Intercomparison and Accuracy Assessment of **OpenET in California**

July 2026

OpenET Technical Series, No. 6

Prepared by

John M. Volk[4], Charles Morton[4], Blake Minor[4], Adam Purdy[1,9], Lee Johnson[1,9], Forrest S. Melton[9], Sachiko Sueki[4], Lydia Bleifuss[10], Sara G. Larsen[10], Richard G. Allen[5], Martha Anderson[11], Joshua B. Fisher[2], MacKenzie Friedrichs[12], Justin L. Huntington[4], Ayse Kilic[7], Gabriel E.L. Parrish[12], Christopher Pearson[4], Peter M. Revelle[4], Anderson Ruhoff[6], Gabriel B. Senay[12], Yun Yang[3]

California State University Monterey Bay[1], Chapman University[2], Cornell University[3], Desert Research Institute[4], Evapotranspiration Plus, LLC[5], Federal University of Rio Grande do Sul[6], GreenBlue ET LLC[7], National Aeronautics and Space Administration[9], OpenET, Inc.[10], United States Department of Agriculture[11], United States Geological Survey[12]

Prepared for

California Department of Water Resources

Executive Summary

Remotely sensed evapotranspiration (ET) data are currently being used in California and across the western United States (US) to support data-driven approaches for sustainable water management. OpenET, which comprises a set of six satellite-driven ET models, was developed to generate, distribute, and archive spatially continuous high resolution (30 m) ET data. The system also generates an ensemble value derived from the individual model set (Melton et al., 2022). A prior published study characterized OpenET performance using a nationwide set of ground-based measurements (Volk et al., 2024). This follow-on study describes an assessment involving publicly available data collected by 45 eddy covariance flux towers deployed within California, representing nearly 7,000 site-days and 1,400 site-months over a 20-year period. Results for agricultural lands are emphasized, and results are also summarized for natural land cover types. The OpenET Collection v2.0 ensemble value had mean bias error (MBE) of 1.1% and mean absolute error (MAE) of 19.0% at the monthly timestep for croplands. Exclusion of two highly complex flux tower sites with outlier values reduced the MBE for croplands to -1.7% and MAE to 17.3%. Updates to the OpenET models reflected in OpenET Collection v2.1 further improved the accuracy for croplands to a MBE of -0.6% and a MAE of 16.6% at monthly timesteps. These findings are consistent with a prior OpenET accuracy assessment for the contiguous US (CONUS). Cropland performance was also analyzed with respect to daily, growing season, and annual timesteps. Results for the ensemble value, on average, met error tolerances previously formulated by a diverse user-group for various agricultural water management applications. Monthly performance of the OpenET ensemble ET value for wetlands was comparable to agricultural sites, while accuracies for other natural ecosystems were lower. This analysis is intended to further inform California's water management community of the expected accuracy and variability of OpenET output for different land cover types and regions of the state.

1 Introduction

This research report provides an initial assessment of the accuracy of the six OpenET (https://etdata.org/; Melton et al., 2022) satellite-based remote sensing evapotranspiration (RSET) models as well as the multi-model ensemble value within the state of California. The six OpenET models include: ALEXI/DisALEXI (Anderson et al., 2018), eeMETRIC (Allen et al., 2007), geeSEBAL (Laipelt et al., 2021), PT-JPL (Fisher et al., 2008), SIMS (Melton et al., 2012; Pereira et al., 2020), and SSEBop (Senay et al., 2023), as well as their ensemble value. OpenET is an operational cloud-based system that provides publicly accessible, field-scale (30 m) evapotranspiration (ET) data throughout the United States (U.S.) at daily and monthly time steps. Spatially aggregated results corresponding to individual agricultural field boundaries are also available. While previous inter-model comparisons and accuracy assessments have addressed OpenET performance more broadly across the U.S. (Volk et al., 2024; Melton et al., 2022), there has yet to be a California focused assessment to guide regional use and applications.

The RSET models employed by OpenET are well established and have been applied in various water resource management contexts, including irrigation scheduling, water accounting, allocation, and rights administration (Karimi et al., 2013; Knipper et al., 2019; Senay et al., 2016; OpenET for Water Managers, OpenET for Growers). Such applications are particularly vital in California, where agriculture accounts for a major share of the nation's food production, yet surface water supplies are vulnerable to periodic drought, and recent legislation provides a framework for increased groundwater sustainability. In earlier stages of OpenET development, a user group comprised of a variety of stakeholders including farmers, irrigation managers, and local/state water agencies, formulated a consensus baseline and optimal accuracy goals for agricultural water management applications (Table 1).

Table 1. User-defined ET data application accuracy and data-type attributes (expanded from Volk et al., 2024).

Application	**Baseline Accuracy**	**Optimal Accuracy**	**Frequency**	**Ensemble or multiple values?**
Irrigation scheduling	+/- 25% daily	+/- 10-15% daily	1-3x per week	One value
Water accounting / demand projections	+/- 20% monthly +/- 15% annual	+/- 15%	Monthly to annual	Ensemble
Drought monitoring / impact assessment	+/- 20% monthly +/- 15% annual	+/- 15%	Monthly or seasonal reports	One value
Water transfers	+/- 20% monthly +/- 15% annual	+/- 10%	Monthly and annual reports	One value
Water budgets	+/- 15% annual	+/- 10%	Annual reports	One value
Calibration of groundwater models	+/- 15% annual	+/- 15%	Retrospective analyses	Ensemble
Conservation planning	+/- 15% annual	+/- 15%	Retrospective analyses	Ensemble
Water rights admin / Regulatory compliance	+/- 20% monthly +/- 15% annual	+/- 5-10%	Monthly and annual reports	One value

To evaluate model accuracy in California, this report compares OpenET data to ET measurements collected from eddy covariance (EC) micrometeorological systems (generally described by Baldocchi, 2014; Baldocchi et al., 2001). The report details how ground-based EC measurements were sourced, quality-controlled, and post-processed. In addition, we detail how

RSET data were sampled to compare with ground-based measurements. Results are summarized for croplands and natural land covers at various timesteps. Results from this analysis provide ET data users with information about the expected accuracy and variability of OpenET data for different land cover types and regions of California.

2 Background and methods

2.1 Eddy Covariance ("Flux") ET data

For this accuracy assessment, OpenET data were directly compared against *in situ* measurements of ET collected from EC systems. EC systems make direct and independent measurements of turbulent fluxes (sensible heat, latent heat) and, when properly sited and maintained, can provide a robust ground-based method for continuous ET measurement at the scale of a single agricultural field (Billesbach et al., 2024). The comparison focused primarily on ET derived from energy balance corrected EC data, referred to here as "closed flux ET." Here, "unclosed flux ET" refers to ET derived directly from measured latent heat flux before energy balance correction. "Closed flux ET" refers to ET after applying the ONEFlux/FLUXNET2015 (Pastorello et al., 2020) energy balance closure correction used in our processed flux dataset. This correction is not a single constant adjustment but varies through time with the observed energy balance ratio (the ratio of latent and sensible heat flux to net radiation minus ground heat flux) that is applied over moving windows to the latent energy (Volk et al., 2023a). To provide site level context on closure-correction magnitude, Appendix Table C1 summarizes mean closure percentage and the average difference between closed and unclosed ET for each flux tower site.

Energy imbalance is a well-known issue with EC measurements, where the sum of latent and sensible heat fluxes typically underestimates the available energy by 10–30% (Foken, 2008; Mauder et al., 2020). This lack of closure due to environmental factors or sensor limitations introduces uncertainty into ground-based measurements and its application for model validation. Comparisons with both closed and unclosed EC ET values are presented here to provide added context for OpenET model performance.

Datasets from EC stations in various parts of the state (Figure 1, Table 2) were carefully curated for this evaluation. Most sites were sourced from a publicly archived, post-processed benchmark flux ET dataset (Volk et al., 2023a; Volk et al., 2023b) largely derived from the AmeriFlux network (Novick et al., 2018). More recent in-state EC data from 23 stations from AmeriFlux, USDA Agricultural Research Service, and California State University Monterey Bay supplemented the previously published benchmark ET dataset. In total, this analysis featured 45 EC stations, 28 of which were agricultural (cropland) sites. Two agricultural sites were grazed pastures on islands in the Sacramento–San Joaquin River Delta that were not irrigated due to their proximity to open water and shallow water tables (US-Snf and US-Snd). All other agricultural sites were managed crops subject to irrigation. A total of 6,844 paired model-measurement days (3,351 for cropland) supported this analysis, representing 1,369 site-months (714 for cropland) from 2003–2024. Monthly timestep results were analyzed for all landcover types. Additional results at daily and growing season timesteps are shown for croplands. Although four EC towers used in this study were in short-season vegetable crops (site IDs MB_BRC_S1, MB_BRC_S2, MB_LET_S1, and MB_LET_S2), none had three or more months of paired data (our minimum requirement for computing monthly error metrics), and therefore these were only included in daily results.

Table 2. List of eddy covariance stations used in this analysis. Start and end dates refer to the first and last day of paired observations with OpenET data. Sites that were used in this analysis but not included in the CONUS-wide assessment of Volk et al. (2024) are italicized.

Site ID	Latitude	Longitude	Data Source	General Classification	Land Cover Details	Start	End
Almond_High	36.1697	-120.2010	USDA-ARS	Croplands	Almond	2016-10-18	2019-10-04
Almond_Low	36.9466	-120.1024	USDA-ARS	Croplands	Almond	2016-10-11	2019-10-04
Almond_Med	36.1777	-120.2026	USDA-ARS	Croplands	Almond	2016-10-02	2019-10-04
BAR007	38.7530	-122.9800	USDA GRAPEX	Croplands	Vineyard	2020-02-06	2020-12-22
BAR012	38.7510	-122.9750	USDA GRAPEX	Croplands	Vineyard	2017-05-20	2020-12-23
MandarinCitrus	36.4231	-119.2530	USDA-ARS	Croplands	Mandarin orange	2020-11-23	2023-09-13
MB_BRC_S1	36.4620	-121.3898	CSUMB	Croplands	Broccoli	2023-07-26	2023-09-28
MB_BRC_S2	36.4593	-121.3877	CSUMB	Croplands	Broccoli	2024-07-27	2024-10-15
MB_LET_S1	36.4569	-121.3827	CSUMB	Croplands	Lettuce	2023-04-20	2023-06-08
MB_LET_S2	36.4567	-121.3822	CSUMB	Croplands	Lettuce	2024-04-15	2024-06-10
MB_Pch	36.4587	-119.5801	USDA ARS	Croplands	Peach	2012-04-07	2015-12-12
MB_WG_ZAB	36.3471	-121.2966	CSUMB	Croplands	Winegrape	2023-04-05	2025-01-28
NavalCitrus	36.2347	-119.9250	USDA-ARS	Croplands	Navel orange	2020-11-22	2023-09-13
RIP760	36.8390	-120.2100	USDA GRAPEX	Croplands	Vineyard	2017-05-22	2020-12-09
Scheid	36.4150	-121.2800	CSUMB	Croplands	Winegrape	2020-04-28	2022-09-17
SLM001	38.2890	-121.1180	USDA GRAPEX	Croplands	Vineyard	2013-04-08	2020-12-08
SLM002	38.2800	-121.1180	USDA GRAPEX	Croplands	Vineyard	2013-04-08	2020-12-08
US-Bi1	38.0992	-121.4993	AmeriFlux	Croplands	Alfalfa	2016-01-27	2024-04-30
US-Bi2	38.1091	-121.5351	AmeriFlux	Croplands	Corn	2017-01-05	2024-04-30

US-Blo	38.8953	-120.6328	AmeriFlux	Evergreen Forests	Ponderosa pine plantation, mixed-ever green forest	2003-10-07	2007-10-02
US-CZ2	37.0311	-119.2566	AmeriFlux	Evergreen Forests	Oak/pine forest	2010-01-28	2016-09-25
US-CZ3	37.0674	-119.1951	AmeriFlux	Evergreen Forests	Pine/fir forest	2008-08-02	2016-09-25
US-Dmg	38.0015	-121.6691	AmeriFlux	Wetlands	Restored tidal freshwater wetland	2021-01-16	2024-04-30
US-DS1	38.1335	-121.5390	AmeriFlux	Croplands	Corn	2018-02-01	2019-12-30
US-DS2	38.1375	-121.5140	AmeriFlux	Croplands	Corn	2020-01-06	2021-03-13
US-DS3	38.1235	-121.5490	AmeriFlux	Croplands	Rice	2021-01-16	2023-08-26
US-EDN	37.6156	-122.1140	AmeriFlux	Wetlands	Permanent wetland	2018-02-01	2021-06-01
US-Hsm	38.2368	-122.0212	AmeriFlux	Wetlands	Diked marsh turned restored in wetland in Fall 2021	2021-01-08	2023-11-06
US-RGF	37.6995	-121.1369	AmeriFlux	Croplands	Corn-whea t rotation	2023-01-22	2023-12-09
US-SCf	33.8079	-116.7717	AmeriFlux	Mixed Forests	Oak/pine forest	2006-01-19	2014-12-27
US-SCg	33.7365	-117.6946	AmeriFlux	Grasslands	Grassland	2006-01-19	2016-09-27
US-SCs	33.7343	-117.6959	AmeriFlux	Shrublands	Coastal sage scrub	2006-01-19	2016-09-27
US-SCw	33.6047	-116.4527	AmeriFlux	Shrublands	Pinyon/Jun iper woodland	2006-01-12	2016-06-23
US-Snd	38.0366	-121.7540	AmeriFlux	Croplands	Peatland pasture	2007-03-15	2014-09-02
US-Sne	38.0369	-121.7547	AmeriFlux	Wetlands	Restored wetland	2016-01-27	2020-12-07
US-Snf	38.0402	-121.7272	AmeriFlux	Croplands	Pasture	2018-02-09	2020-07-08
US-SO2	33.3738	-116.6228	AmeriFlux	Shrublands	Chaparral, severe fire in 2003	2004-01-14	2006-12-14

US-SO3	33.3771	-116.6226	AmeriFlux	Shrublands	Chaparral, severe fire in 2003	2004-01-14	2006-12-29
US-SO4	33.3845	-116.6406	AmeriFlux	Shrublands	Old-growth chaparral ecosystem	2004-01-14	2006-12-29
US-Srr	38.2006	-122.0264	AmeriFlux	Wetlands	Brackish tidal marsh	2014-01-05	2017-09-26
US-Tw2	38.1047	-121.6433	AmeriFlux	Croplands	Corn on peat soil	2012-01-16	2013-04-16
US-Tw3	38.1151	-121.6467	AmeriFlux	Croplands	Alfalfa	2013-01-02	2018-06-01
US-Twt	38.1087	-121.6531	AmeriFlux	Croplands	rice	2009-01-31	2017-04-03
US-Var	38.4133	-120.9508	AmeriFlux	Grasslands	C3 annual grasses	2003-10-07	2024-04-07
US-xTE	37.0058	-119.0060	AmeriFlux	Evergreen Forests	Red and white fir	2018-02-03	2023-08-12

