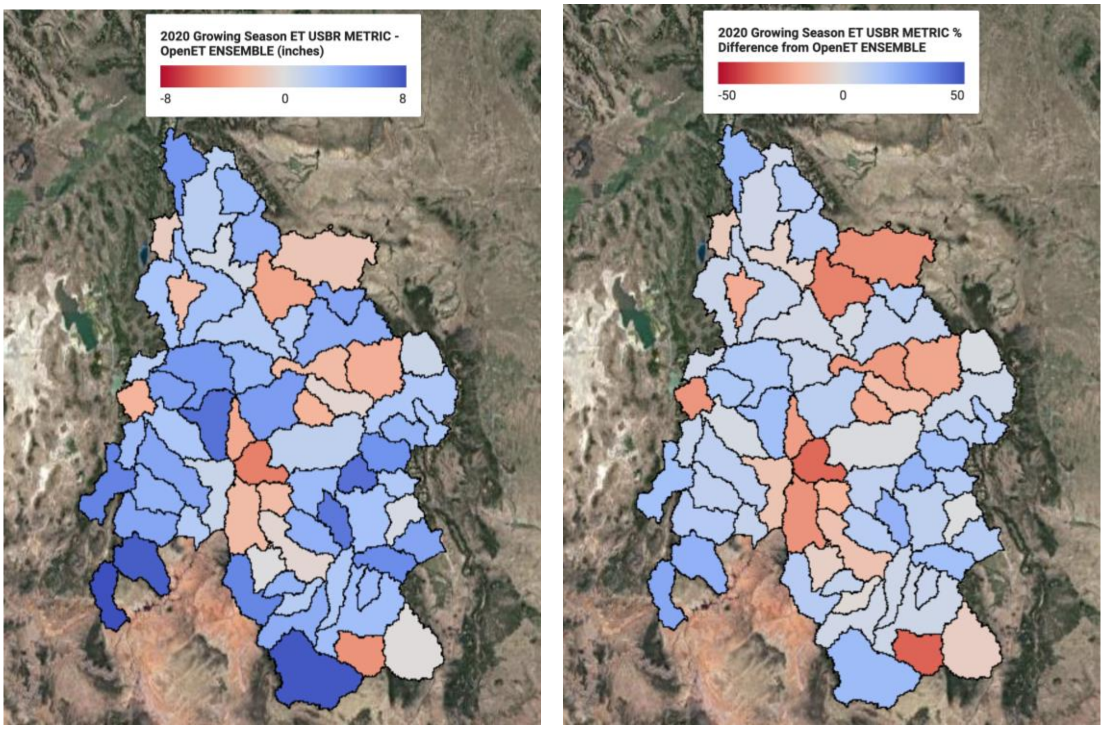


Figure 8. Maps of 2020 growing season ET differences between the USBR METRIC and OpenET MAD Ensemble model estimates. The left panel shows absolute ET differences while the right panel shows percent differences. Pixel level differences were averaged to the HUC8 level after masking to include only cropland estimates.

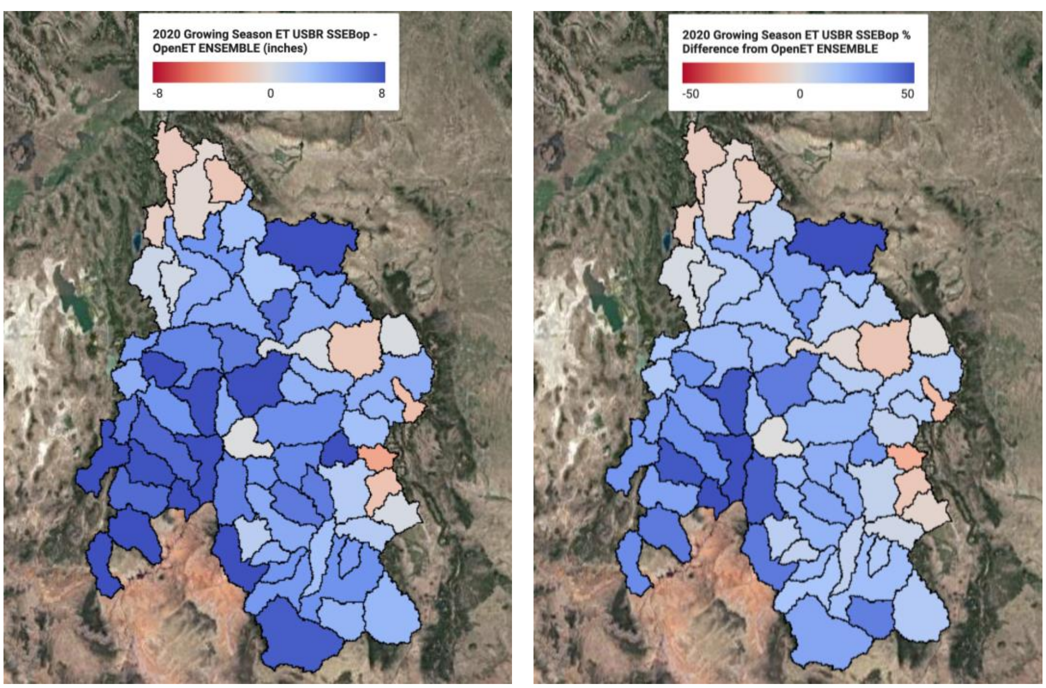


Figure 9. Maps of 2020 growing season ET differences between the USBR SSEBop and OpenET MAD Ensemble model estimates. The left panel shows absolute ET differences while the right panel shows percent differences. Pixel level differences were averaged to the HUC8 level after masking to include only cropland estimates.

Comparison with UCRB Eddy Covariance Stations

Regression analyses between OpenET eeMETRIC and SSEBop and EC data collected between 2018 and 2020 show relatively good correlation and accuracy (Figures 10-13). Bias metrics, indicated by the slope of the best fit line through zero (slope), show that OpenET eeMETRIC and SSEBop results fall within 8% (slope values of 0.92 to 1.08) of in-situ closed energy balance ET estimates at all sites with the exception of Big Piney, WY. Energy balance closure calculations are not possible at the Big Piney station due to persistent flood irrigated conditions that caused inaccurate soil heat flux measurements. To provide comparisons more in line with results from other stations, closed EC estimates at Big Piney were estimated by adjusting the monthly unclosed estimates by the average monthly closure adjustment from two of the three other UCRB EC monitoring stations (average closure adjustment of 1.23). The average monthly closure for 2018-2020 data within the growing season (April through October) for Vernal and Palisade stations ranged from 0.75 to 0.77 with relatively consistent results from month to month, however, Bloomfield closure results were much lower (around 66 percent) and were therefore not included in the adjustment for the Big Piney station. eeMETRIC and SSEBop results were within 16% of Big Piney EC estimates after applying the closure adjustment. OpenET eeMETRIC and SSEBop show high correlation (r-squared values of 0.80 to 0.96) with EC results indicating good skill in predicting ET variability over time. RMSE (root mean square error) values for OpenET eeMETRIC and SSEBop ranged from 0.55 to 1.35 inches per month. Importantly, both OpenET eeMETRIC and SSEBop produce results with similar accuracy to USBR manual and automated METRIC and SSEBop model implementations.

The difference between the daily and monthly EC station source area sampling methods for USBR model ET typically shows a low to moderate improvement in accuracy using daily sampling, particularly for SSEBop and RMSE and r-squared values (Figures 10-13). Using daily ET estimates and source areas to calculate monthly ET totals is expected to result in improved accuracy compared to monthly aggregated source areas and ET because the daily variability in source areas and ET are explicitly estimated. There are a few exceptions where the accuracy statistics are slightly improved using the OpenET monthly footprint, such as for manual METRIC at the Bloomfield EC station.

OpenET ensemble averages are in line with both METRIC, SSEBop, and EC estimates with RMSE values of approximately 0.47 to 1.82 inches/mo. R-squared values ranged from 0.89 to 0.95 and slope through zero values ranged from 0.89 to 0.95. The OpenET ensemble average is generally biased low at all sites (slopes < 1). Low ET estimates from PT-JPL, geeSEBAL, and DisALEXI influence the ensemble mean in areas where large ranges in ET estimates exist, and therefore no outliers are eliminated by the MAD outlier detection approach. These lower estimates are somewhat balanced by the estimates from SIMS, which are generally on the upper end of the ensemble range, but still result in low ET estimates relative to the closed ET estimates at the four EC locations (i.e., slopes <1). Given this low bias trend at the EC stations (Figures 11-13), it is likely that the OpenET ensemble average is biased low across the UCRB (Figures 4-5). Performance and findings from OpenET implementations of SIMS, PT-JPL, and DisALEXI are described in more detail below.

Overall, SIMS agrees well with EC ET estimates, with the slope of the best fit line ranging from 0.96 to 1.10, RMSE values of 0.55 to 1.16 inches per month, and r-squared values of 0.82 to 0.95 (Figures 10-13). However, these sites are all generally well-irrigated, and many fields within the UCRB experience water limitations due to farming practices and during periods of shortage. SIMS was originally developed to support irrigation scheduling and uses a reflectance-based vegetation index approach to estimate ET that assumes that croplands are well-irrigated. As a result, SIMS is expected to overestimate ET for fields that experience intermittent or short-term deficit irrigation that may not result in sustained declines in the extent and condition of the crop canopy. While the SIMS consumptive use estimate is the highest of the OpenET ensemble for UCRB, estimates are within 7% of the OpenET SSEBop model, and 12% of the USBR METRIC and USBR SSEBop models for 2020 for the entire UCRB. SIMS also agrees reasonably well with these models in NM, WY, and UT. In areas and time periods where water limitations are not wide-spread, SIMS is expected to perform and compare well as illustrated from comparisons to EC, METRIC, and SSEBop ET estimates. However, during periods when water limitations are expected to be widespread, it would be reasonable to exclude SIMS from the OpenET model ensemble since it tends to estimate ET under well-watered conditions.

Unlike other models, PT-JPL, geeSEBAL, and DisALEXI consistently underestimate ET at all EC stations with the exception of Big Piney, WY (Figures 10-13). Low ET estimates are also evident in basin and state-wide consumptive use comparisons, where PT-JPL, geeSEBAL, and DisALEXI estimates ranged from 40% to 83% of the OpenET ensemble ET value and USBR METRIC and SSEBop totals. The PT-JPL model relies on the Priestley-Taylor wet environment potential ET equation as an upper limit on estimated ET (Priestley and Taylor, 1972). Priestley-Taylor (PT) evapotranspiration was originally developed to represent wet environment potential ET from a “horizontally uniform saturated surface (land and water)”, sufficiently extended to obviate any significant advection of energy from outside the area of interest (Priestley and Taylor, 1972; Lhomme et al., 1996). In arid and semi-arid environments, ET from well-watered areas is commonly well above the PT potential ET rate due to advection of warm and dry air from dry regions toward well-watered surfaces. Several studies have discussed the limitation of the PT equation and underestimation of ET for irrigated environments in arid to semi-arid areas (Jensen et al., 1990; McAneney and Itier, 1996; Weiß and Menzel, 2008; Tabari and Talaee, 2011). To improve PT-JPL ET estimates for croplands, wetlands, and riparian areas in arid and semiarid environments where advection is prevalent, a PT aridity adjustment spatial layer was developed based on the ratio of ASCE reference ET and PT ET, following principles of the complementary relationship of evaporation (Kahler and Brutsaert, 2006; Szilagyi, 2007; Huntington et al., 2011). ASCE grass reference ET and PT ET estimates were developed using the gridMET dataset. The alpha adjustment layer was calculated as the ratio of the growing season average bias-corrected ASCE grass reference ET to the growing season average PT ET. Results from the OpenET Accuracy Assessment and Intercomparison study (Melton et al., 2021) indicated that application of the PT aridity adjustment layer increased overall accuracy of PT-JPL for cropland, wetland, and riparian areas, however, ET estimates generally remain near or below the ensemble average for these areas. Even after this aridity correction, PT-JPL estimates still fall well below the 1:1 line when compared to EC station estimates of ET within the UCRB, and are consistently below METRIC, SSEBop, and SIMS consumptive use estimates across the UCRB and for each state (Figures 11-13 and 4-5).

Estimates from geeSEBAL and DisALEXI are also consistently low when compared to the EC data and the ensemble estimate of consumptive use across the UCRB. Potential issues related to geeSEBAL automated hot and cold pixel selection and internal calibration have been identified as a potential cause for the relatively low estimates of ET, along with the fact that within geeSEBAL, sensible heat flux is assumed to be at or near zero for the cold pixel well-watered field condition. While this assumption is valid for more moderate to humid regions, it is not a robust assumption for all conditions within arid and semi-arid regions such as the UCRB. Sensible heat flux for well-watered areas within arid and semi-arid environments is often negative, resulting in ET being greater than available energy. Because geeSEBAL does not allow for ET to be greater than available energy, it tends to underestimate ET, as shown from comparisons to the EC station data (Figures 11-13) and intercomparison of the OpenET ensemble of models (Figures 4-5). OpenET and partners plan to enhance geeSEBAL by improving performance in complex terrain, and allowing for negative sensible heat flux (i.e., advection) in future implementations of geeSEBAL within OpenET.

Based on biases and inconsistencies occurring during OpenET intercomparisons, a geolocation issue was identified with the gridding of the GOES thermal imagery used in ALEXI, which produces the baseline 4 km ET results that are disaggregated to 30 m scale with DisALEXI. This issue is exacerbated with the transition to the GOES-R satellite series in 2018 for GOES-West, most prominently impacting ALEXI from 2020-present. Misregistration of LST and LAI inputs to ALEXI preferentially reduces ET fluxes over landscapes with small features of high contrast in temperature and vegetation cover, a key characteristic of irrigated areas in the UCRB. This registration problem is being addressed for implementations of disALEXI within OpenET. In addition, corrections accounting for topographic variability in LST and net radiation, as well as adjustment of the soil resistance factor for sparsely vegetation surfaces, will be implemented in ALEXI and DisALEXI and will further increase ET in the UCRB region.

OpenET intercomparison and Accuracy Assessment

In addition to the four UCRB EC station comparisons, OpenET has conducted one of the largest intercomparison and accuracy assessments of field-scale satellite-driven ET models to date (Melton et al., 2021). Satellite-derived OpenET data were compared against ET measurements collected by 142 EC flux stations and four precision weighing lysimeters located across the continental U.S. The key results for croplands are summarized below (Table 5), and information for other land cover types is available at https://openetdata.org/accuracy. Full details will be provided in the forthcoming OpenET Intercomparison and Accuracy Assessment Report, along with the technical details of the methods used in the assessment.

For croplands, the ensemble average performed as well as or better than any individual model across most accuracy metrics, though nearly all models demonstrated high accuracy measures for croplands. Accuracy metrics for the slope of the best fit line through the origin (slope), mean bias error (MBE), mean absolute error (MAE), root mean squared error (RMSE), and R-squared (r^2) value are summarized in the table below for the water year (October-September) and growing season timesteps. Only 14 sites were compared for water years due to strict restrictions on total numbers of missing days allowed during a year. Descriptions of each of these accuracy assessment metrics and how to interpret them are provided in the *Explanation of Accuracy Metrics* supplemental section below.

At annual timescales, 14 sites had at least one full water year of data with 48 total water years, and the OpenET ensemble value had a slope of 0.93, and MAE and RMSE values of <10%, with an R-squared value of 0.90, demonstrating excellent overall agreement with the flux station data. The MBE was -7.0%, which is still very good, but indicates a negative 7 percent bias in the ensemble ET value relative to the ground-based ET, and this bias should be accounted for in evaluating annual ET totals.