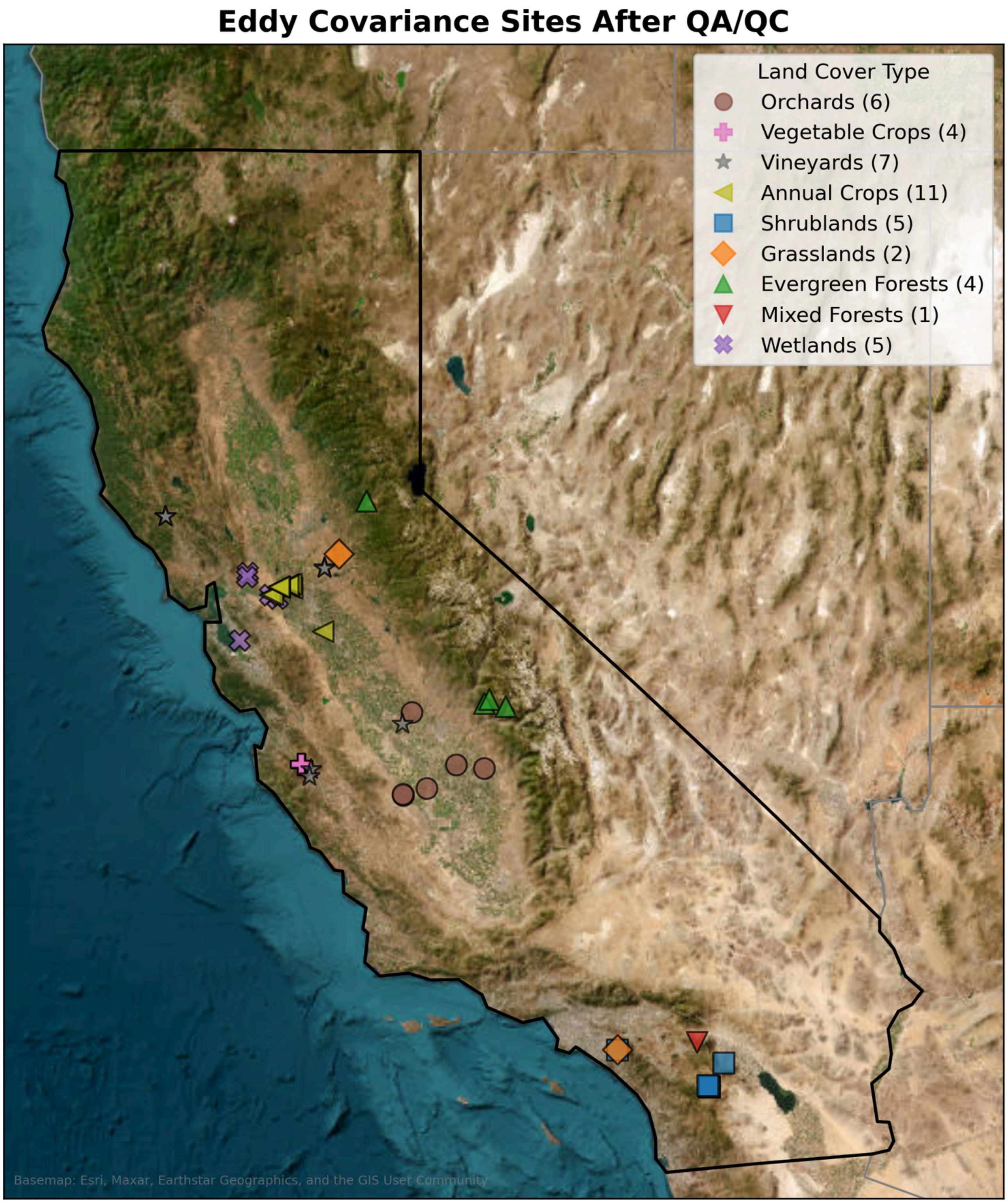


Figure 1. Locations and land cover types of eddy covariance stations used in this study. Note, the actual coordinates of some stations were slightly adjusted to reduce symbol overlap.

2.2 Flux Data Post-Processing and Quality Assurance / Quality Control

To ensure only high-quality EC data were included in this analysis, post-processing and quality assurance/quality control (QA/QC) using modern open-source Python tools (Volk et al., 2021) was performed. A variety of post-processing steps including filtering criteria, measurement thresholds related to the energy balance variables, and total number of observations within each day (Volk et al., 2023a). Whenever possible, we used only half-hourly flux data (latent, sensible, and ground heat flux, and net radiation) that had been approved and quality-checked by site principal investigators (PIs). PI-provided quality control flags were also applied to remove lower-quality or questionable flux observations. Heat storage terms were included wherever available, such as soil heat storage above flux plates.

Gap-filling of half-hourly and daily ET was performed conservatively. No more than 4 hours were gap-filled per day when calculating a daily ET total. No more than 5 days per month were gap-filled when calculating a monthly ET total. Missing values were assigned for daily/monthly ET aggregated values when these thresholds were exceeded. Daily energy balance closure correction followed the moving-window (5, 11, 15-day) energy balance ratio method used in FLUXNET2015/ONEFlux to be consistent with the broader science community and promote standardization. Modifications were introduced to the ONEFlux pipeline for daily flux processing, such as excluding days where the correction would halve or more than double the original latent heat flux (Volk et al., 2023a; Pastorello et al., 2020). Additional refinements included adjusting the latent heat of vaporization based on air temperature measured at the flux station (Volk et al., 2023b; Harrison 1963) and using gridded reference ET data combined with observed station EToF (fraction of grass reference ET) to fill short gaps in daily ET (only for monthly aggregation). These steps were implemented through the flux-data-qaqc software package (Volk et al., 2021) as used in prior OpenET assessments.

The automated post-processing and QA/QC steps described above were complemented by visual inspection to identify and manually remove anomalous data, such as flatlines or spikes caused by instrument errors. One AmeriFlux site (US-Lin) was excluded after manual review because its summer ET was anomalously low for a mature, drip-irrigated Valencia orange orchard (Fares et al., 2012). Summer EToF was only about 0.3–0.4, substantially below the citrus crop coefficient of 0.65 and historical Lindcove summer ET estimates reported by Schwankl et al. (2007), and the cause of this discrepancy could not be established with sufficient confidence. To further reduce uncertainty in ET related to energy balance closure, candidate sites with average growing-season energy balance closure below 75%, or below 60% in the non-growing season, were excluded before compiling the final dataset shown in Table 2. The sites excluded by this screening were: US-xSP, US-Dia, US-xSJ, US-Tw4, US-Ton, US-RGB, US-CGG, US-RGo, US-PSH, and US-PSL. Growing season periods were defined for each site using a cumulative growing degree day and killing frost threshold approach (Volk et al., 2023a). Before closure correction, the mean daily energy balance closure across all EC sites was 85%, implying a 15% underestimation of turbulent fluxes, particularly pronounced on high energy days (net radiation – ground heat flux > 150 W/m^2). According to consensus among experts in the field of eddy covariance measurements and theory, a mean daily energy imbalance of 85% indicates data of reasonable quality, i.e., not indicative of substantial measurement error or very high uncertainty (e.g., AGU Chapman Conference on Energy Balance Closure Problem, 2025).

2.3 OpenET RSET data

Model data were extracted for each EC station using the current (as of Dec. 2025) operational version (Collection v2.0 produced with CIMIS reference ET) of the six OpenET models implemented via the Google Earth Engine Python API (Gorelick et al., 2017). Model data were also extracted for the more recently released OpenET Collection v2.1 (Appendix A). The six models share many common input datasets (e.g., Landsat surface reflectance or temperature, gridded weather data, and land cover) but differ in their assumptions, parameterizations, and computational frameworks (Melton et al., 2022; Volk et al., 2024). Five of the six models constrain components of the surface energy balance using land surface temperature (LST), typically derived from Landsat Collection 2 (Crawford et al., 2023), in combination with gridded weather and land cover datasets. The exception is SIMS, which retrieves ET using a surface reflectance-based crop coefficient approach that assumes well-watered conditions, along with a soil water balance model based on precipitation that currently applies to monthly and growing season data.

OpenET model data were derived from Landsat imagery after scene screening, including removal of cloud-contaminated observations. As a result, the number and timing of valid overpass observations varied by site and season, which can influence the strength of daily and monthly comparisons when cloud-free data are reduced.

For monthly ET data, OpenET models interpolate daily ET between Landsat overpass dates to integrate monthly ET totals. In eeMETRIC, SIMS, and SSEBop, overpass day ET is computed as the fraction of reference ET (EToF) multiplied by reference ET (ETo). In geeSEBAL and PT-JPL, ET is computed directly on overpass days, EToF is then calculated as ET divided by ETo, and EToF is linearly interpolated between overpasses. For all five of these models, the interpolated daily EToF values are multiplied by ETo to produce daily ET. DisALEXI differs in that it estimates ET on overpass days but uses a solar radiation, rather than ETo and EToF, for time integration. Thus, monthly ET reflects both overpass day accuracy and the assumptions and data used to interpolate between overpass days.

Brief descriptions and the primary citation(s) of each OpenET model are provided below:

- ALEXI/DisALEXI: A two-source energy balance model that partitions ET into vegetation and soil components using LST and radiative fluxes (Anderson et al., 2018).

- eeMETRIC: An energy balance model adapted for Google Earth Engine, which uses reference ET to internally calibrate ET data based on surface energy balance (Allen et al., 2007).
- geeSEBAL: A surface energy balance model using satellite-derived radiative and aerodynamic variables, without the need for local calibration (Laipelt et al., 2021).
- PT-JPL: A Priestley-Taylor-based model that derives ET by applying stress scalars to partition evaporation and transpiration (Fisher et al., 2008).
- SIMS: Satellite Irrigation Management Support model calculates crop coefficients (Kc) from normalized difference vegetation index (NDVI) and vegetation density calculated from Landsat surface reflectance under well-watered assumptions, and calculates ET as Kc × reference ET (ETo). For monthly outputs, a simplified soil water balance model, driven by precipitation, is included to partition ET into crop transpiration and bare soil evaporation (Melton et al., 2012; Pereira et al., 2020).
- SSEBop: The Operational Simplified Surface Energy Balance model is driven by LST and reference ET and parameterized with NDVI and surface psychrometric constant using the principle of satellite psychrometry to directly compute daily actual ET (Senay et al., 2023).

The OpenET ensemble ET value is computed using a robust outlier rejection approach based on the median absolute deviation (MAD) method (Hampel, 1974; Leys et al., 2013; Rousseeuw & Hubert, 2011). Up to two model outliers are identified and excluded at the pixel level (for each individual pixel and date) before computing the mean for daily and monthly ET. In non-agricultural areas, SIMS is excluded from the ensemble because the current implementation was not designed to address natural or unmanaged vegetation. In such pixels, only one model may be removed by the MAD filter so that at least four models are used in the ensemble calculation. Additional technical details are also available for each model in OpenET's online documentation.

For daily model data comparisons with flux ET, we limited our analysis to days of satellite overpass, whereas for monthly and longer time periods, monthly aggregated model ET data were used. Monthly ET values from OpenET models incorporate daily interpolation techniques that use gridded reference ET (Spatial CIMIS) or solar radiation (for DisALEXI) (Melton et al., 2022; Volk et al., 2024).

2.4 Sampling OpenET RSET Pixels at Flux Sites

Model data were extracted at each flux tower using either temporally static or dynamic spatial sampling approaches based on flux footprints. Static grids (3×3, 5×5, or 7×7 30 m Landsat pixels) were defined using long-term wind speed and direction data, while dynamic footprints were derived using the two-dimensional flux footprint parameterization of Kljun et al. (2015), and weighted by hourly reference ET derived from NLDAS-2 forcing data (Xia et al., 2012) to better account for hourly variability in source area influence (Volk et al., 2023a). Typically, dynamic footprints could not be produced due to missing input parameters (13 sites out of 46 had sufficient data, 10 of which were croplands). In those cases, a 7x7 static footprint was used to extract ET data. A prior analysis of 87 sites across the contiguous U.S. showed that a 7×7 static grid captured approximately 92% of the dynamic footprint area on average (Volk et al., 2023a). For certain locations with high land cover heterogeneity (i.e., smaller sites), we used smaller 5x5 or 3x3 static footprints to ensure the OpenET model ET values best represent the EC measurement source area, following Volk et al. (2024; 2023a).

In addition to the footprint methods described above, we also evaluated the impact of additional wind-direction filtering to prevent flux contributions from adjacent, non-representative land-cover types from influencing aggregate results. Appendix B provides additional details on the methods and results of this analysis. Although the additional filtering resulted in slight improvements in relative accuracy overall and meaningful improvements at some flux tower sites, the additional filtering also reduced the total number of paired model-measured data. For this reason, the results of this analysis are presented in Appendix B, while the main body of the report prioritizes use of a larger sample size of sites.

2.5 Model Evaluation Metrics

Five commonly used goodness-of-fit metrics were used to compare modeled and measured ET at each flux station:

(1) Slope of a linear fit forced through the origin, such that zero model ET corresponds to zero measured ET. Slope summarizes proportional bias whereby values >1 indicate positive bias (model over-prediction) and values <1 indicate negative bias (under-prediction).

(2) MBE (mean bias error) calculated as modeled minus measured ET. In the equations below, *P* refers to modeled or predicted ET values, and *O* refers to measured or observed ET values. Positive values indicate model over-prediction, negative indicates under-prediction, and zero indicates perfect agreement between modeled and measured values.

$$MBE = (1/n)\sum_{i=1}^{n}\left(P_i - O_i\right)$$

(3) MAE (mean absolute error) calculated as the absolute value of modeled minus measured ET. Values are always positive, and zero indicates perfect agreement between modeled and measured values.

$$MAE = (1/n)\sum_{i=1}^{n}\left|P_i - O_i\right|$$

(4) RMSE (root mean square error). RMSE is similar to MAE but penalizes large differences to a greater degree. Values are always positive, and zero indicates perfect agreement between modeled and measured values.

$$RMSE = \sqrt{(1/n)\sum_{i=1}^{n}\left(P_i - O_i\right)^2}$$

5) r² (coefficient of determination) provides a measure of the proportion of the variance in a dependent variable that is explained by the independent variable, and is calculated as the square of the Pearson correlation between modeled and measured ET. The value is independent of the origin-forced fit. Range is 0-1, with higher values indicating a stronger relationship between modeled and measured values.

We note that while both Slope and MBE measure bias, Slope gives more weight to higher ET values as it forces the regression through the origin. As a result, there are instances in this study where the two metrics disagree on bias direction (positive vs. negative).
For MBE, MAE, and RMSE, grouped statistics were calculated as weighted means, weighting each station by the square root of its number of paired data points (sample size) to limit the influence of unusually short or long records (Obrecht, 2019). For Slope and r^2, grouped values were derived from a single calculation using all paired data combined. Grouped statistics were reported by major land cover class (e.g., croplands, shrublands). To ensure robustness, grouped weighted-mean error statistics for MBE, MAE, and RMSE included only stations with at least 3 months of data for monthly analyses or at least 6 days for daily analyses. Site metadata and date ranges are provided in Table 2, while site-level daily and monthly accuracy statistics and paired sample counts are provided in Appendix C (Tables C2–C3).