For the growing season, 38 sites had at least one complete growing season with a total of 151 complete growing seasons across these sites. The OpenET ensemble value had a slope of 1.0, MBE of -1.7%, MAE of 13.2%, RMSE of 15.2%, and R-squared of 0.88. These results demonstrate very strong agreement between the ensemble value and the ground-based ET datasets for growing seasons. The slope of 1.0, MBE of -1.7%, and R-squared of 0.88 indicates that on average, the ensemble ET value demonstrates good accuracy with an average error of less than 2% across a wide range of crops and meteorological conditions. It should also be noted that there were far fewer cropland EC stations available in more arid and semi-arid regions across the western U.S. If more stations were available, it would be anticipated that the OpenET ensemble slope would be lower than shown in Table 5 due to the low bias of several OpenET models in arid and semi-arid areas as demonstrated for the UCRB. Unfortunately, none of the UCRB EC stations meet the OpenET energy balance closure criteria of having greater than 75% closure or inclusion in OpenET's Intercomparison and Accuracy Assessment.

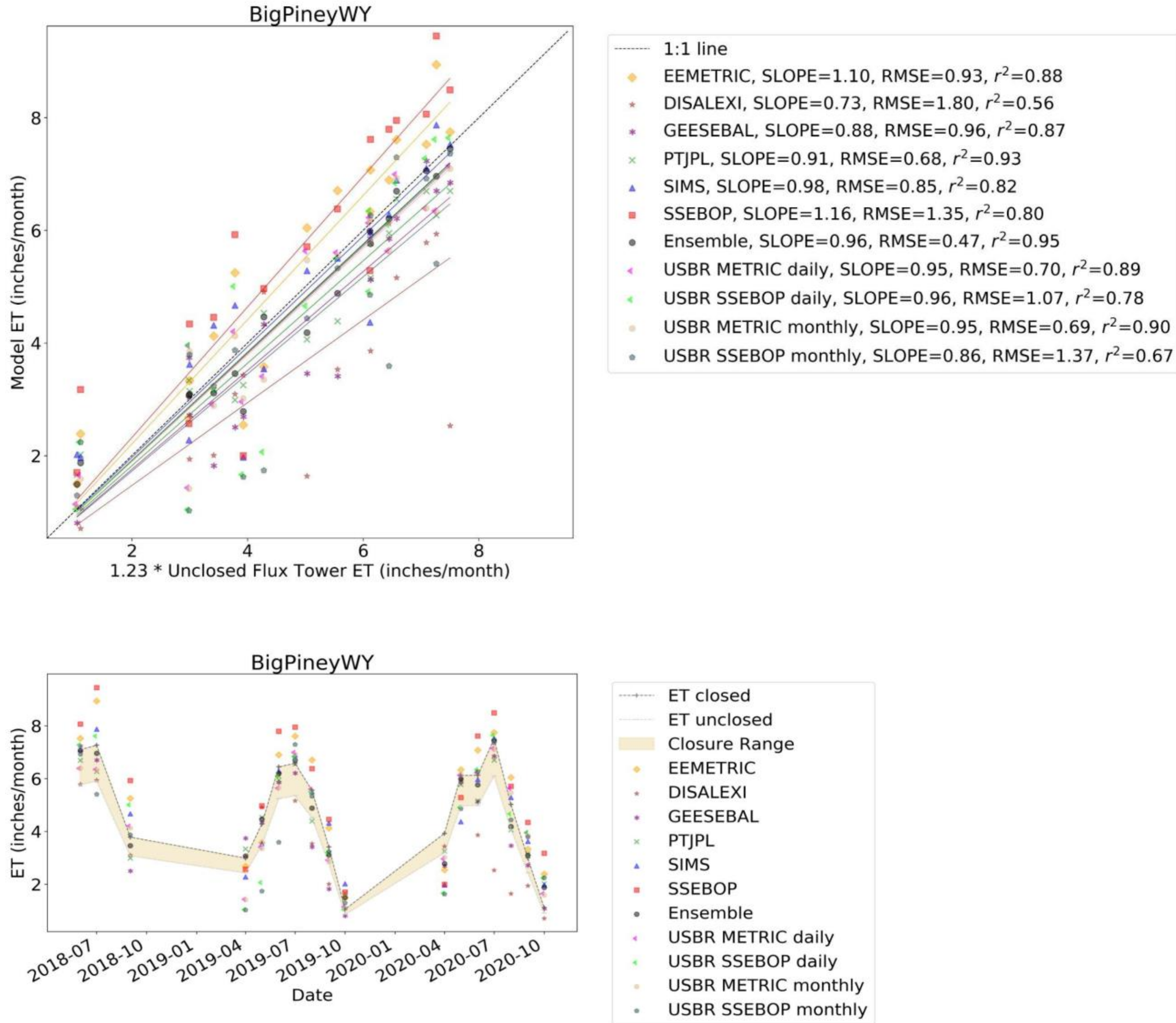


Figure 10. Scatter plot (top) and time series (bottom) comparisons of monthly growing season (Apr-Oct) UCRB EC results to each ET model for the Big Piney, WY monitoring station from 2018-2020. Closed energy balance results were estimated by applying the average closure adjustment of 1.23 from the Vernal and Palisade UCRB EC monitoring sites. USBR “daily” and “monthly” values refer to results generated using data extraction methods that relied on daily and monthly EC source areas, respectively. All OpenET results utilized monthly EC source areas for data extraction.

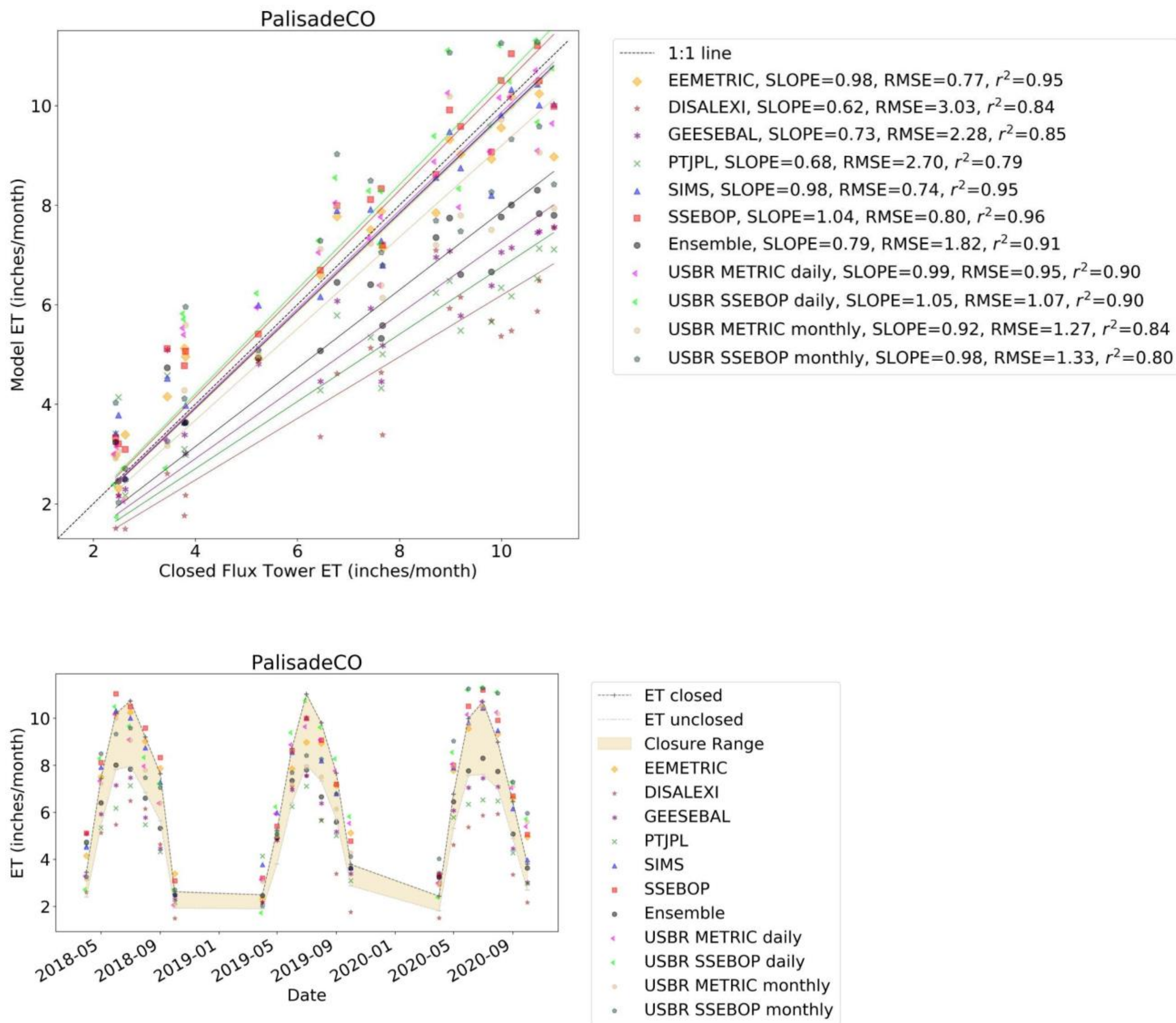


Figure 11. Scatter plot (top) and time series (bottoms) comparisons of monthly growing season (Apr-Oct) UCRB EC results to each ET model for the Palisade, CO monitoring station from 2018-2020. USBR "daily" and "monthly" values refer to results generated using data extraction methods that relied on daily and monthly EC source areas, respectively. All OpenET results utilized monthly EC source areas for data extraction.