It should be recognized that for purposes of this study, the EC data were considered error-free, and all errors were attributed to OpenET. In practice, the EC data also have inherent measurement error and uncertainty, especially during periods when atmospheric conditions violate the assumptions of the eddy covariance approach (e.g., low wind speed, precipitation, extreme atmospheric instability, etc.). The term 'error' conveys the 'difference' between these two measurement sets, each of which is subject to uncertainty.

3 Results and Discussion

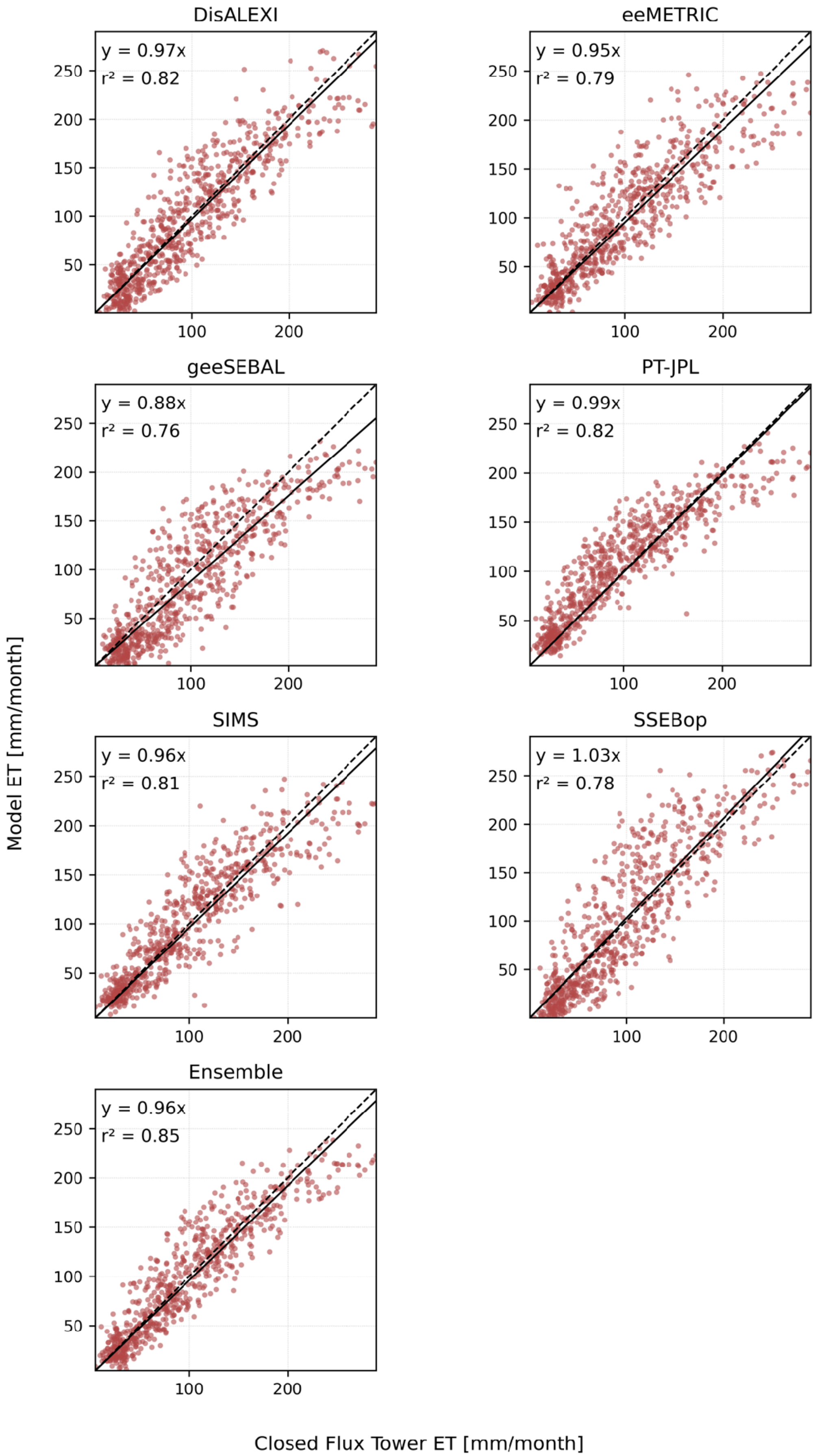
3.1 Model Accuracy in Croplands

Figure 2 shows scatter plots comparing OpenET Collection v2.0 monthly ET versus closed flux tower ET for all cropland sites. For monthly ET retrievals, the ensemble value showed a Slope of 0.96 and MBE of 1.1 mm/month (Table 3), both indicating very low bias. The ensemble MAE and RMSE across all available sites were 18.7 and 23.4 mm/month, respectively, and $r^2 = 0.85$. The ensemble ET value generally outperformed any single model, with lower MAE, RMSE, and higher r^2. Errors associated with the ensemble value on average met stated tolerances for agricultural and water resources management applications (Table 1).

Table 3. Accuracy metrics between modeled-measured ET for California cropland closed flux tower ET using paired data aggregated at daily, monthly, growing season, and water year temporal periods from OpenET Collection v2.0. Growing season periods were determined for each site using the cumulative growing degree day and killing frost approach described in Volk et al. (2023a); growing season total ET values were then accumulated using the monthly ET data from the models and paired with the EC monthly data. The values for weighted mean MBE, MAE, and RMSE are also provided as percentages of the weighted mean measured monthly ET. These are referred to as "normalized" error metrics in the text (i.e., NMBE, NMAE, and NRMSE). For MBE, MAE, and RMSE, we required a minimum of 6 paired model-observation data points per flux station for daily metrics and 3 paired data points per station for monthly metrics. For this reason, the number of data points differs slightly in some instances compared to slope and r^2, which had no minimum selection criteria following Volk et al. (2024). Daily results for SIMS exclude its soil evaporation component, which applies to monthly and growing season aggregations only. The ensemble value may deviate from arithmetic mean of individual models due to outlier filtering and rounding.

Aggregation	Mean measured ET	Statistic	Ensemble	DisALEXI	eeMETRIC	geeSEBAL	PT-JPL	SIMS	SSEBop	N sites	N data points
Daily	3.5 (mm)	Slope	0.95	0.95	0.96	0.87	0.97	0.94	1.07	28	3351
		MBE (mm)	0.11 (3.1%)	0.09 (2.7%)	0.15 (4.2%)	-0.19 (-5.6%)	0.37 (10.6%)	0.03 (1.0%)	0.37 (10.8%)	27	3348
		MAE (mm)	0.83 (24.0%)	0.93 (26.8%)	1.04 (30.2%)	1.16 (33.4%)	1.03 (29.9%)	0.87 (25.3%)	1.10 (31.8%)	27	3348
		RMSE (mm)	1.06 (30.6%)	1.17 (33.8%)	1.32 (38.3%)	1.48 (42.8%)	1.29 (37.4%)	1.12 (32.5%)	1.38 (40.0%)	27	3348
		R-squared	0.77	0.72	0.67	0.61	0.68	0.73	0.69	28	3351
Monthly	98 (mm)	Slope	0.96	0.97	0.95	0.88	0.99	0.96	1.03	25	714
		MBE (mm)	1.1 (1.1%)	1.2 (1.2%)	-0.7 (-0.7%)	-6.9 (-7.0%)	9.5 (9.7%)	1.5 (1.6%)	5.2 (5.3%)	23	712
		MAE (mm)	18.7 (19.0%)	21.5 (21.9%)	22.7 (23.1%)	26.3 (26.8%)	21.9 (22.4%)	20.6 (20.9%)	27.5 (28.0%)	23	712
		RMSE (mm)	23.4 (23.8%)	26.1 (26.6%)	28.8 (29.4%)	32.3 (32.9%)	27.3 (27.8%)	26.2 (26.7%)	32.7 (33.3%)	23	712
		R-squared	0.85	0.82	0.79	0.76	0.82	0.81	0.78	25	714
Growing season	692 (mm)	Slope	0.99	0.99	1.01	0.87	1.07	1.01	1.01	24	80
		MBE (mm)	45.4 (6.6%)	47.4 (6.8%)	50.1 (7.2%)	-30.0 (-4.3%)	112.7 (16.3%)	56.0 (8.1%)	65.9 (9.5%)	24	80
		MAE (mm)	126.3 (18.3%)	150.5 (21.8%)	149.2 (21.6%)	161.2 (23.3%)	157.7 (22.8%)	151.0 (21.8%)	160.6 (23.2%)	24	80
		RMSE (mm)	154.6 (22.3%)	169.6 (24.5%)	184.1 (26.6%)	176.5 (25.5%)	191.9 (27.7%)	188.7 (27.3%)	187.0 (27.0%)	24	80
		R-squared	0.82	0.78	0.77	0.75	0.79	0.75	0.75	24	80
Water year	1177 (mm)	Slope	0.92	0.92	0.96	0.79	1.01	0.97	0.93	10	31
		MBE (mm)	-97.5 (-8.3%)	-101.0 (-8.6%)	-62.7 (-5.3%)	-248.2 (-21.1%)	10.8 (0.9%)	-34.9 (-3.0%)	-91.7 (-7.8%)	10	31
		MAE (mm)	110.8 (9.4%)	162.2 (13.8%)	156.7 (13.3%)	248.2 (21.1%)	108.8 (9.2%)	128.4 (10.9%)	144.2 (12.2%)	10	31
		RMSE (mm)	117.0 (9.9%)	166.2 (14.1%)	165.6 (14.1%)	253.8 (21.6%)	114.3 (9.7%)	134.3 (11.4%)	152.2 (12.9%)	10	31
		R-squared	0.83	0.61	0.41	0.81	0.73	0.55	0.81	10	31

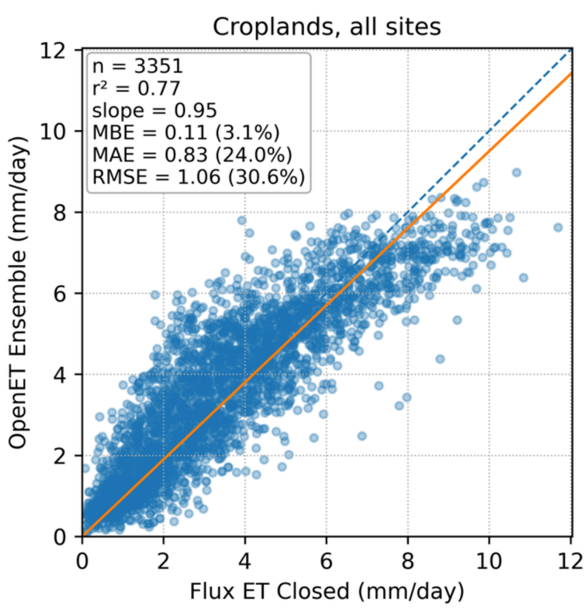
Table 3 shows the performance of individual models and the ensemble value for different temporal aggregation periods ranging from daily to water year. Individual model MBE ranges from -0.19 mm/day to 0.37 mm/day, the MAE ranges from 0.87 to 1.16 mm/day, and the explained variance ranges from 0.61–0.73 for daily data. Among the individual models, SIMS showed the lowest MBE, lowest MAE, and greatest r^2. For monthly data, the individual model MBE ranges from -7.0 mm to 9.7 mm, the MAE ranges from 20.6-27.5 mm, and the explained variance ranges from 0.76-0.82. eeMETRIC showed the lowest MBE (-0.7%) and SIMS showed the lowest MAE (20.9%) among the individual models. Three models explained 80% or more of the variance at this time aggregation. At the growing season aggregations, the individual model MBE ranges from -30 mm (-4.3%) to 112.7 mm (16.3%), and the MAE ranges from 149.2-161.2 mm. geeSEBAL showed the lowest MBE (-4.3%) and DisALEXI, eeMETRIC, and SIMS had similar and low MAE (~21.8%). At water year aggregations, the individual model MBE ranges from -248.2 mm (-21.1%) to 10.8 mm (0.9%), and the MAE ranges from 108.8-248.2 mm. PT-JPL showed the lowest MBE (0.9%) and MAE (9.2%). While all individual model statistics are shown, certain model ET values were excluded from the ensemble ET calculation due to outlier detection and filtering. As a result, the ensemble ET value is not necessarily equivalent to the arithmetic mean of the individual models. The ensemble value frequently has lower errors and greater r^2 relative to individual models across all time-aggregation periods.



DisALEXI
y = 0.97x
r² = 0.82
eeMETRIC
y = 0.95x
r² = 0.79
geeSEBAL
y = 0.88x
r² = 0.76
PT-JPL
y = 0.99x
r² = 0.82
SIMS
y = 0.96x
r² = 0.81
SSEBop
y = 1.03x
r² = 0.78
Ensemble
y = 0.96x
r² = 0.85
Model ET [mm/month]
Closed Flux Tower ET [mm/month]

Figure 2. Comparison of all paired OpenET and closed flux tower data from all cropland stations for all months of record. Included for each model is the result of least squares linear regression forced through the origin (solid black line). The coefficient in each regression equation is the Slope through the origin and the r^2 (Pearson correlation coefficient) is also shown. The black dashed line is 1:1.

With the relatively small number of cropland flux tower sites available in California, grouped statistics are sensitive to the inclusion of a few atypical sites. Table 3 reports cropland results for all sites, whereas Table 4 and Figure 3 additionally show a sensitivity analysis excluding the spatially complex sites: NavalCitrus, and US-Snd. These two sites were retained in the full cropland summary for completeness but were also evaluated separately because they exhibited consistently large positive bias across OpenET models and have site characteristics that complicate both eddy covariance measurements and model evaluation. NavalCitrus is a citrus orchard with heterogeneous tree–interrow structure that was subject to a major thinning event in 2022 (Dhungel et al., 2024); it is also adjacent to a solar farm that may alter local energy balance and aerodynamics. US-Snd is a drained peatland pasture on a subsided Delta island surrounded by canals, levees, and patches of open water, making it less representative of typical irrigated cropland and more sensitive to footprint and wind-direction effects. We therefore interpret Table 4 as more representative of typical California cropland performance, while retaining Table 3 to show the influence of these complex sites on the full cropland dataset. Exclusion of two spatially complex sites (NavalCitrus and US-Snd), as shown in Figure 3, improves the monthly MAE from 19.0% to 17.3% for the ensemble ET value, while the growing season MBE improves to 4.5% and the MAE to 16.7%.