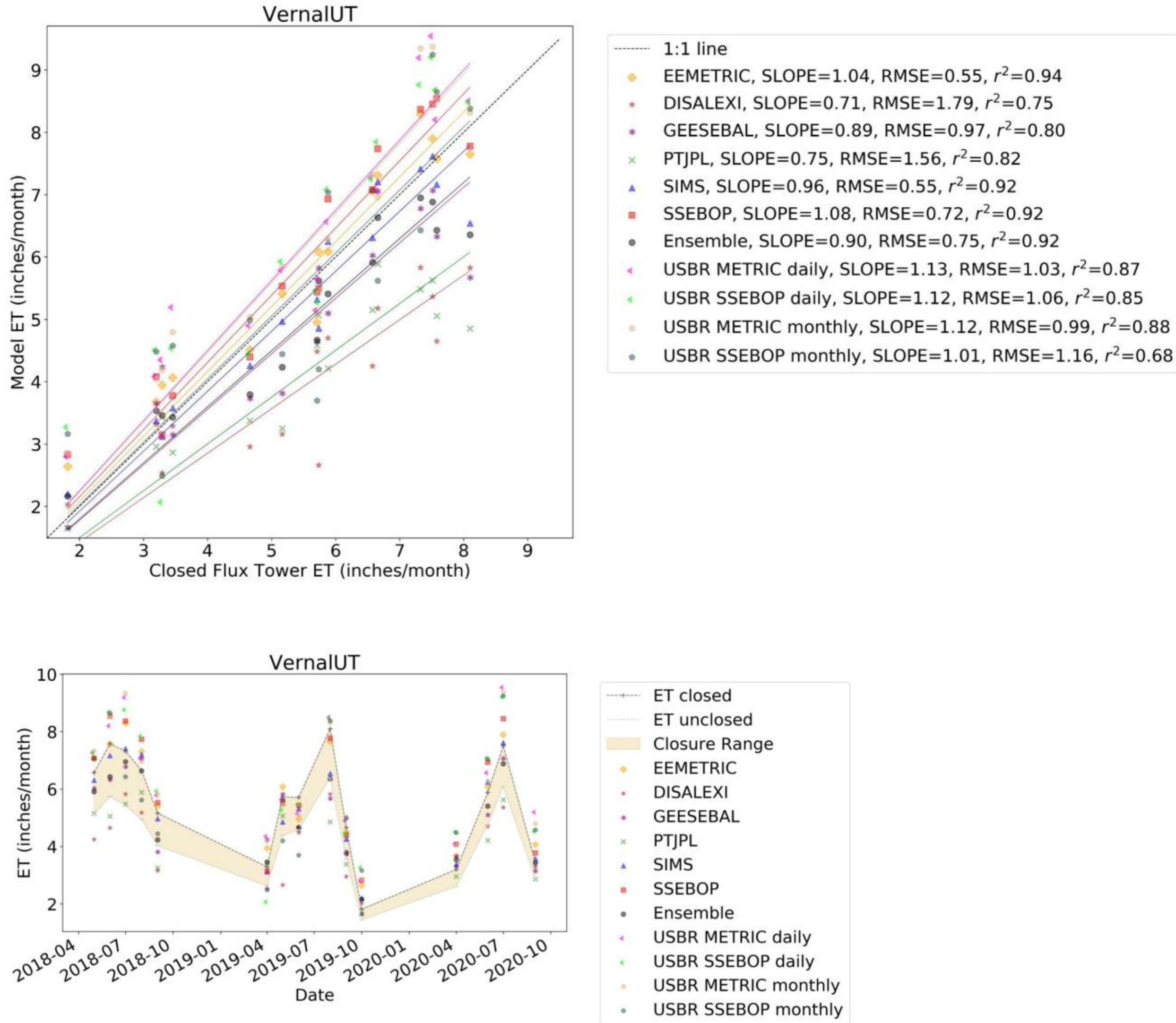


Figure 12. Scatter plot (top) and time series (bottom) comparison of monthly growing season (Apr-Oct) UCRB EC results to each ET model for the Vernal, UT monitoring station from 2018-2020. USBR “daily” and “monthly” values refer to results generated using data extraction methods that relied on daily and monthly EC source areas, respectively. All OpenET results utilized monthly EC source areas for data extraction.