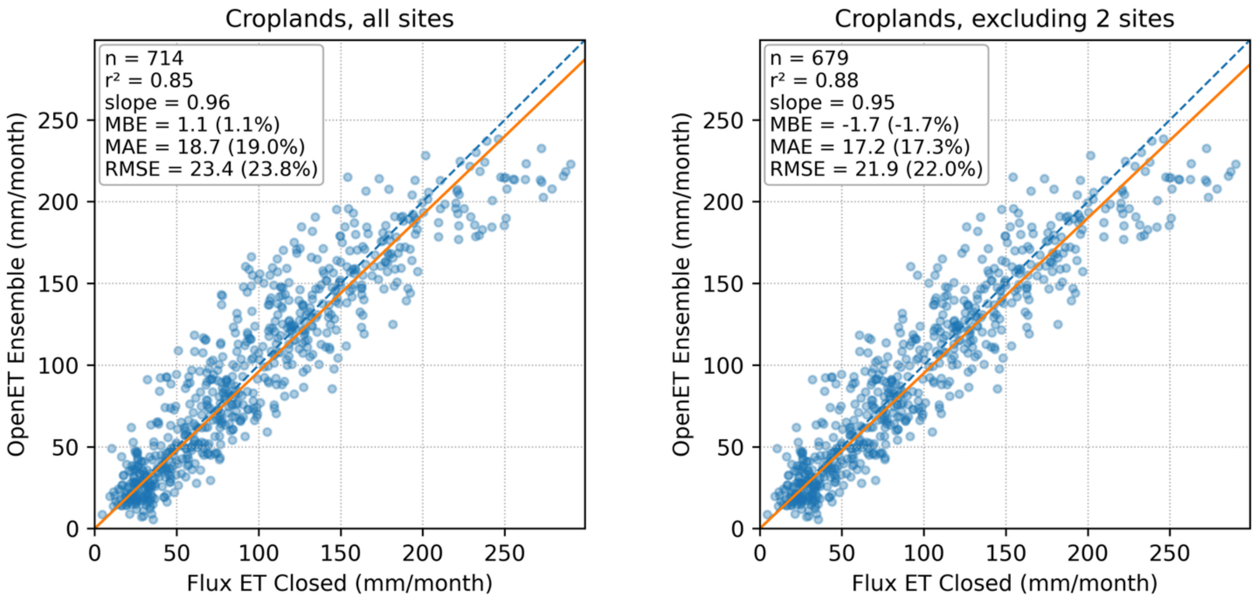


Figure 3. Comparisons between the OpenET ensemble daily (top) and monthly (bottom) ET and eddy covariance ET for all cropland stations with and with the exclusion of two flux stations (the NavalCitrus orange orchard and the US-Snd peat pasture) that are atypical for CA agriculture and exhibit large bias in all OpenET models. Included are the same summary statistics that are used throughout this report: r^2, Slope through origin, MBE, MAE, and RMSE both in mm and normalized by mean flux tower ET.

Two cropland sites (NavalCitrus and US-Snd) exhibited unusually large positive model biases relative to the rest of the cropland EC. For example, at NavalCitrus, the ensemble had a monthly Slope of 1.42, MBE of 27.9 mm $month^{-1}$ (41.4%), and MAE of 28.2 mm $month^{-1}$ (41.9%). Site-level monthly statistics are provided in Appendix C (Table C3). One possible contributing factor is a mismatch between overpass time conditions and the full day ET response at woody crop sites, where diurnal hysteresis and canopy–soil partitioning may be more pronounced. We note this as a plausible hypothesis, although an initial preliminary analysis of half-hourly ground measurements at the two citrus orchard sites used in this assessment did not provide convincing evidence.

The differences between Table 3 and Table 4 highlight the importance of improving OpenET values for citrus, increasing the number of long-term flux tower records at cropland sites in California, and investigating the influence of spatial heterogeneity and wind direction on the selection of static and dynamic footprints computed at each flux tower site. For US-Snd in particular, shifting the tower less than 100 m to the west in 2014 (when the field was restored to a natural wetland and renamed as US-Sne) dramatically improved agreement between the EC data and OpenET data in the same field. Overall, the accuracy metrics in Table 4 are expected to be more representative of cropland areas in California, with the possible exception of citrus crops.

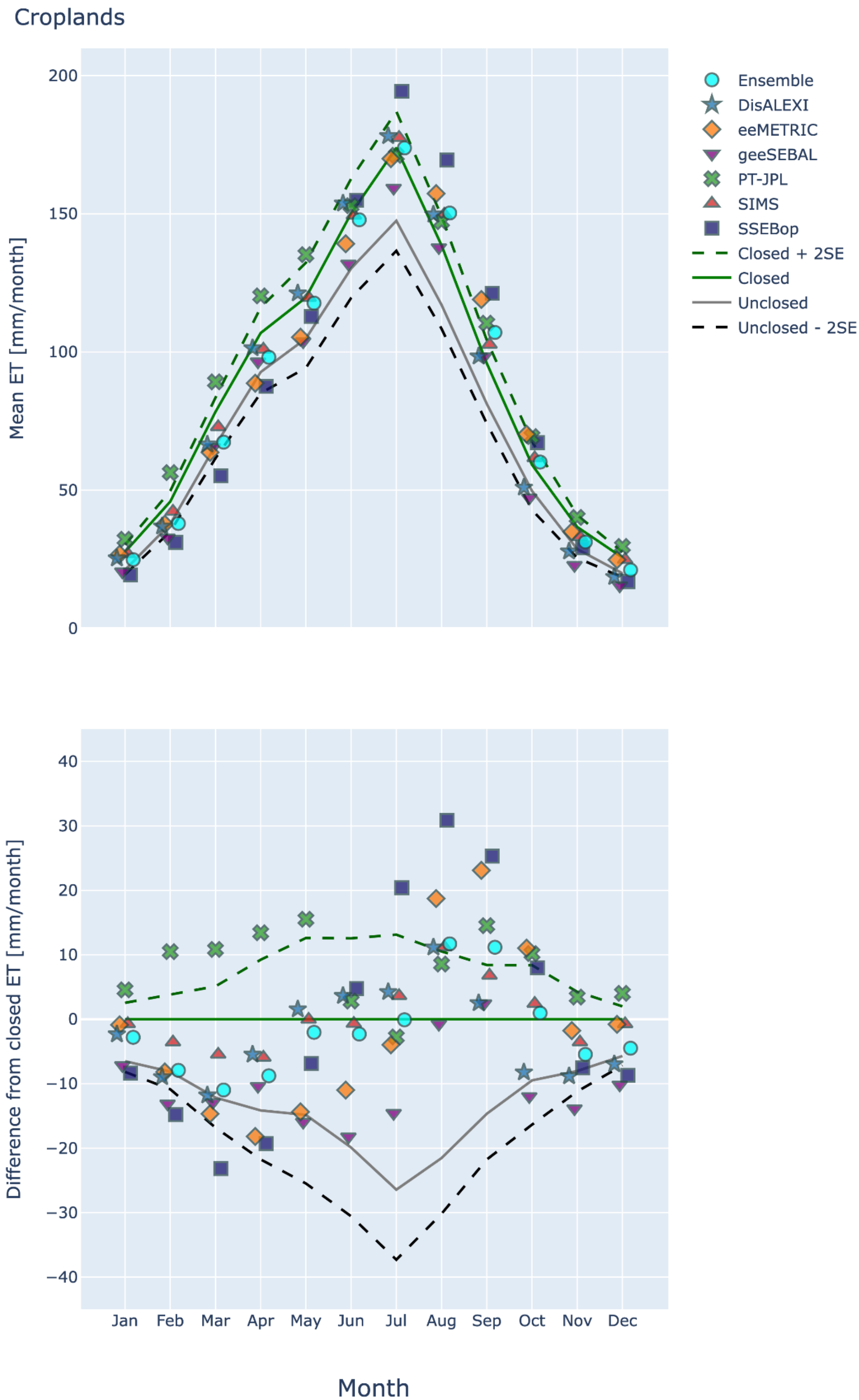
Table 4. Accuracy metrics between modeled-measured ET for California cropland closed flux tower ET excluding two complex sites (NavalCitrus and US-Snd) using paired data aggregated at daily, monthly, growing season, and water year temporal periods from OpenET Collection v2.0. See description of Table 3 for more details on how data are aggregated and how statistics are computed.

Aggregation	Mean measured ET	Statistic	Ensemble	DisALEXI	eeMETRIC	geeSEBAL	PT-JPL	SIMS	SSEBop	N sites	N data points
Daily	3.5 (mm)	Slope	0.94	0.94	0.94	0.85	0.95	0.92	1.05	26	3093
		MBE (mm)	0.01 (0.1%)	-0.00 (-0.0%)	0.04 (1.2%)	-0.31 (-8.8%)	0.28 (7.8%)	-0.08 (-2.2%)	0.26 (7.5%)	25	3090
		MAE (mm)	0.78 (22.2%)	0.90 (25.4%)	1.00 (28.5%)	1.14 (32.4%)	1.00 (28.3%)	0.83 (23.5%)	1.04 (29.5%)	25	3090
		RMSE (mm	1.01 (28.7%)	1.14 (32.2%)	1.28 (36.4%)	1.46 (41.5%)	1.26 (35.7%)	1.08 (30.5%)	1.31 (37.1%)	25	3090
		R-squared	0.8	0.74	0.7	0.64	0.71	0.76	0.72	26	3093
Monthly	99 (mm)	Slope	0.95	0.95	0.94	0.86	0.98	0.95	1.01	23	679
		MBE (mm)	-1.7 (-1.7%)	-1.6 (-1.6%)	-3.3 (-3.3%)	-9.9 (-10.0%)	7.4 (7.4%)	-1.3 (-1.3%)	1.8 (1.8%)	21	677
		MAE (mm)	17.2 (17.3%)	20.2 (20.4%)	21.5 (21.6%)	25.6 (25.8%)	20.7 (20.9%)	19.2 (19.3%)	25.6 (25.7%)	21	677
		RMSE (mm)	21.9 (22.0%)	24.8 (25.0%)	27.6 (27.8%)	31.7 (31.9%)	26.2 (26.3%)	24.9 (25.1%)	30.5 (30.7%)	21	677
		R-squared	0.88	0.84	0.81	0.78	0.83	0.83	0.8	23	679
Growing season	741 (mm)	Slope	0.98	0.98	1	0.86	1.06	1	1	22	70
		MBE (mm)	33.2 (4.5%)	36.4 (4.9%)	39.2 (5.3%)	-48.7 (-6.6%)	107.4 (14.5%)	45.1 (6.1%)	49.8 (6.7%)	22	70
		MAE (mm)	123.5 (16.7%)	151.5 (20.4%)	149.7 (20.2%)	164.7 (22.2%)	157.6 (21.3%)	151.1 (20.4%)	155.4 (21.0%)	22	70
		RMSE (mm	153.5 (20.7%)	171.4 (23.1%)	187.1 (25.2%)	180.7 (24.4%)	193.6 (26.1%)	191.0 (25.8%)	182.6 (24.6%)	22	70
		R-squared	0.79	0.74	0.72	0.71	0.76	0.7	0.71	22	70
Water year	1177 (mm)	Slope	0.92	0.92	0.96	0.79	1.01	0.97	0.93	10	31
		MBE (mm)	-97.5 (-8.3%)	-101.0 (-8.6%)	-62.7 (-5.3%)	-248.2 (-21.1%)	10.8 (0.9%)	-34.9 (-3.0%)	-91.7 (-7.8%)	10	31
		MAE (mm)	110.8 (9.4%)	162.2 (13.8%)	156.7 (13.3%)	248.2 (21.1%)	108.8 (9.2%)	128.4 (10.9%)	144.2 (12.2%)	10	31
		RMSE (mm)	117.0 (9.9%)	166.2 (14.1%)	165.6 (14.1%)	253.8 (21.6%)	114.3 (9.7%)	134.3 (11.4%)	152.2 (12.9%)	10	31
		R-squared	0.83	0.61	0.41	0.81	0.73	0.55	0.81	10	31

3.2 Seasonal Variability in Croplands

To evaluate seasonal effects on model accuracy, monthly means were calculated from paired OpenET and EC datasets. Figure 4 shows the paired OpenET data juxtaposed to the energy-balance corrected (closed) and uncorrected (unclosed) EC data. Inclusion of unclosed flux data provides context regarding the inherent measurement uncertainty in the eddy covariance method (Volk et al., 2024; Ingwersen et al., 2015). As an additional illustration of uncertainty in measured ET, dashed lines represent ±2 standard errors of the monthly closed and unclosed EC values. The envelope provides a rough approximation of the plausible range in measured ET, assuming errors are normally distributed around monthly ET measurements.

Clear patterns of seasonal variability were observed in OpenET model performance, with the strongest agreement between modeled and measured ET during the peak growing season months. The smallest relative deviations from measured ET were during mid-summer (June–August), a period when irrigation demand is highest and ET rates are typically at their peak. Seasonal deviations associated with individual models were generally ±20 mm/month. The ensemble ET value consistently fell between the closed and unclosed flux ET values, with monthly differences typically within ±10 mm/month of the closed flux ET.



Croplands
Mean ET [mm/month]
Difference from closed ET [mm/month]
Month
Ensemble
DisALEXI
eeMETRIC
geeSEBAL
PT-JPL
SIMS
SSEBop
Closed + 2SE
Closed
Unclosed
Unclosed - 2SE
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

Figure 4. Monthly mean ET from OpenET RSET models and compared with measured flux tower ET at all cropland sites (top). Same data presented as differences vs. closed flux ET (bottom). Closed and Unclosed refer to flux tower ET with and without energy balance closure correction. Dashed lines indicate ±2 standard errors of the mean (SE) for the closed and unclosed eddy covariance data and represent the approximate range of variability in monthly mean flux ET. Note that markers' x-spacing is slightly adjusted to avoid overlapping.

To evaluate potential seasonality in satellite comparison availability, we counted the number of dates with valid daily OpenET ensemble values in the paired daily dataset for each site-month. Availability was lowest in winter and highest during the growing season: winter site-months averaged 2.2 paired daily values, compared with 3.1 in spring, 4.1 in summer, and 3.5 in fall. Only 34% of winter site-months had three or more paired daily values, compared with 74% in summer and 68% in fall. These results suggest that seasonal differences in cloud-free satellite image availability, along with the temporal interpolation methods used to produce continuous daily ET values, may contribute to some of the seasonal patterns observed in Figure 4.

The consistent agreement between the ensemble and EC data across seasons highlights the value of combining multiple models to reduce uncertainty in monthly ET data retrievals. Some seasonal trends in individual model bias were evident relative to EC data in Figure 4. Both SSEBop and eeMETRIC were relatively low (outside the unclosed-closed EC uncertainty range) during spring (March–May) and relatively high during late summer and early fall (July–September). The geeSEBAL values were relatively low in late spring (May–June), while PT-JPL was relatively high during late winter and early spring (February–May).

3.3 Results for Crop Types and Examples at Individual Sites

Table 5 presents ET data accuracy statistics for individual crop categories such as annuals (wheat, soy, corn, rice, alfalfa, and hay), orchards, and vineyards. Results for annual crops were similar to those of all croplands presented in Table 3. The OpenET ensemble shows strong performance with NMBE of –0.3%, -3.0%, and 9.9% for annual crops, orchards, and vineyards respectively. Annual crops make up the majority of the EC sites used in the study and thus have more weight in the grouped accuracy metrics.