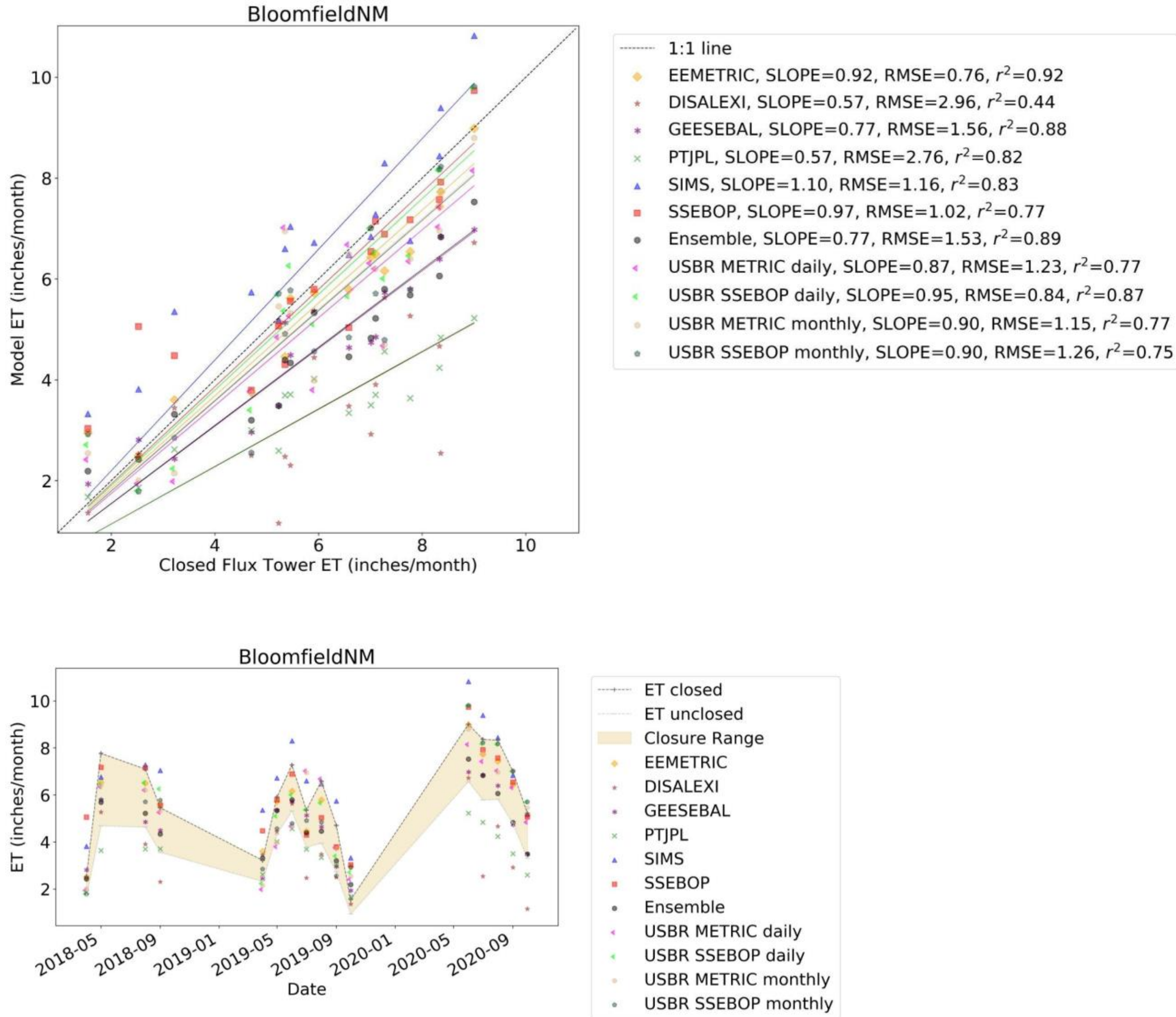


Figure 13. Scatter plot (top) and time series (bottom) comparisons of monthly growing season (Apr-Oct) UCRB EC results to each ET model for the Bloomfield, NM monitoring station from 2018-2020. USBR "daily" and "monthly" values refer to results generated using data extraction methods that relied on daily and monthly EC source areas, respectively. All OpenET results utilized monthly EC source areas for data extraction.

Table 5: OpenET ensemble cropland accuracy summary statistics showing both water year and growing season time periods.

Time Period	Slope	Mean Bias Error	Mean Absolute Error	Root Mean Squared Error	R-Squared	Mean flux station ET
Water Year: 14 sites with 48 total water years)	0.93	-2.81 in (-7.0%)	3.59 in (8.9%)	3.96 in (9.8%)	0.90	40.3 in
Growing Season: 38 sites with 151 growing seasons	1.0	-0.398 in (-1.7%)	3.16 in (13.2%)	3.65 in (15.2%)	0.88	24.0 in

CONCLUSIONS

The lack of consistent, accurate information on ET and consumptive use of water by irrigated agriculture is one of the most important data gaps for water managers in the western U.S. The ability to easily access information on ET is central to improving water budgets, advancing the use of data-driven water management strategies, and expanding incentive-driven conservation programs. Recent advances in remote sensing of ET have led to the development of multiple approaches for field-scale ET mapping that have been used for local and regional water resource management applications by US state and federal agencies. The OpenET project is a community-driven effort that is building upon these advances to develop an operational system for generating and distributing ET data at field scale using an ensemble of six well-established satellite-based approaches for mapping ET. Key objectives of OpenET include increasing access to remotely sensed ET data through a web-based data explorer and data services, supporting use of ET data for a range of water resource management applications, and development of use cases and training resources for agricultural producers and water resource managers.

OpenET is built upon the Google Earth Engine platform using open source software tools to increase transparency and to facilitate collaboration across three federal agencies, six ET modeling teams, and a wide range of public and private entities collaborating with OpenET as use case partners and advisors. Google Earth Engine allows for efficient processing, storage, scalability in time and space, and automated operational updates with new observations that would otherwise be costly and burdensome for State and federal agencies to develop, operate, and maintain. This enhanced efficiency and utility of OpenET is evident in the fact that the SSEBop team utilized the OpenET framework to develop estimates of ET for this study, resulting in reduced staff time and costs for development of ET datasets across the UCRB. The collaborative, community-driven effort has accelerated progress on alignment of model inputs and pre-processing routines, and completion of a joint model intercomparison and accuracy assessment. The ability to easily compare model results at scale have accelerated the ability of the ET modeling community to identify and understand differences across the ensemble of approaches used by OpenET, as highlighted in this intercomparison report for the UCRB.

Conclusions from this study are based on 1) comparisons against ET estimates from four eddy covariance (EC) flux stations in the UCRB; 2) model-to-model comparisons within the OpenET ensemble and against the USBR-supported implementations of METRIC and SSEBop; and 3) an understanding of the representation of physical processes within each model. The following conclusions acknowledge the challenges associated with quantifying the accuracy of satellite-based ET models using only four EC stations to represent over 700,000 acres of irrigated land. Based on this intercomparison study for irrigated lands within the UCRB, along with results from the broader OpenET accuracy assessment, there is strong evidence that the OpenET implementations of the eeMETRIC and SSEBop models are currently performing well and providing accurate data for the UCRB, with uncertainties in growing season and annual ET of ± 15% or better. OpenET eeMETRIC and SSEBop demonstrated good performance when compared to UCRB EC stations, and showed temporal and spatial consistency, and compared as expected to other OpenET models throughout the UCRB given the underlying assumptions and current shortcomings of other OpenET models. Results from manual USBR METRIC and automated SSEBop model implementations showed good agreement with UCRB EC station data and OpenET implementations of these models. Differences in model calibration, reference ET sources and reference ET types, and satellite image filtering contributed to the observed differences between USBR and OpenET model results for METRIC and SSEBop.

Model physics, assumptions, intended uses, and study area climatic conditions currently limit the performance or consistency of some of the satellite-based models within the UCRB, despite good agreement with the EC data for some models. These limitations are primarily due to arid to semi-arid conditions and frequent water shortages that occur within irrigated lands of the UCRB. SIMS was among the best performing models at the relatively well-watered UCRB EC stations. Those UCRB stations were located within generally well-managed and well-irrigated areas over the study period, and representative of conditions for which SIMS was specifically developed. SIMS uses a reflectance-based vegetation index approach to estimate ET that assumes that croplands are well-irrigated, and was originally developed to support irrigation scheduling. While many fields within the UCRB are well-watered, many experience water limitations due to farming practices and periods of shortage. As a result, SIMS is expected to overstate ET for fields that experience intermittent or short-term deficit irrigation that do not result in sustained declines in the extent and condition of the crop canopy. While the SIMS consumptive use estimate is the highest of the OpenET ensemble for the UCRB for 2020, estimates are within 7% of the OpenET SSEBop model, and 12% of the USBR METRIC and USBR SSEBop models at the basin scale. OpenET results from PT-JPL, geeSEBAL, and DisALEXI indicate that these models were biased low when compared to UCRB EC station estimates of ET. These models also had the lowest consumptive use estimates across the UCBR and for each state. Low biases may be due to factors related to aridity, advection, complex topography, and disaggregation of regional ET information, though further investigation is needed.

The OpenET ensemble average is intended to better inform practitioners regarding ET model agreement and disagreement, and to take advantage of strengths of the different ET methods across different regions and land cover types. The use of multiple models allows for efficient identification of anomalies and outliers through intercomparison, especially in areas with limited to no in-situ monitoring. From close inspection of the OpenET ensemble average, median, individual model ET estimates across space and time, and comparisons to EC station estimates, it is clear that all models can, at times, produce erroneous ET estimates, and that these errors can be both random and systematic. The OpenET team applied the Median Absolute Deviation (MAD) approach to identify and eliminate model outliers at monthly time steps, and for every pixel. However, when the ensemble sample size is small and range of modeled ET is large, the utility of the MAD outlier detection approach (and others) is limited, and outliers or models with known systematic biases are often not removed from the ensemble average. As a result, it is possible for models with local or regional systematic biases to be included in the calculation of the OpenET ensemble value for some locations or time periods. At this time, based on the results of the analyses performed for the UCRB, the OpenET ensemble average is not recommended for use within the UCRB. As improvements are made to individual models over the coming year, this recommendation may change. In addition, an important aspect and benefit of OpenET is access to all model results, so that users may create their own custom ensemble products using the best models for specific use cases and areas of interest.

Future OpenET efforts will continue to focus on model improvements and intercomparisons throughout the U.S. Over the next year, the NASA Western Water Applications Office (WWAO), in coordination with the OpenET team, will be conducting a more detailed intercomparison and accuracy assessment for the UCRB than the one presented in this report. In closing, the most promising qualities of OpenET for stakeholders in the UCRB are increased transparency, better understanding of the strengths and weaknesses of different modeling approaches, a community to support further research and development, and supporting operational production of ET data that can be generated in a consistent way, at scale and at low cost. Most importantly, OpenET provides data that are reproducible in time and space. The OpenET team looks forward to continued collaboration with the UCRB Consumptive Use Feasibility team moving ahead.