Table 5. Mean monthly statistics between modeled-measured ET for California cropland closed flux tower ET, grouped by major crop categories: annuals (wheat, corn, soy, rice), orchards (almond, peach, citrus), and winegrape vineyards. The values for mean MBE, MAE, and RMSE are also provided as percentages of the weighted mean measured monthly ET, referred to as "normalized" error metrics in the text (NMBE, NMAE, and NRMSE). For MBE, MAE, and RMSE, we require a minimum of 6 paired model-observation data points per flux station for daily metrics and 3 paired data points per station for monthly metrics. For this reason, the number of data points differs slightly in some instances compared to slope and r^2, which had no minimum selection criteria following Volk et al. (2024).

Land cover type	Statistic	Ensemble	DisALEXI	eeMETRIC	geeSEBAL	PT-JPL	SIMS	SSEBop	N sites	N data points
	Slope	0.97	0.98	1.04	0.82	1.01	0.98	0.97	11	406
Annual crops	MBE (mm)	-0.3 (-0.3%)	-0.3 (-0.4%)	6.1 (7.0%)	-15.4 (-17.8%)	7.5 (8.6%)	2.8 (3.3%)	-4.3 (-5.0%)	11	406
Mean station ET	MAE (mm)	17.3 (20.0%)	22.0 (25.4%)	21.3 (24.5%)	26.9 (31.0%)	16.2 (18.7%)	19.8 (22.8%)	27.1 (31.2%)	11	406
= 87 (mm)	RMSE (mm)	21.0 (24.2%)	26.0 (30.0%)	27.1 (31.3%)	31.6 (36.4%)	20.6 (23.7%)	24.6 (28.4%)	31.6 (36.4%)	11	406
	R-squared	0.87	0.83	0.82	0.79	0.87	0.81	0.8	11	406
	Slope	0.9	0.92	0.84	0.88	0.91	0.94	0.99	6	183
Orchards	MBE (mm)	-3.4 (-3.0%)	-1.5 (-1.3%)	-12.4 (-10.7%)	-4.9 (-4.2%)	2.3 (2.0%)	0.9 (0.8%)	2.7 (2.3%)	6	183
Mean station ET =	MAE (mm)	21.9 (18.9%)	23.6 (20.4%)	27.6 (23.8%)	25.4 (22.0%)	26.8 (23.2%)	22.7 (19.6%)	23.1 (19.9%)	6	183
116 (mm)	RMSE (mm)	28.7 (24.8%)	29.9 (25.9%)	34.9 (30.2%)	33.3 (28.8%)	33.6 (29.1%)	30.2 (26.1%)	31.1 (26.9%)	6	183
	R-squared	0.87	0.84	0.82	0.82	0.83	0.84	0.83	6	183
	Slope	1.05	1.05	0.95	1.04	1.14	0.97	1.28	7	124
Vineyards	MBE (mm)	9.9 (9.9%)	7.9 (7.9%)	0.1 (0.1%)	9.2 (9.2%)	23.5 (23.5%)	-0.6 (-0.6%)	29.4 (29.5%)	6	123
Mean station ET	MAE (mm)	17.3 (17.4%)	17.4 (17.4%)	19.3 (19.3%)	26.2 (26.3%)	28.0 (28.1%)	19.5 (19.5%)	34.3 (34.4%)	6	123
= 100 (mm)	RMSE (mm)	21.5 (21.6%)	21.1 (21.2%)	24.4 (24.4%)	32.4 (32.5%)	33.4 (33.5%)	24.6 (24.7%)	37.3 (37.4%)	6	123
	R-squared	0.81	0.81	0.73	0.63	0.74	0.71	0.71	7	124

In annual crop sites (wheat, soy, corn, and rice), individual model metrics ranged from -15.4 mm/month to 7.5 mm/month for MBE, 16.2-27.1 mm/month for MAE, and the explained variance ranged from 0.79-0.87. Of the individual models, DisALEXI had the lowest MBE (-0.4%), while PT-JPL had the best MAE (18.7%) and r2 (0.87). Table 5 provides more details about each individual model's accuracy for different crop types.

Six EC towers were located in orchards (3 almond, 1 peach, 2 citrus). Overall, the models showed a somewhat negative bias with the ensemble having a Slope of 0.90 (Table 5). The ensemble ET value showed the lowest MAE and RMSE. For orchards, individual model accuracy ranged from -12.4 mm/month to 2.7 mm/month for MBE, and 22.7 mm/month to 27.6 mm/month for MAE, and the explained variance ranged from 0.82 to 0.84. SIMS had the lowest MBE (0.8%) among the individual models of the individual models in the ensemble, SIMS also had the lowest MAE (19.6%). SSEBop has the best Slope value of 0.99. All models had r^2 values between 0.82 and 0.84. The ensemble value had the lowest MAE (18.9%) and RMSE (24.8%) and the best r^2 (0.87) in orchards.

Although one of the two citrus sites showed strong positive monthly bias, the grouped orchard statistics in Table 5 reflect all six orchard towers, not only the highlighted citrus example. Negative ensemble bias at the four non-citrus orchard sites offset the positive citrus bias, so the grouped orchard metrics remained slightly negative overall, with an ensemble slope below 1.0.

In vineyard and orchard sites, inter-model performance was more varied than in annual crops. In vineyards, the ensemble value had a Slope of 1.06 and NMAE and NRMSE of 17.4% and 21.6%, respectively (Table 5). DisALEXI closely matched the ensemble performance, likely due to its dual-source formulation that partitions the surface energy balance between soil and canopy, a feature particularly beneficial in row-structured crops like vineyards (Kustas et al., 2016). For one (of two) citrus orchards sites the ensemble tended to overstate ET. This aligns with previous findings in deficit-irrigated citrus (Dhungel et al., 2024). One of the citrus sites in our dataset (NavalCitrus) may have been affected by localized heat inputs from an adjacent solar farm, potentially leading to enhanced ET and closure correction errors not captured by the flux tower. In one of its three-year deployment, where overestimation was highest, the Naval EC site was subject to a major thinning event where the orchard was subject to 50% tree removal (Dhungel et al., 2024). This further complicates the site and strengthens the notion that it is atypical for California citrus agriculture; it also introduces additional uncertainty in the EC measurements due to the aerodynamic complexities and increased land cover heterogeneity that resulted.

In addition to crop type groups, we highlight monthly time series from four crops and regions: 1) an almond orchard in the San Joaquin Valley near the foothills in CIMIS ETo zone 12 2) a mandarin citrus orchard in the San Joaquin Valley near Visalia in ETo zone 12 (Dhungel et al., 2024); 3) a winegrape vineyard in the North Coast near Cloverdale in ETo zone 8 (Kustas et al., 2016); and 4) alfalfa site in the Sacramento–San Joaquin River Delta in ETo zone 14.

Figure 5 compares EC and OpenET for an almond site. Typically, the ensemble fell between the closed and unclosed flux ET with SSEBop and a few other models overestimating ET during the fall periods when ET rates fall off sharply. At this site, eeMETRIC consistently under-stated ET compared to the closed flux ET, instead, tracking closely to the unclosed monthly flux ET values. The remaining models and the ensemble typically fell within the closure range during growing season periods.

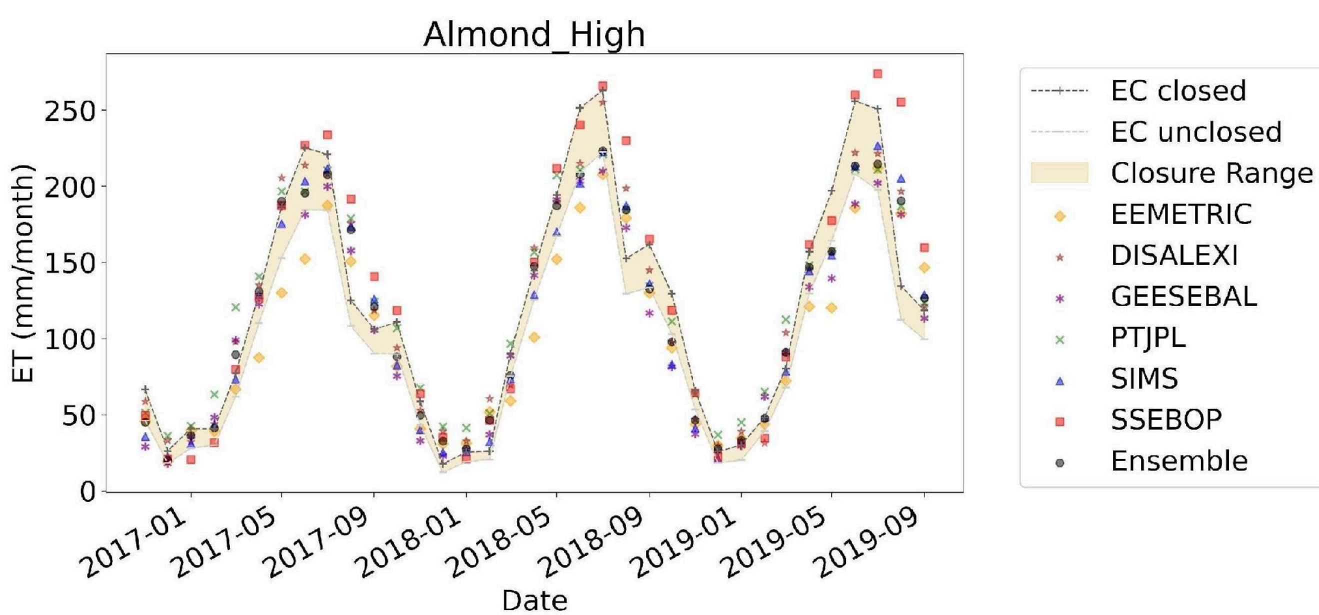


Figure 5. Time series comparison of monthly ET between OpenET models and flux ET from the Almond_High (for high salinity) site in the San Joaquin Valley foothills.

Figure 6 shows monthly ET at a mandarin citrus site. Most models were biased high during spring and fall, with closer agreement during the peak growing season. This site shows that year-to-year individual model biases can vary. For instance, highest monthly ET was frequently associated with geeSEBAL in 2021, DisALEXI in 2023, and varied among models in 2022.

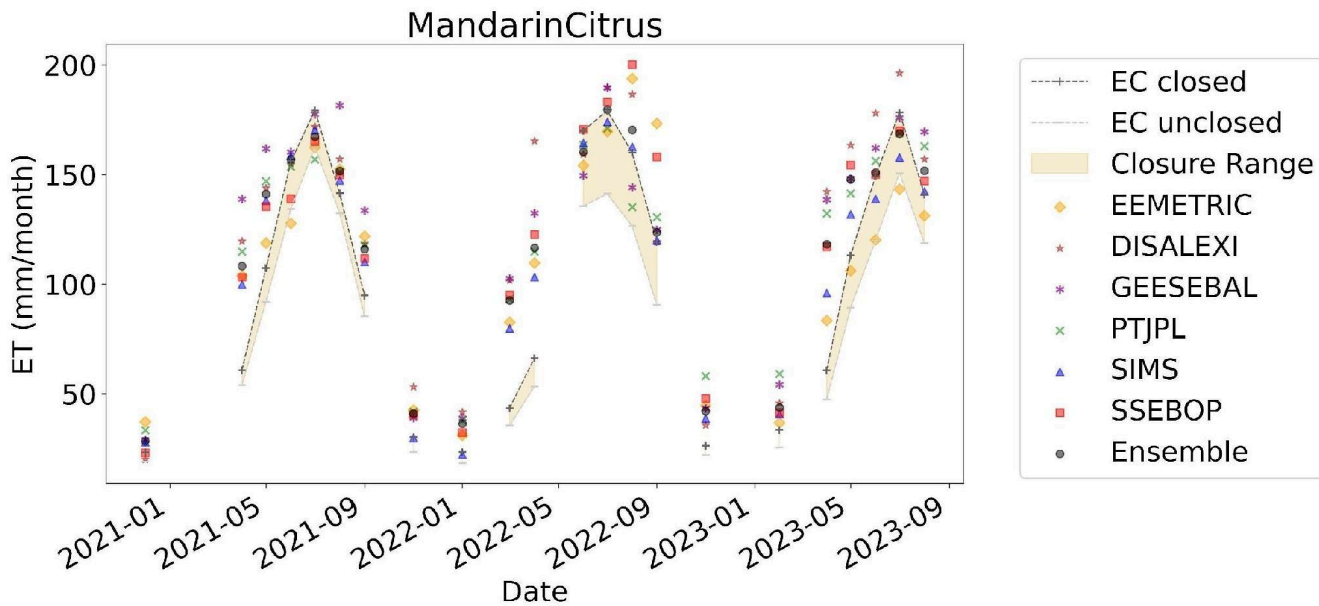


Figure 6. Time series comparison of monthly ET between OpenET models and flux from the Mandarin Citrus site near Visalia.

At a vineyard site in the North Coast region, we found increased model spread with PT-JPL showing a positive bias and eeMETRIC a negative bias (Figure 7). However, the model spread was mostly evenly distributed around the closed flux ET making the RSET values appear as random errors when taken together. However, as with several other vineyard sites, DisALEXI had slightly lower error than the ensemble value.

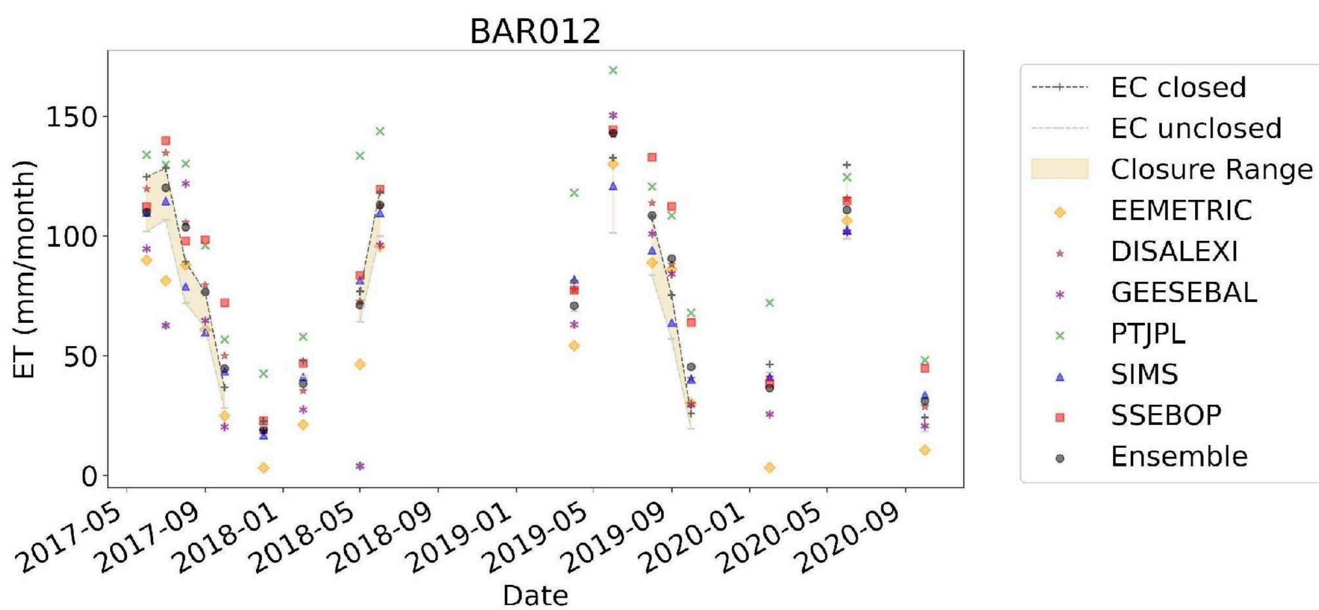


Figure 7. Time series comparison of monthly ET between OpenET models and flux ET from the BAR012 North Coast winegrape vineyard site.