SUPPLEMENTAL INFORMATION AND FIGURES

Explanation of Accuracy Metrics

OpenET selected five key metrics to characterize the accuracy of OpenET data:

Slope: The slope of the best-fit line forced through the origin provides a measure of overall agreement between the satellite data and ground-based ET data. A slope of less than 1.0 indicates that the satellite ET data are, on average, less than the ground-based ET data. A slope greater than 1.0 indicates that, on average, the satellite ET data are greater than the ground-based ET data. A slope of 1.0 indicates no overall bias, and values between 0.9 and 1.1 are generally considered excellent for ET estimation using any method.

Mean Bias Error (MBE): As the name suggests, MBE provides a measure of overall bias and is closely related to the slope. MBE quantifies the bias in the satellite ET data relative to the ground-based ET data, and characterizes the expected overall error when the remotely sensed ET data are aggregated over large areas. A negative bias indicates that the satellite ET data are, on average, less than the ground-based ET data. A positive bias indicates that, on average, the satellite ET data are greater than the ground-based ET data. A value of zero would represent no bias, and values in the range of +/- 10% are considered excellent for remotely sensed ET.

Mean Absolute Error (MAE): MAE provides a measure of the expected error for a given location and time period, and quantifies the average absolute difference between the satellite and ground-based ET data. Since MAE provides the absolute error, the value is often considered to characterize typical expected error both above and below the reference data. A value of zero represents perfect agreement and values in the range of 0-15% (interpreted as up to +/-15%) are considered excellent for remotely sensed ET data over agricultural lands. Accuracy requirements for remotely sensed ET vary by application, however, values in the range of 0-25% are considered to be very good and acceptable for many applications. MAE includes random error in both the satellite-based ET data and the ground-based ET data that are used as a reference. When MBE is much lower than MAE, it may indicate that random error in the satellite data is the primary contributor to error, or that error in the ground-based reference ET is also contributing to the MAE. When MAE and MBE are similar in magnitude, it usually indicates that error is due to a persistent bias in the satellite-based ET data.

Root Mean Squared Error (RMSE): RMSE is another widely used measure of the expected error for a given time and location. It is interpreted in the same way as MAE, but provides greater weight to outliers or large ‘misses’ in the satellite-based ET data. RMSE may be used as a measure of expected error when the intended use of the data is sensitive to large errors in the satellite data for any given location or time period. A value of zero represents perfect agreement, and values in the range of 0-25% RMSE are generally considered to be very good for satellite-based ET data.

r-squared: The r-squared value provides a measure of the proportion of the variance in the ground-based ET data that is explained or reproduced by the satellite-based ET data. It can be interpreted as a measure of the ability of the satellite-based ET to accurately estimate the relative variation in the ground-based ET data. A value of 1.0 would mean that the satellite-based ET data explains 100% of the variance in the ground-based ET data. Values above 0.6 are considered to be good, and values above 0.85 for remotely sensed ET data are considered to be excellent. It is important to note that a model can have an r-squared value near 1.0, and still have a consistent high or low bias.

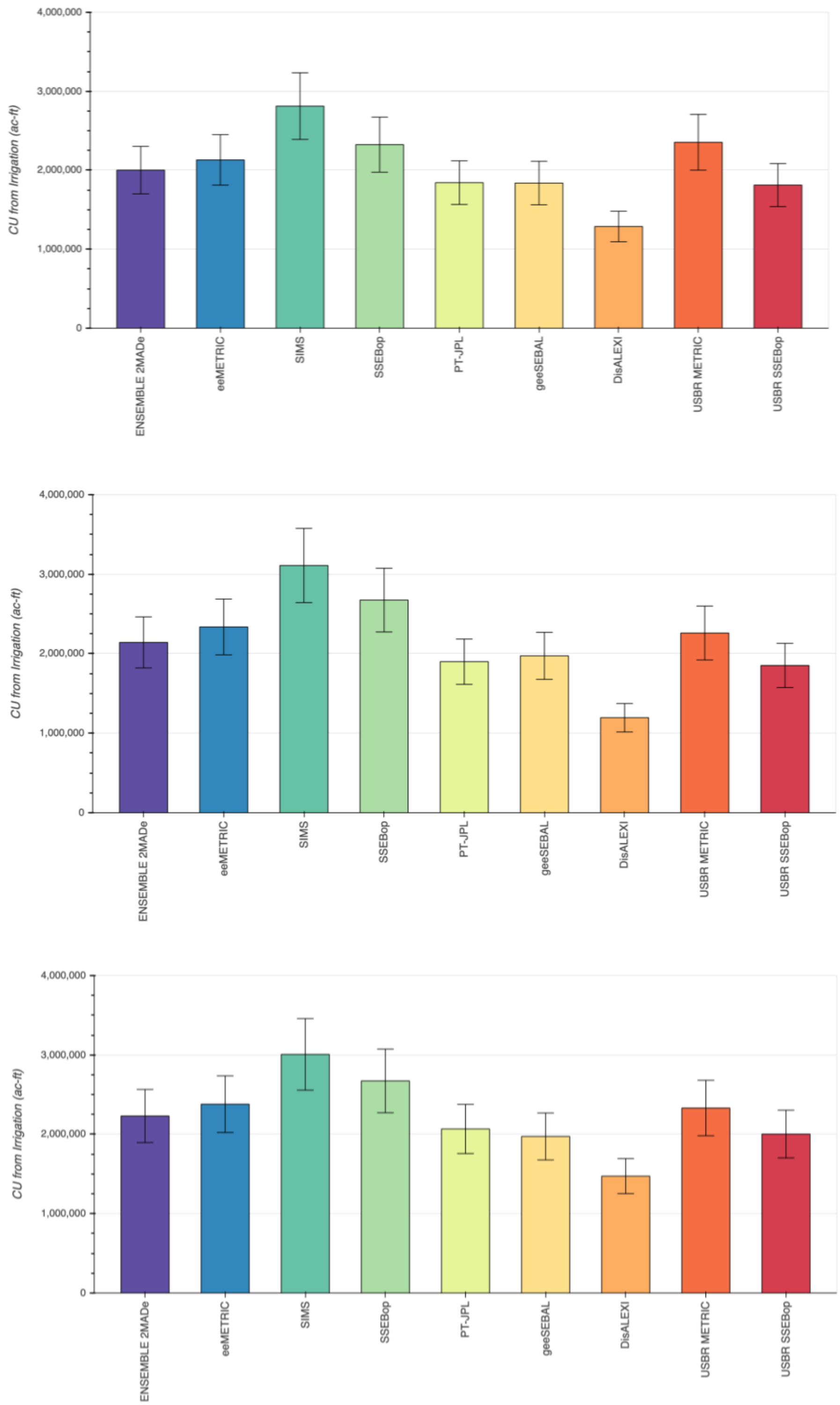


Figure S1: Basin-wide 2017 (top), 2018 (middle), and 2019 (bottom) growing season totals of consumptive use from each model. +/- 15% error bars are shown for illustration and comparisons purposes.

.