A comparison of OpenET-EC ET for an alfalfa field in the Sacramento–San Joaquin River Delta is shown in Figure 8. This location is well watered and in a relatively cool and homogeneous agricultural area without advection from surrounding drylands. PT-JPL, SIMS, and eeMETRIC tend to show the best agreement in terms of Slope and RMSE with the flux tower ET. PT-JPL has a lower RMSE than the ensemble. geeSEBAL and DisALEXI show negative bias. SIMS has a slight positive bias. geeSEBAL tends to track near the unclosed flux ET during the growing season, and several models are slightly lower during the non-growing season. Overall, however, most models produce monthly ET values that fall within the flux ET closure range.

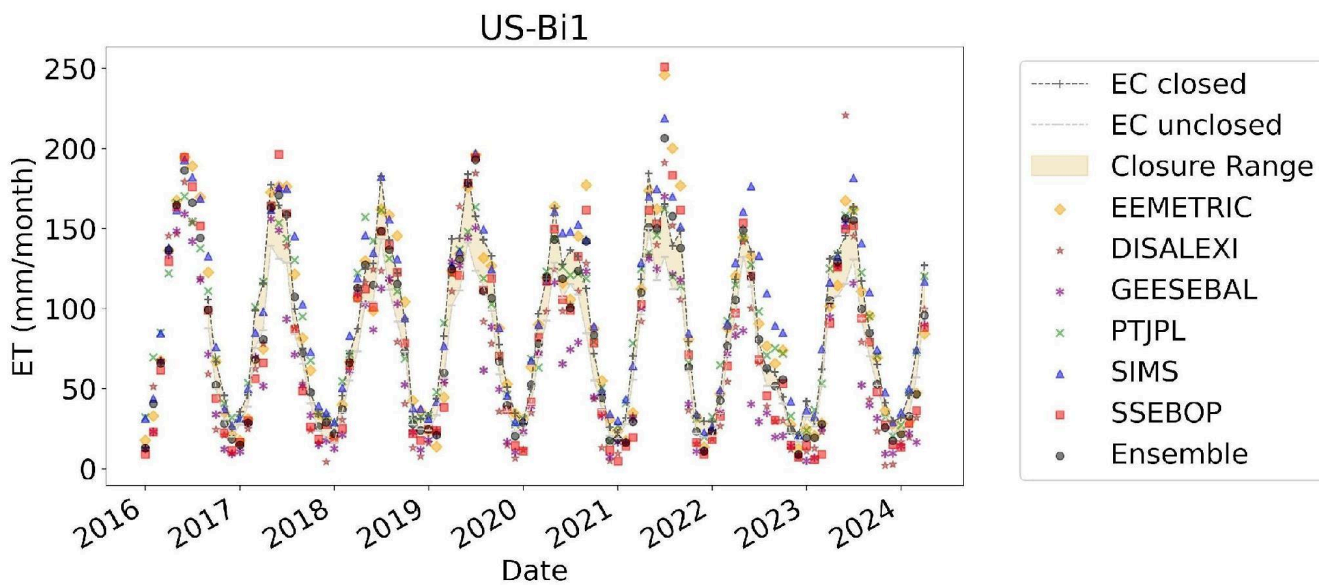


Figure 8. Time series comparison of monthly ET between OpenET models and flux ET from the US-Bi1 alfalfa site in the Delta.

From this brief review of individual sites, it is evident that each model has unique behaviors in different crops and regions. Even within a single site, changes in meteorological conditions or land surface conditions can sometimes lead to yearly variations in bias for a single model at a given site. This result provides an additional reason for the use of the ensemble approach for reducing individual model uncertainty. Technical details about the known issues and biases in different climates and land cover types for each model are further described on the OpenET Known-Issues webpage.

3.4 Error Reduction due to Ensemble Averaging and Temporal Aggregation

The ensemble value consistently had lower MAE and higher r^2 than any individual model for all temporal aggregation periods (Table 3). Slope and MBE metrics were relatively strong as well. The reduction of absolute errors in the ensemble value becomes more apparent as the temporal aggregation period increases. For example, monthly results for croplands showed substantial improvement compared to daily (date of overpass) results. At daily intervals, similar patterns are seen between models and the ensemble value, although increased variability is present. Temporal aggregation can serve to cancel some random errors (Allen et al., 2007) and reduce the effect of outliers that may exist due to daily cloud cover (and associated reliance on interpolation methods) or other high frequency error factors.

Diminishing returns were realized when aggregating from monthly to growing season intervals. This may be due to the smaller sample size. All models tended to perform similarly in terms of MAE and RMSE at growing season timesteps (NMAE ≤ 24% and NRMSE ≤ 27%), whereas at monthly timesteps there was a split with DisALEXI, PT-JPL, and SIMS having lower MAE and RMSE than geeSEBAL and SSEBop, and eeMETRIC fell in between (Table 3). Although, again, the ensemble value showed the least error at both timesteps.

3.5 Performance in Natural Ecosystems

Natural ecosystems in this analysis included shrubland, grassland, evergreen forest, mixed forest, and wetland sites (Figure 1; Table 2). Apart from wetland sites, OpenET showed substantially higher error in natural ecosystems as compared to croplands (Figure 9).

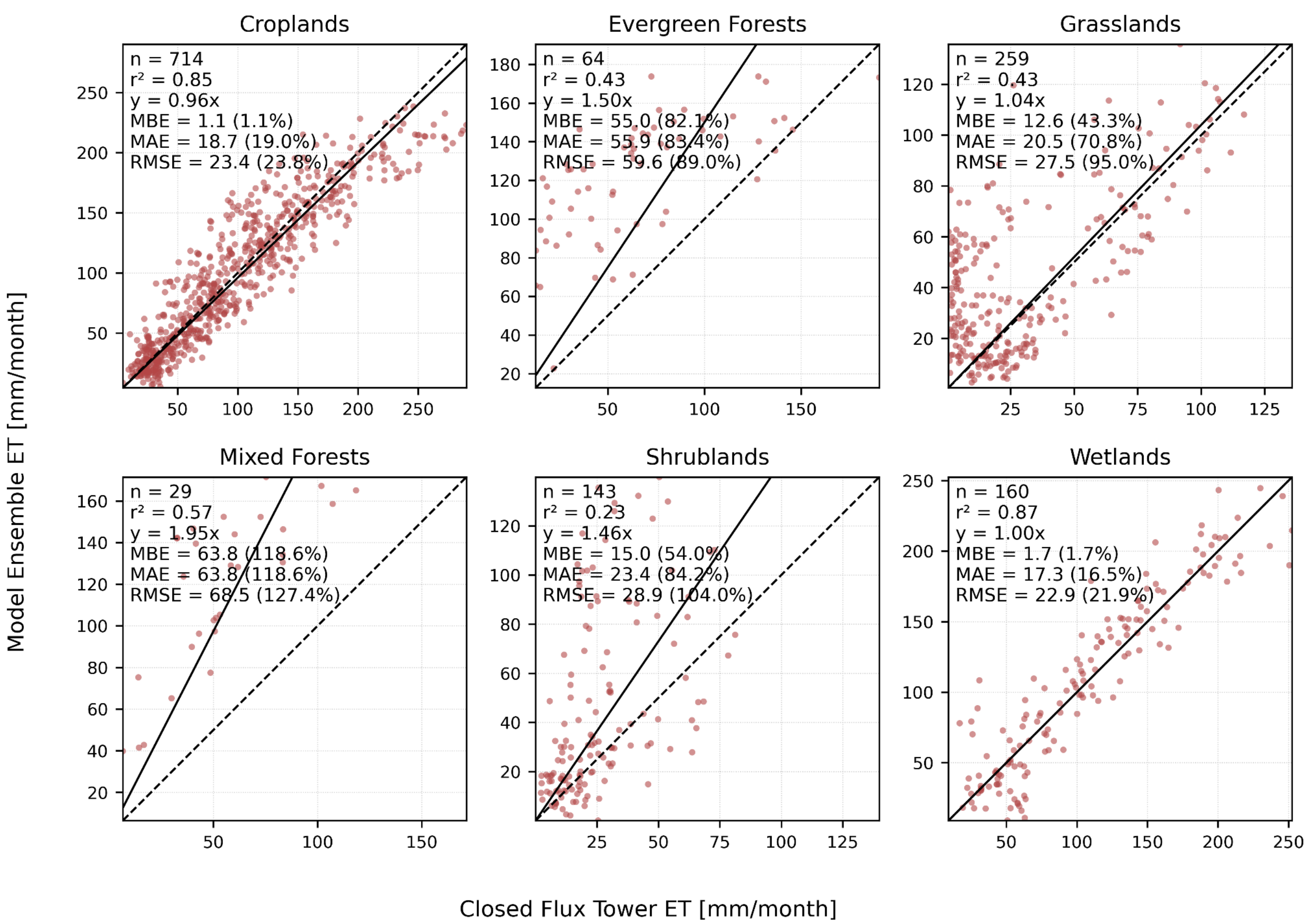


Figure 9. Monthly comparison of all paired OpenET ensemble values versus closed flux tower ET from all stations for all months of record, grouped by land cover type. Included for each model is the result of the least square linear regression forced through the origin (black solid line), r^2, number of paired model-measurements (n), and MBE, MAE, and RMSE calculated using the same methods used throughout

this report including their values normalized by the mean flux tower ET. Represented are 3 Evergreen Forests; 2 Grasslands; 1 Mixed Forest; 5 Shrublands; and 5 Wetland sites. Black dashed line is 1:1

Wetland sites showed the lowest bias (slope near 1 and MBE near 0) among the natural land cover types, with the ensemble and SSEBop both performing well in terms of all error metrics (Table 6). These sites are generally free of water stress and have relatively dense vegetation during the growing season and a higher ET range. Performance in these systems suggests that the simplified surface energy balance assumptions of SSEBop, which use reference ET to scale actual ET, may work to advantage when canopy cover is relatively uniform and dense, and soil moisture is not limited.

In contrast, shrubland and grassland, where vegetation may be sparse or patchy, or are highly responsive to episodic precipitation, posed greater challenges for all models. SSEBop was by far the best performing model in shrublands (e.g., Slope=0.95, NMBE=-7.2%), well surpassing the ensemble in all metrics. In general, however, these water-limited systems showed higher error (e.g., Figure 9 shows that the ensemble value had a NMBE ≥ 43% and NMAE ≥ 70% in shrublands and grasslands), underscoring both the difficulty of measuring and modeling in fragmented, low-ET environments. These findings point to the need for improved process representation of soil moisture controls and vegetation dynamics in shrublands. Integration of observations from ECOSTRESS to complement Landsat's morning overpass times may also help to account for increasing stomatal resistance and reduced ET rates later in the day for grassland and shrubland ecosystems in California's mediterranean climate.

Table 6. Mean monthly statistics between modeled-measured ET for stations for non-agricultural land cover types. The values for weighted mean MBE, MAE, and RMSE are also provided as percentages of the weighted mean measured monthly ET. Note that SIMS is not currently run for natural ecosystems.

Land cover type	Statistic	Ensemble	DisALEXI	eeMETRIC	geeSEBAL	PT-JPL	SSEBop	N sites	N data points
	Slope	1.5	1.43	1.44	1.65	1.4	1.59	3	64
Evergreen Forests	MBE (mm)	55.0 (82.1%)	50.0 (74.7%)	48.2 (71.9%)	65.9 (98.4%)	48.9 (72.9%)	62.8 (93.7%)	3	64
Mean station ET =	MAE (mm)	55.9 (83.4%)	52.5 (78.4%)	50.5 (75.4%)	66.2 (98.7%)	51.8 (77.4%)	63.1 (94.2%)	3	64
67.0 (mm)	RMSE (mm)	59.6 (89.0%)	56.4 (84.1%)	56.2 (83.9%)	71.0 (105.9%)	55.6 (83.0%)	67.4 (100.6%)	3	64
	R-squared	0.43	0.42	0.38	0.37	0.38	0.36	3	64
	Slope	1.04	0.93	1.07	1.1	1.33	0.91	2	259
Grasslands	MBE (mm)	12.6 (43.3%)	10.0 (34.5%)	8.4 (28.9%)	21.6 (74.6%)	25.5 (87.8%)	5.6 (19.2%)	2	259
Mean station ET =	MAE (mm)	20.6 (70.9%)	22.2 (76.6%)	17.4 (60.2%)	31.3 (108.1%)	26.4 (90.9%)	17.8 (61.5%)	2	259
29.0 (mm)	RMSE (mm)	27.5 (95.0%)	29.4 (101.2%)	23.2 (79.9%)	42.6 (146.8%)	32.8 (113.0%)	23.1 (79.7%)	2	259
	R-squared	0.43	0.31	0.61	0.16	0.58	0.39	2	259
	Slope	1.95	1.96	1.86	2.25	1.78	1.97	1	29
Mixed Forests	MBE (mm)	63.8 (118.6%)	64.9 (120.7%)	53.9 (100.2%)	81.3 (151.0%)	56.2 (104.4%)	69.2 (128.6%)	1	29
Mean station ET =	MAE (mm)	63.8 (118.6%)	64.9 (120.7%)	54.3 (100.9%)	81.3 (151.0%)	56.2 (104.4%)	69.2 (128.6%)	1	29
54.0 (mm)	RMSE (mm)	68.5 (127.4%)	69.5 (129.2%)	65.5 (121.7%)	87.6 (162.9%)	60.5 (112.4%)	72.7 (135.1%)	1	29
	R-squared	0.57	0.56	0.5	0.54	0.51	0.49	1	29
	Slope	1.46	1.34	1.52	1.65	1.71	0.95	5	143
Shrublands	MBE (mm)	15.0 (54.0%)	10.7 (38.4%)	17.8 (64.0%)	21.1 (76.0%)	24.5 (88.2%)	-2.0 (-7.2%)	5	143
Mean station ET =	MAE (mm)	23.4 (84.2%)	23.4 (84.1%)	25.4 (91.4%)	31.4 (113.2%)	27.3 (98.2%)	13.9 (49.9%)	5	143
28.0 (mm)	RMSE (mm)	28.9 (104.0%)	28.1 (101.0%)	31.3 (112.8%)	40.4 (145.5%)	31.2 (112.4%)	17.1 (61.6%)	5	143
	R-squared	0.23	0.25	0.26	0.16	0.22	0.44	5	143
	Slope	1.01	1.07	0.97	0.96	0.9	1.06	5	160
Wetlands	MBE (mm)	1.7 (1.7%)	8.1 (7.7%)	-2.6 (-2.4%)	-3.1 (-3.0%)	-2.0 (-1.9%)	5.5 (5.2%)	4	158
Mean station ET =	MAE (mm)	17.3 (16.5%)	21.1 (20.1%)	19.8 (18.9%)	24.9 (23.8%)	25.8 (24.7%)	18.7 (17.9%)	4	158
105.0 (mm)	RMSE (mm)	22.9 (21.9%)	26.7 (25.5%)	25.3 (24.1%)	31.9 (30.4%)	36.2 (34.6%)	24.2 (23.1%)	4	158
	R-squared	0.87	0.86	0.84	0.77	0.64	0.89	5	160

In forests, ET was consistently overstated (Figure 9), similar to the broader geographic findings of Volk et al. (2024). Models may over-respond to increased surface temperature or vegetation greenness in evergreen forests, leading to inflated ET when actual transpiration remains moderate. This may be related to canopy complexity, biomass heat storage, and climate buffering effects that reduce the direct relationship between surface temperature and transpiration (e.g., Lindroth et al., 2010; Trebs et al., 2021). These findings should be viewed as preliminary, given the limited number of sites, but they suggest a need for model refinement, integration of additional satellite sensors with an afternoon overpass time, or development of post-processing corrections.

3.6 Significance and Interpretation of Error

The average monthly error for the ensemble ET value across all California cropland sites evaluated, including the two spatially complex sites, was approximately 1.1% (MBE), 19.0% (MAE) and 23.8% (RMSE), normalized by the mean closed flux ET (Table 3). These values fall within the target accuracy range for many OpenET applications and use cases (Table 1), which are typically 10–20% for monthly ET and 15–25% for daily ET, (Volk et al., 2024; Melton et al., 2022) and are within expected uncertainties associated with both EC and RSET methods (Allen et al., 2011).

Across croplands, the ensemble meets the Table 1 target accuracy thresholds for daily, monthly, and annual ET. Seasonal targets were not defined by the user group and are not evaluated here. It should be recognized that while the stated tolerances of Table 1 are fixed, the error metrics from this study are averages and actual results for specific instances may vary and could be more or less favorable. Notably, these error rates are comparable to the inherent uncertainty of EC flux data in general, which can range from 10–15% under favorable conditions, as well as typical RSET uncertainties of 5-15% for energy balance models (10-30% for reflectance-based approaches) (Allen et al., 2011). It is possible that these typical uncertainties have narrowed due to technical advances made since publication of that report, and we would encourage evaluation of the current study in light of any updated expectations that might appear.

We note that with the exclusion of two complex EC sites (Figure 3 and Table 4) the cropland accuracy for the ensemble value's daily ET, in terms of MBE, improves from 3.1% to 0.1% and MAE improves from 24.0% to 22.2%. At monthly timescales, exclusion of these sites results in an improvement in the OpenET ensemble value's MAE from 19.0% to 17.3%. At both timescales the MBE values are lower than those in the U.S. study of Volk et al. (2024), and the MAE values are similar.

Most of the errors observed in croplands were random rather than systematic bias (i.e., not consistently positive or negative), which is important since such errors tend to average across spatial and temporal aggregation. This effect enhances the utility of OpenET for seasonal and annual accounting. Additionally, there is spatial uncertainty associated with both eddy covariance measurements and the associated 30 m satellite pixel sampling areas, however, we applied methods to reduce those mismatches using dynamic flux footprint predictions for the 13 out of 45 total sites (10 of which were croplands) for which the required input data was available from the flux tower (Volk et al., 2023a), and evaluated additional wind direction filtering for all sites (Appendix B). Overall, the OpenET ensemble data for croplands appear to be approaching the practical limits of validation precision due to constraints imposed by measurement uncertainty associated with current state-of-the-art EC (or other ground-based) systems (AGU Chapman Conference on Energy Balance Closure Problem, 2025). However, comparisons for individual sites and key crop and land cover types (e.g., citrus and grasslands) provide data that will inform research priorities for the OpenET Science Team moving forward.

The majority of EC deployments to date have tended to emphasize commodity crops, both in California and beyond. California in particular would benefit from additional ground-based monitoring of specialty crops since it is a major producer, and particularly short-season, high-value vegetables which have received relatively little attention in this regard. For natural ecosystems (other than wetlands), continued data collection and model refinement efforts are recommended, at least for the model set currently operating under OpenET.

4.0 Conclusions

This report evaluated the accuracy of OpenET evapotranspiration (ET) data using high-quality eddy covariance (EC) data from 45 sites across California comprising >1,300 site-months of paired data, including 28 agricultural locations (>700 site-months). Performance metrics included Slope, mean absolute error (MAE), mean bias error (MBE), root mean square error (RMSE) and r^2. For combined croplands, the results for the OpenET ensemble ET value were: MBE of 1.1% and MAE of 19.0% relative to EC measurements of monthly ET. When two complex sites were removed and considered separately, MBE changed to -1.7% and MAE was reduced to 17.3%. These averages were within the error tolerance specified by OpenET user-groups for various water management applications. Overall study findings were generally consistent with prior performance characterization for CONUS. Collection of additional high quality, open access ground-based ET data is recommended to more thoroughly evaluate performance for individual crop types of interest, such as vegetable crops and citrus orchards, as well as for key groundwater basins across the state.

Agreement among OpenET constituent models was strongest during peak growing-season months. Among individual models, eeMETRIC had the lowest MBE and SIMS had lowest MAE and RMSE for combined crops for monthly data. The ensemble value frequently outperformed individual models for combined crops and across individual crop types (annuals, orchards, vineyards) for most metrics. This finding underscores the advantages of the ensemble approach, which tends to reduce the effect of model biases and random errors, and thus stabilize performance across crop types and time periods.

Results for natural vegetation types were more mixed. Performance on wetland sites was similar to croplands. Grasslands, forests, and shrublands, however, yielded reduced accuracy statistics. Additional research on those land cover types is needed to help reduce model bias error and random error. Additional research on characterization of expected error in EC measurements, and the impact of use of different flux tower footprints to subset the satellite-derived ET data would also help to increase understanding of differences between satellite and EC-based ET datasets.

Considering the state's diversity of growing regions and crops, the current body of publicly available EC or other *in situ* ET measurements could be expanded. As additional ground-based ET data become available, continued model intercomparisons and performance analyses may support future model refinements to OpenET and similar satellite-based ET mapping technologies.

References

Allen, R. G., Pereira, L. S., Howell, T. A., & Jensen, M. E. (2011). Evapotranspiration information reporting: I. Factors governing measurement accuracy. Agricultural Water Management, 98, 899–920.

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

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

Baldocchi, D. (2014). Measuring fluxes of trace gases and energy between ecosystems and the atmosphere: The state and future of the eddy covariance method. Global Change Biology, 20, 3600–3609.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., et al. (2001). FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bulletin of the American Meteorological Society, 82, 2415–2434.

Billesbach, D. P., Arkebauer, T. J., & Sullivan, R. C. (2024). Intercomparison of sensible and latent heat flux measurements from combined eddy covariance, energy balance, and Bowen ratio methods above a grassland prairie. Scientific Reports, 14, 21866. https://doi.org/10.1038/s41598-024-67911-z

Crawford, C. J., Roy, D. P., Arab, S., Barnes, C., Vermote, E., Hulley, G., Gerace, A., Choate, M., Engebretson, C., et al. (2023). The 50-year Landsat Collection 2 archive. Science of Remote Sensing, 100103. https://doi.org/10.1016/j.srs.2023.100103

Dhungel, R., Anderson, R. G., French, A. N., Skaggs, T. H., Ajami, H., & Wang, D. (2024). Intercomparison of citrus evapotranspiration among eddy covariance, OpenET ensemble models, and the Water and Energy Balance Model (BAITSSS). Agricultural Water Management, 304, 109066.

Fisher, J. B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., McCabe, M. F., Hook, S., Baldocchi, D., Townsend, P. A., Kilic, A., Tu, K., Miralles, D. D., Perret, J., Lagouarde, J.-P., Waliser, D., Purdy, A. J., French, A., Schimel, D., Famiglietti, J. S., Stephens, G., & Wood, E. F. (2017). The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. Water Resources Research, 53(4), 2618–2626.

Fisher, J. B., Tu, K. P., & 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, 901–919.

Foken, T. (2008). The energy balance closure problem: An overview. Ecological Applications, 18, 1351–1367.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18–27.

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

Harrison, L. P. (1963). Fundamental concepts and definitions relating to humidity. In A. Wexler (Ed.), Humidity and Moisture (Vol. 3). Reinhold.

Howitt, R. E., Medellín-Azuara, J., MacEwan, D., Lund, J. R., & Sumner, D. A. (2015). Economic analysis of the 2015 drought for California agriculture. UC Davis Center for Watershed Sciences.

Huntington, J., Pearson, C., Minor, B., Volk, J., Morton, C., Melton, F., & Allen, R. (2022). Upper Colorado River Basin OpenET intercomparison summary. https://doi.org/10.13140/RG.2.2.21605.88808

Ingwersen, J., Imukova, K., Högy, P., & Streck, T. (2015). On the use of the post-closure methods uncertainty band to evaluate the performance of land surface models against eddy covariance flux data. Biogeosciences, 12, 2311–2326.

Fares, S., Weber, R., Park, J.-H., Gentner, D., Karlik, J., & Goldstein, A. H. (2012). Ozone deposition to an orange orchard: Partitioning between stomatal and non-stomatal sinks. Environmental Pollution, 169, 258–266.

Karimi, P., Bastiaanssen, W. G. M., & Molden, D. (2013). Water Accounting Plus (WA+): A water accounting procedure for complex river basins based on satellite measurements. Hydrology and Earth System Sciences, 17, 2459–2472.

Kljun, N., Calanca, P., Rotach, M. W., & Schmid, H. P. (2015). A simple two-dimensional parametrisation for flux footprint prediction (FFP). Geoscientific Model Development, 8, 3695–3713.

Knipper, K. R., Kustas, W. P., Anderson, M. C., Alsina, M. M., Hain, C. R., Alfieri, J. G., Prueger, J. H., Gao, F., McKee, L. G., & Sanchez, L. A. (2019). Using high-spatiotemporal thermal satellite ET retrievals for operational water use and stress monitoring in a California vineyard. Remote Sensing, 11, 2124.

Kustas, W. P., Nieto, H., Morillas, L., Anderson, M. C., Alfieri, J. G., Hipps, L. E., Villagarcía, L., Domingo, F., & Garcia, M. (2016). Revisiting the paper "Using radiometric surface temperature for surface energy flux estimation in Mediterranean drylands from a two-source perspective." Remote Sensing of Environment, 184, 645–653.

Laipelt, L., Kayser, R. H. B., Fleischmann, A. S., Ruhoff, A., Bastiaanssen, W., Erickson, T. A., & 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, 81–96.

Leys, C., Ley, C., Klein, O., Bernard, P., & 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, 764–766.

Lindroth, A., Mölder, M., & Lagergren, F. (2010). Heat storage in forest biomass improves energy balance closure. Biogeosciences, 7(1), 301–313.

Mauder, M., Foken, T., & Cuxart, J. (2020). Surface-energy-balance closure over land: A review. Boundary-Layer Meteorology, 177, 395–426.

Melton, F. S., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., Erickson, T., Allen, R., Anderson, M., Fisher, J. B., et al. (2022). OpenET: Filling a critical data gap in water management for the western United States. JAWRA Journal of the American Water Resources Association, 58, 971–994.

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., et al. (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, 1709–1721.

Novick, K. A., Biederman, J. A., Desai, A. R., Litvak, M. E., Moore, D. J. P., Scott, R. L., & Torn, M. S. (2018). The AmeriFlux network: A coalition of the willing. Agricultural and Forest Meteorology, 249, 444–456.

Obrecht, N. A. (2019). Sample size weighting follows a curvilinear function. Journal of Experimental Psychology: Learning, Memory, and Cognition, 45, 614.

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

Pereira, L. S., Paredes, P., Melton, F., Johnson, L., Wang, T., López-Urrea, R., Cancela, J. J., & 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, 106197.

Rousseeuw, P. J., & Hubert, M. (2011). Robust statistics for outlier detection. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1, 73–79.

Schwankl, L. J., Prichard, T. L., Hanson, B. R., & Elkins, R. B. (2007). *Understanding your orchard's water requirements* (ANR Publication 8212). University of California Agriculture and Natural Resources.

Senay, G. B., Friedrichs, M., Singh, R. K., & Velpuri, N. M. (2016). Evaluating Landsat 8 evapotranspiration for water use mapping in the Colorado River Basin. Remote Sensing of Environment, 185, 171–185.

Senay, G. B., Parrish, G. E. L., Schauer, M., Friedrichs, M., Khand, K., Boiko, O., Kagone, S., Dittmeier, R., Arab, S., & Ji, L. (2023). Improving the Operational Simplified Surface Energy Balance evapotranspiration model using the forcing and normalizing operation. Remote Sensing, 15, 260.

Trebs, I., Mallick, K., Bhattarai, N., Sulis, M., Cleverly, J., Woodgate, W., Silberstein, R., Hinko-Najera, N., Beringer, J., Meyer, W. S., Su, Z., & Boulet, G. (2021). The role of aerodynamic resistance in thermal remote sensing-based evapotranspiration models. Remote Sensing of Environment, 264, 112602.

Volk, J. M., Huntington, J. L., Allen, R. G., Melton, F. S., Anderson, M. C., & Kilic, A. (2021). fluxdataqaqc: A Python package for energy balance closure and post-processing of eddy flux data. Journal of Open Source Software, 6(66), 3418. https://doi.org/10.21105/joss.03418

Volk, J. M., Huntington, J. L., Melton, F. S., Allen, R., Anderson, M., Fisher, J. B., Kilic, A., Ruhoff, A., Senay, G. B., Minor, B., et al. (2024). Assessing the accuracy of OpenET satellite-based

evapotranspiration data to support water resource and land management applications. Nature Water, 2, 193–205.
Volk, J. M., Huntington, J., Melton, F. S., Allen, R., Anderson, M. C., Fisher, J. B., Kilic, A., Senay, G., Halverson, G., Knipper, K., Minor, B., Pearson, C., Wang, T., Yang, Y., Evett, S., French, A. N., Jasoni, R., & Kustas, W. (2023a). Development of a benchmark eddy flux evapotranspiration dataset for evaluation of satellite-driven evapotranspiration models over the CONUS. Agricultural and Forest Meteorology, 331, 109307.
Volk, J. M., Huntington, J. L., Melton, F. S., Minor, B., Wang, T., Anapalli, S., Anderson, R. G., Evett, S., French, A., Jasoni, R., et al. (2023b). Post-processed data and graphical tools for a CONUS-wide eddy flux evapotranspiration dataset. Data in Brief, 48, 109274.
Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., & Lohmann, D. (2012). Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. Journal of Geophysical Research: Atmospheres, 117(3). https://doi.org/10.1029/2011JD016051

Appendix

Appendix A: Evaluation of OpenET Collection v2.1 Data

To supplement this intercomparison and accuracy assessment, we include select results using the same data and methods, but using OpenET Collection v2.1 data (released Dec 2025) as opposed to 2.0 (please see https://etdata.org/methods/ for additional information on OpenET Collections). We briefly highlight cropland results due to their greater importance to water budgeting in the state and because there are not many observational networks in natural ecosystems currently available that meet our data quality criteria for eddy covariance data. Table A1 shows analogous results of daily, monthly, growing season, and water year statistics for all cropland stations used in this report with the OpenET Collection v2.1 data.