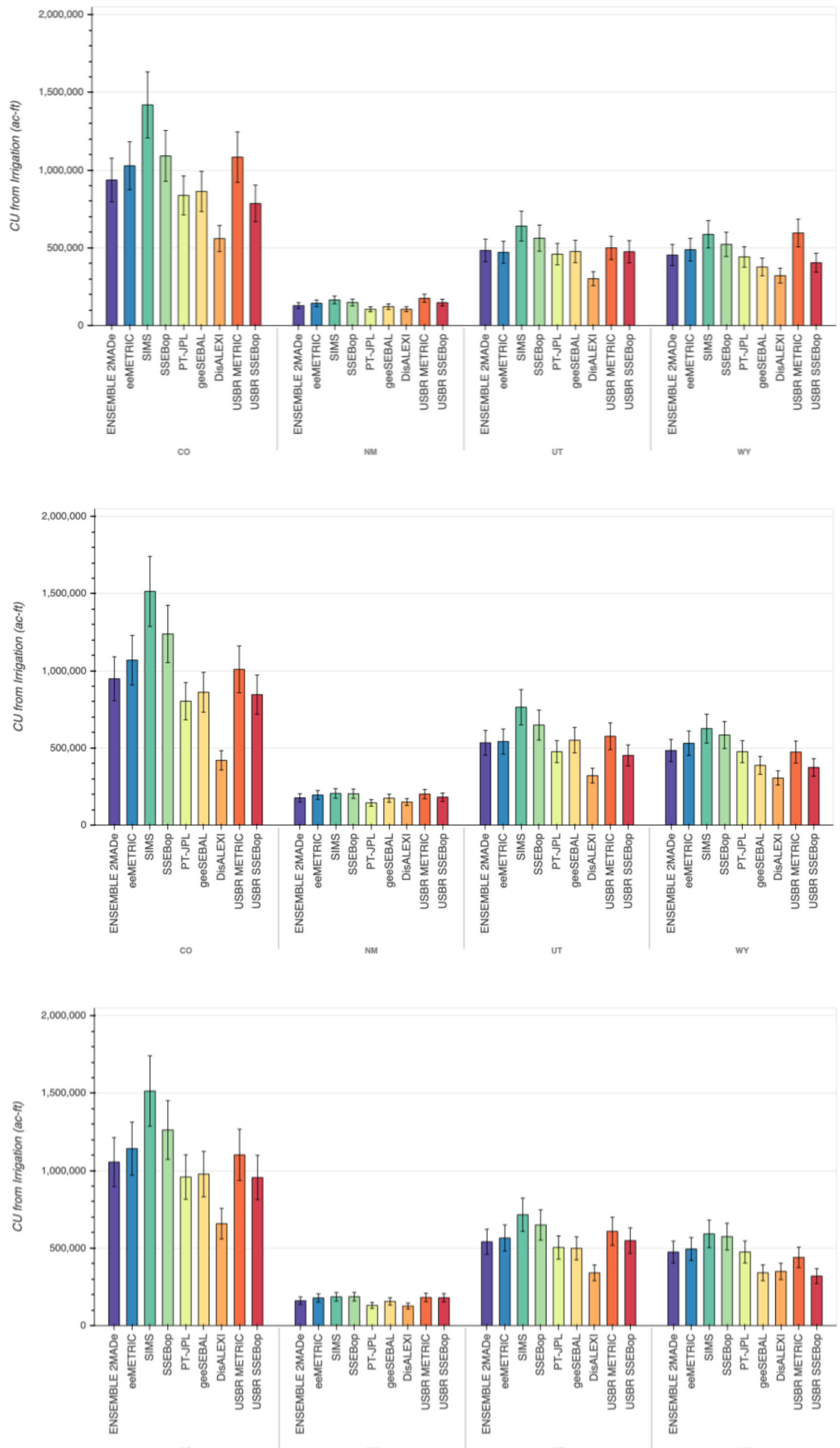


Figure S2: 2017 (top), 2018 (middle), and 2019 (bottom) state growing season totals of consumptive use from each model. +/- 15% error bars are shown for illustration and comparisons purposes. Precipitation.

ACKNOWLEDGEMENTS

The authors of this Appendix would like to acknowledge financial support from USBR and OpenET to conduct this analysis. The authors would also like to thank OpenET team members listed below for their thoughtful contributions and feedback, and for making OpenET what it is.

Robyn Grimm[4], Jamie Herring[5], Maurice Hall[4], Dana Rollison[4], Tyler Erickson[6], Martha Anderson[8], Joshua B. Fisher[9], Ayse Kilic[10], Gabriel Senay[11], Christopher Hain[12], Lee Johnson[1,2], Anderson Ruhoff[13], Philip Blankenau[14], Matt Bromley[3], Will Carrara[1,2], Britta Daudert[3], Conor Doherty[15], Christian Dunkerly[3], MacKenzie Friedrichs[11], Alberto Guzman[1.2], Gregory Halverson[16], Jody Hansen[3], Jordan Harding[5], Yanghui Kang[17], David Ketchum[18], Charles Morton[3], Samuel Ortega-Salazar[10], Thomas Ott[3], Mutlu Ozdogan[17], Peter Revelle[7,10], Mitch Schull[8,19], Carlos Wang[20], Yun Yang[8,19], Ray G. Anderson[8]

[1]NASA Ames Research Center, [2]California State University Monterey Bay, [3]Desert Research Institute, [4]Environmental Defense Fund, [5]Habitat Seven, [6]Google Inc., [7]University of Idaho, [8]USDA Agricultural Research Service, [9]Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, [10]University of Nebraska Lincoln, [11]United States Geological Survey, [12]NASA Marshall Space Flight Center, [13]Universidade Federal do Rio Grande do Sul, [14]Idaho Department of Water Resources, [15]Stanford University, [16]Jet Propulsion Laboratory, California Institute of Technology, [17]University of Wisconsin, [18]University of Montana, [19]University of Maryland, [20]University of California, Berkeley

REFERENCES

Abatzoglou, J.T., 2013. Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, *33*(1), pp.121-131.

Ali, M. H., & Mubarak, S., 2017. Effective rainfall calculation methods for field crops: an overview, analysis and new formulation. *Asian Research Journal of Agriculture*, 1-12.

Allen, R. G., & Robison, C. W. (2007). Evapotranspiration and net irrigation water requirements for Idaho. University of Idaho reported submitted to Idaho Department of Water Resources. Retrieved from www.kimberly.uidaho.edu/ETIdaho

Allen, R.G., Tasumi, M., Morse, A. and Trezza, R., 2005. A Landsat-based energy balance and evapotranspiration model in Western US water rights regulation and planning. *Irrigation and Drainage Systems*, *19*(3-4), pp.251-268.

Allen, R.G., Tasumi, M. and Trezza, R., 2007. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)—Model. *Journal of Irrigation and Drainage Engineering*, *133*(4), pp.380-394.

Allen, R., Irmak, A., Trezza, R., Hendrickx, J.M., Bastiaanssen, W. and Kjaersgaard, J., 2011. Satellite-based ET estimation in agriculture using SEBAL and METRIC. *Hydrological Processes*, 25(26), pp.4011-4027.

Allen, R.G., Pereira, L.S., Raes, D. and Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao, Rome*, 300(9), p.D05109.

Allen, R.G., Robison, C.W., Huntington, J., Wright, J.L. and Kilic, A., 2020. Applying the FAO-56 Dual Kc Method for Irrigation Water Requirements over Large Areas of the Western US. Transactions of the ASABE, 63(6), pp.2059-2081.

Anderson, M.C., Norman, J.M., Mecikalski, J.R., Otkin, J.A. and Kustas, W.P., 2007. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. *Journal of Geophysical Research: Atmospheres*, *112*(D10).

Anderson, M., Gao, F., Knipper, K., Hain, C., Dulaney, W., Baldocchi, D., Eichelmann, E., Hemes, K., Yang, Y., Medellin-Azuara, J. and Kustas, W., 2018. Field-scale assessment of land and water use change over the California Delta using remote sensing. *Remote Sensing*, 10(6), p.889.

Arsenault, R., Gatien, P., Renaud, B., Brissette, F., & Martel, J. L. (2015). A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. *Journal of Hydrology*, *529*, 754-767.

ASCE (2005). The ASCE Standardized Reference Evapotranspiration Equation. Task committee on standardization of reference evapotranspiration. Walter, I.A., Allen, R.G., Elliott, R., Itenfisu, D., Brown, P., Jensen, M.E., Mecham, B., Howell, T.A., Snyder, R., Eching, S. and Spofford, T., ASCE: Reston, VA, USA. Retrievable from https://www.uidaho.edu/cals/kimberly-research-and-extension-center/research/water-resources/standardization

Bastiaanssen, W.G., Menenti, M., Feddes, R.A. and Holtslag, A.A.M., 1998. A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. *Journal of Hydrology*, 212, pp.198-212.

Brânzei, R., Tijs, S. and Timmer, J., 2001. Collecting information to improve decision-making. *International Game Theory Review*, *3*(01), pp.1-12.

Bos, Marinus G., et al. "Effective precipitation." *Water Requirements for Irrigation and the Environment* (2009): 81-101.

Dastane, N. G., 1974. Precipitación efectiva en la agricultura de regadío. FAO, Rome.

Feddes, R. A., et al. "Modelling soil water dynamics in the unsaturated zone—state of the art." *Journal of hydrology* 100.1-3 (1988): 69-111.

Fisher, J.B., Tu, K.P. and Baldocchi, D.D., 2008. Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. *Remote Sensing of Environment*, *112*(3), pp.901-919.

Foken, T., 2008. The energy balance closure problem: an overview, *Ecological Applications*, 18: 1351–1367, 2008b, doi:10.1890/06-0922.1.

Gorelick, Noel, et al. "Google Earth Engine: Planetary-scale geospatial analysis for everyone." *Remote sensing of Environment* 202 (2017): 18-27.

Hawkins, R.G., A.T. Hjelmfelt, and A.W. Zevenbergen. 1985. Runoff probability, storm depth, and curve numbers. *Journal of Irrigation and Drainage Engineering 111*(4):330–340.

Hampel, F.R., 1974. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, *69*(346), pp.383-393.

Huber, P.J. and Ronchetti, E.M., 1981. Robust statistics. John Wiley & Sons. *New York*, *1*(1).

Huntington, J.L., Szilagyi, J., Tyler, S.W. and Pohll, G.M., 2011. Evaluating the complementary relationship for estimating evapotranspiration from arid shrublands. *Water Resources Research*, *47*(5).

Huntington, J.L., Gangopadhyay, S., Spears, M., Allen, R. King, D., Morton, C., Harrison, A., McEvoy, D., and A. Joros, (2015). West-Wide Climate Risk Assessments: Irrigation Demand and Reservoir Evaporation Projections. U.S. Bureau of Reclamation, Technical memorandum No. 68-68210-2014-01, 196p., 841 app.

Irmak, S., Kilic, A., Howell, T. A., Martin, D., Payero, J. O., and Copeland, K. S. (2008). Variability Analyses of Alfalfa-Reference to Grass-Reference Evapotranspiration Ratios in Growing and Dormant Seasons. Biological Systems Engineering: Papers and Publications. 82. https://digitalcommons.unl.edu/biosysengfacpub/82

Jensen, M.E., Burman, R.D. and Allen, R.G., 1990. Evapotranspiration and irrigation water requirements. ASCE Manuals and Reports on Engineering Practices No. 70, New York, 332 p.

Jensen M.E and Allen R.G., 2016. Evaporation, evapotranspiration, and irrigation water requirements. ASCE Manual of Practice 70, 2nd edn. American Society of Civil Engineers, Reston, VA. 744 p.

Kahler, D. M. and Brutsaert, W., 2006. Complementary relationship between daily evaporation in the environment and pan evaporation. *Water resources research*, *42*(5).

Kirtman, B.P., Min, D., Infanti, J.M., Kinter, J.L., Paolino, D.A., Zhang, Q., Van Den Dool, H., Saha, S., Mendez, M.P., Becker, E. and Peng, P., 2014. The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, *95*(4), pp.585-601.

Kljun, N., Calanca, P., Rotach, M.W. and H.P. Schmid. 2015. A simple two-dimensional parameterization for Flux Footprint Prediction (FFP). *Geoscientific Model Development*, 8(11), 3695.

Kuan, Joseph. *Learning Highcharts*. Packt Publishing Ltd, 2012.

Kumar, R., Kumar, M., Farooq, Z., & Dadhich, S. M. (2017). Evaluation of models and approaches for effective rainfall in irrigated agriculture–An overview. *Journal of Soil and Water Conservation*, *16*(1), 32-36.

Laipelt, L., Kayser, R.H.B., Fleischmann, A.S., Ruhoff, A., Bastiaanssen, W., Erickson, T.A. and Melton, F., 2021. Long-term monitoring of evapotranspiration using the SEBAL algorithm and Google Earth Engine cloud computing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *178*, pp.81-96.

Leys, C., Ley, C., Klein, O., Bernard, P. and Licata, L., 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), pp.764-766.

Leuning, R., E. Van Gorsel, W. J. Massman, and P. R. Isaac, 2012, Reflections on the surface energy imbalance problem, *Agricultural and Forest Meteorology*, 156: 65–74, 2012, doi:10.1016/j.agrformet.2011.12.002.

Lhomme, J.P., 1997. A theoretical basis for the Priestley-Taylor coefficient. *Boundary-Layer Meteorology*, *82*(2), pp.179-191.

McAneney, K.J. and Itier, B., 1996. Operational limits to the Priestley-Taylor formula. *Irrigation Science*, *17*(1), 37-43.

Melton, F.S., Johnson, L.F., Lund, C.P., Pierce, L.L., Michaelis, A.R., Hiatt, S.H., Guzman, A., Adhikari, D.D., Purdy, A.J., Rosevelt, C. and Votava, P., 2012. Satellite irrigation management support with the terrestrial observation and prediction system: A framework for integration of satellite and surface observations to support improvements in agricultural water resource management. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*(6), pp.1709-1721.

Melton, F. S., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., ... & Anderson, R. G. (2021). OpenET: Filling a critical data gap in water management for the Western United States. *JAWRA Journal of the American Water Resources Association, https://doi.org/10.1111/1752-1688.12956*

Patwardhan, A.S., Nieber, J.L. and Johns, E.L., 1990. Effective rainfall estimation methods. *Journal of irrigation and Drainage Engineering*, 116(2), pp.182-193.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M. and Isaac, P., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Scientific data, 7(1), pp.1-27.

Pereira, L.S., Paredes, P., Melton, F., Johnson, L., Wang, T., López-Urrea, R., Cancela, J.J. and Allen, R.G., 2020. Prediction of crop coefficients from fraction of ground cover and height. Background and validation using ground and remote sensing data. *Agricultural Water Management*, 241, p.106197.

Priestley, C.H.B. and Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly weather review*, *100*(2), 81-92.

Senay, G.B., Bohms, S., Singh, R.K., Gowda, P.H., Velpuri, N.M., Alemu, H. and Verdin, J.P., 2013. Operational evapotranspiration mapping using remote sensing and weather datasets: A new parameterization for the SSEB approach. JAWRA *Journal of the American Water Resources Association*, 49(3), pp.577-591.

Senay, G.B., 2018. Satellite psychrometric formulation of the Operational Simplified Surface Energy Balance (SSEBop) model for quantifying and mapping evapotranspiration. *Applied Engineering in Agriculture*, *34*(3), pp.555-566.

Severance, Charles. "John resig: Building jquery." *Computer* 48.5 (2015): 7-8.

Stamm, G.G., 1967. Problems and Procedures in Determining Water Supply Requirements for Irrigation Projects. Chapter 40 in Irrigation of Agricultural Lands by Hagan et al., Wisconsin, American Society of Agronomy, II, 769-785.

Szilagyi, J., 2007. On the inherent asymmetric nature of the complementary relationship of evaporation. *Geophysical Research Letters*, *34*(2).

Tabari, H. and Talaee, P. H., 2011. Local calibration of the Hargreaves and Priestley-Taylor equations for estimating reference evapotranspiration in arid and cold climates of Iran based on the Penman-Monteith model. *Journal of Hydrologic Engineering*, *16*(10), 837-845.

Thompson, P.D., 1977. How to improve accuracy by combining independent forecasts. *Monthly Weather Review*, *105*(2), pp.228-229.

Twine, T., Kustas, W.P., Norman, J., Cook, D., Houser, P., Teyers, T.P., Prueger, J.H., Starks, P., and Wesely, M. 2000. Correcting Eddy-Covariance Flux Underestimates over a Grassland. *Agricultural and Forest Meteorology, 103*(3), pp.279-300.

USDA, 1967. Irrigation Water Requirements. Tech. Release No. 21, United States Department of Agriculture, Soil Management Unit. 1967; 59: 67-75.

Weiß, M. and Menzel, L., 2008. A global comparison of four potential evapotranspiration equations and their relevance to stream flow modelling in semi-arid environments. *Advances in Geosciences*, *18*, pp.15-23.