Table A1. Accuracy metrics between modeled-measured ET for California cropland closed flux tower ET using paired data aggregated at daily, monthly, growing season, and water year temporal periods from OpenET Collection v2.1. Growing season periods were determined for each site using the cumulative growing degree day and killing frost approach described in Volk et al. (2023a); growing season total ET values were then accumulated using the monthly ET data from the models and paired with the EC monthly data. The values for weighted mean MBE, MAE, and RMSE are also provided as percentages of the weighted mean measured monthly ET referred to as "normalized" error metrics in the text (i.e., NMBE, NMAE, and NRMSE). For MBE, MAE, and RMSE, we required a minimum of 6 paired model-observation data points per flux station for daily metrics and 3 paired data points per station for monthly metrics. This is why the number of data points differs slightly for those statistics compared to Slope and R-squared which had no minimum criteria following Volk et al. (2024). Daily results for SIMS exclude its soil evaporation component, which applies to monthly and growing season aggregations only.

Aggregation	Mean measured ET	Statistic	Ensemble	DisALEXI	eeMETRIC	geeSEBAL	PT-JPL	SIMS	SSEBop	N sites	N data points
Daily	3.5 (mm)	Slope	0.95	0.96	0.95	0.87	0.98	0.94	1.01	28	3287
		MBE (mm)	0.11 (3.3%)	0.12 (3.5%)	0.12 (3.4%)	-0.19 (-5.6%)	0.45 (12.9%)	0.05 (1.5%)	0.18 (5.2%)	27	3285
		MAE (mm)	0.82 (23.5%)	0.90 (26.0%)	1.06 (30.4%)	1.16 (33.3%)	1.06 (30.6%)	0.86 (24.8%)	0.97 (27.9%)	27	3285
		RMSE (mm	1.04 (30.1%)	1.14 (33.0%)	1.33 (38.4%)	1.48 (42.5%)	1.31 (37.8%)	1.11 (32.0%)	1.23 (35.4%)	27	3285
		R-squared	0.77	0.73	0.67	0.6	0.68	0.73	0.71	28	3287
Monthly	99 (mm)	Slope	0.96	0.97	0.96	0.87	1	0.97	0.99	25	708
		MBE (mm)	2.2 (2.2%)	1.9 (2.0%)	0.8 (0.8%)	-7.2 (-7.3%)	11.4 (11.6%)	2.6 (2.6%)	2.6 (2.6%)	23	706
		MAE (mm)	18.2 (18.5%)	20.6 (20.9%)	23.3 (23.6%)	26.2 (26.5%)	23.1 (23.4%)	20.5 (20.8%)	24.5 (24.9%)	23	706
		RMSE (mm)	22.9 (23.2%)	25.3 (25.6%)	29.3 (29.7%)	32.2 (32.7%)	28.4 (28.8%)	26.2 (26.6%)	29.8 (30.2%)	23	706
		R-squared	0.86	0.83	0.78	0.76	0.81	0.82	0.78	25	708
Growing season	692 (mm)	Slope	1	0.99	1.02	0.87	1.08	1.01	0.99	24	80
		MBE (mm)	50.7 (7.3%)	50.5 (7.3%)	56.9 (8.2%)	-34.3 (-5.0%)	119.3 (17.2%)	60.4 (8.7%)	47.1 (6.8%)	24	80
		MAE (mm)	119.2 (17.2%)	143.5 (20.7%)	146.2 (21.1%)	160.1 (23.1%)	158.8 (22.9%)	151.3 (21.9%)	147.1 (21.3%)	24	80
		RMSE (mm	154.5 (22.3%)	165.3 (23.9%)	182.9 (26.4%)	175.4 (25.4%)	194.9 (28.2%)	187.4 (27.1%)	182.4 (26.4%)	24	80
		R-squared	0.82	0.79	0.77	0.75	0.8	0.75	0.75	24	80
Water year	1214 (mm)	Slope	0.94	0.92	0.97	0.79	1.01	0.98	0.92	8	28
		MBE (mm)	-77.0 (-6.3%)	-101.8 (-8.4%)	-56.2 (-4.6%)	-248.4 (-20.5%)	12.2 (1.0%)	-25.5 (-2.1%)	-102.1 (-8.4%)	8	28
		MAE (mm)	86.8 (7.1%)	150.6 (12.4%)	151.2 (12.5%)	248.4 (20.5%)	99.2 (8.2%)	121.2 (10.0%)	130.9 (10.8%)	8	28
		RMSE (mm)	91.9 (7.6%)	154.9 (12.8%)	161.6 (13.3%)	254.6 (21.0%)	105.5 (8.7%)	127.8 (10.5%)	138.6 (11.4%)	8	28
		R-squared	0.91	0.7	0.45	0.8	0.77	0.57	0.7	8	28

In comparison with Table 3, which shows methods and measured ET data using Collection 2.0 OpenET data, we see improvements in OpenET Collection v2.1 for croplands across all timesteps. SSEBop showed the largest improvements in croplands from Collection 2.0 with a reduction of monthly MBE from 5.2 to 2.6 mm/month (5.3 to 2.6%) and MAE from 27.5 to 24.5 mm/month (28 to 24.9%). DisALEXI and the ensemble ET value also had noticeable improvements from Collection v2.1 compared to v2.0 in croplands. Some minor declines in SIMS performance were observed. Figure A1 shows daily and monthly comparisons for all cropland sites used in the report side by side with Collection v2.0 and v2.1, as well as showing Collection v2.1 with and without the two complex sites where the models showed unusually large biases. More description of why it is useful to exclude these sites is given in section 3.1 and shown for Collection v2.0 in Figure 3.

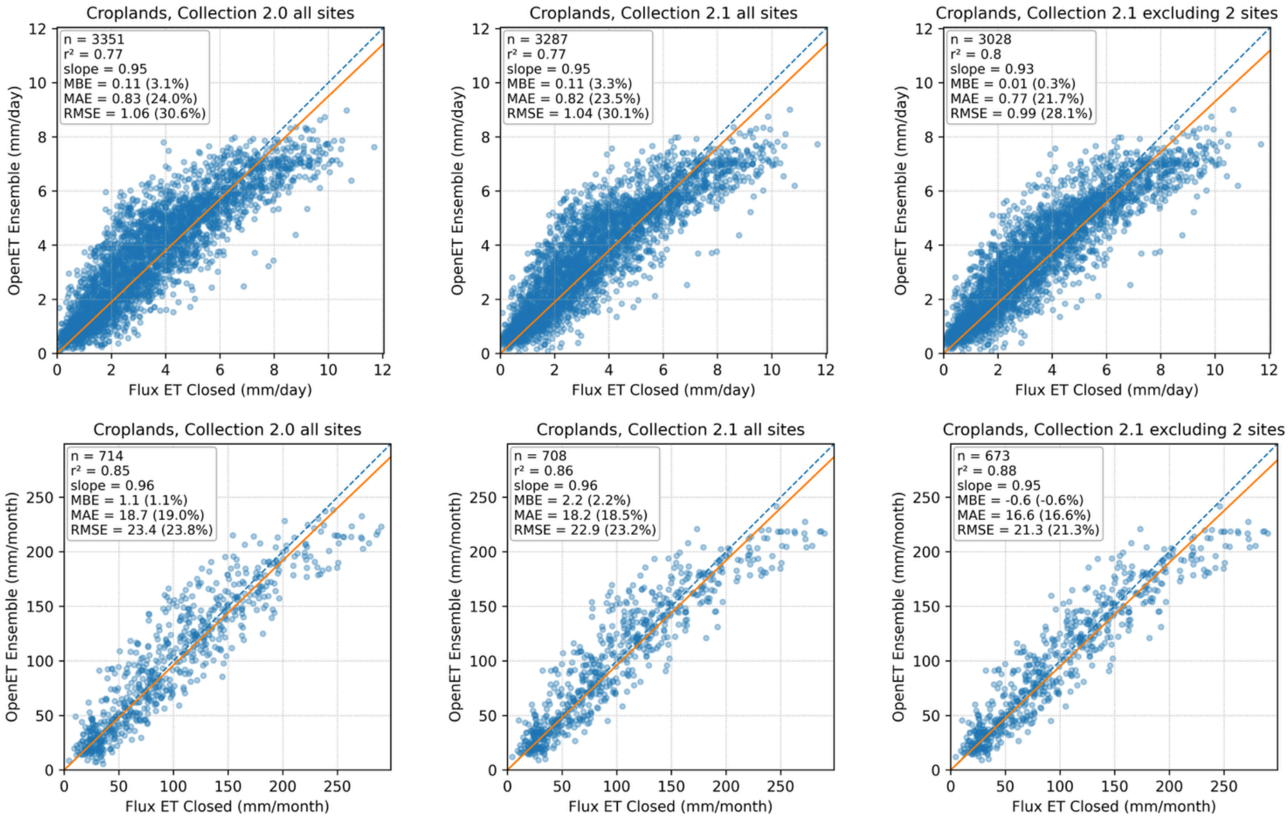


Figure A1. Comparisons between the OpenET ensemble daily (top) and monthly (bottom) ET and eddy covariance ET for all cropland stations using Collection v2.0 with all sites included (left column), Collection v2.1 with all sites included (middle column), and Collection v2.1 with exclusion of two complex sites (NavalCitrus and US-Snd; right column). Included are the same summary statistics that are used throughout this report: r^2, Slope through origin, MBE, MAE, and RMSE in mm and normalized by the mean flux tower ET.

As with OpenET Collection v2.0, exclusion of two complex sites (NavalCitrus and US-Snd) improves the accuracy results for OpenET Collection v2.1. The accuracy statistics summarized in Table A2 indicate that the ensemble ET value, as well as the individual models, improves across nearly all accuracy metrics and time steps.

Table A2. Accuracy metrics between modeled-measured ET for California cropland closed flux tower ET excluding two complex sites (NavalCitrus and US-Snd) using paired data aggregated at daily, monthly, growing season, and water year temporal periods from OpenET Collection v2.1. See description of Table 3 for more details on how data are aggregated and how statistics are computed.

Aggregation	Mean measured ET	Statistic	Ensemble	DisALEXI	eeMETRIC	geeSEBAL	PT-JPL	SIMS	SSEBop	N sites	N data points
Daily	3.5 (mm)	Slope	0.93	0.95	0.93	0.85	0.96	0.92	1	26	3028
		MBE (mm)	0.01 (0.3%)	0.03 (0.7%)	0.01 (0.3%)	-0.31 (-8.7%)	0.33 (9.3%)	-0.06 (-1.6%)	0.11 (3.0%)	25	3026
		MAE (mm)	0.77 (21.7%)	0.87 (24.4%)	1.01 (28.5%)	1.14 (32.3%)	0.99 (28.0%)	0.82 (23.1%)	0.96 (27.0%)	25	3026
		RMSE (mm)	0.99 (28.1%)	1.11 (31.3%)	1.29 (36.3%)	1.46 (41.3%)	1.25 (35.3%)	1.06 (30.0%)	1.21 (34.3%)	25	3026
		R-squared	0.8	0.76	0.7	0.64	0.72	0.77	0.73	26	3028
Monthly	100 (mm)	Slope	0.95	0.96	0.94	0.86	0.99	0.96	0.98	23	673
		MBE (mm)	-0.6 (-0.6%)	-0.8 (-0.8%)	-2.3 (-2.3%)	-10.4 (-10.4%)	8.6 (8.6%)	-0.2 (-0.2%)	0.6 (0.6%)	21	671
		MAE (mm)	16.6 (16.6%)	19.2 (19.2%)	21.6 (21.7%)	25.4 (25.5%)	21.1 (21.1%)	19.1 (19.1%)	24.1 (24.1%)	21	671
		RMSE (mm)	21.3 (21.3%)	23.9 (23.9%)	27.6 (27.7%)	31.5 (31.6%)	26.5 (26.6%)	24.8 (24.8%)	29.4 (29.5%)	21	671
		R-squared	0.88	0.85	0.81	0.79	0.84	0.84	0.79	23	673
Growing season	741 (mm)	Slope	0.99	0.98	1.01	0.86	1.07	1.01	0.98	22	70
		MBE (mm)	39.0 (5.3%)	39.6 (5.3%)	44.8 (6.0%)	-53.5 (-7.2%)	111.8 (15.1%)	50.5 (6.8%)	37.9 (5.1%)	22	70
		MAE (mm)	115.4 (15.6%)	143.4 (19.3%)	144.5 (19.5%)	163.4 (22.1%)	155.9 (21.0%)	151.9 (20.5%)	149.6 (20.2%)	22	70
		RMSE (mm)	153.2 (20.7%)	166.5 (22.5%)	183.6 (24.8%)	179.5 (24.2%)	194.4 (26.2%)	190.3 (25.7%)	187.2 (25.3%)	22	70
		R-squared	0.79	0.75	0.73	0.72	0.76	0.7	0.69	22	70
Water year	1214 (mm)	Slope	0.94	0.92	0.97	0.79	1.01	0.98	0.92	8	28
		MBE (mm)	-77.0 (-6.3%)	-101.8 (-8.4%)	-56.2 (-4.6%)	-248.4 (-20.5%)	12.2 (1.0%)	-25.5 (-2.1%)	-102.1 (-8.4%)	8	28
		MAE (mm)	86.8 (7.1%)	150.6 (12.4%)	151.2 (12.5%)	248.4 (20.5%)	99.2 (8.2%)	121.2 (10.0%)	130.9 (10.8%)	8	28
		RMSE (mm)	91.9 (7.6%)	154.9 (12.8%)	161.6 (13.3%)	254.6 (21.0%)	105.5 (8.7%)	127.8 (10.5%)	138.6 (11.4%)	8	28
		R-squared	0.91	0.7	0.45	0.8	0.77	0.57	0.7	8	28

Appendix B: Evaluation of Additional Wind Direction Flags to Improve Land-use Representation

Logistical constraints often limit where an eddy covariance tower can be sited in a commercial agricultural field. Often, scientists need to balance considerations related to the crop height and predominant wind direction to ensure appropriate fetch, while minimizing the impact on agricultural operations in the field. As a result, towers may occasionally be located near the edge of a field, adjacent to parcels with a different land cover or irrigation management scheme, or different phenology and timing of planting and harvest. Matching the source area to the surface flux measurements remains an important step to evaluate the accuracy of OpenET data relative to micro-meteorological ground observations. Flux footprint modeling facilitates matching the flux measurements with satellite observations within the measurement source area (Kljun et al., 2015), but additional filtering for wind direction ensures the sampled area corresponds to the intended land use type.

For each tower in California, we reviewed the tower location in relation to historical aerial or satellite imagery from Google Earth for each site's measurement records. Wind direction filters were created that were informed by the tower location in relation to the intended measurement unit's boundaries, instrumentation height, crop type, and adjacent land cover. Winds that originated from the flagged directions were used to mask ground-observations. A total of 235 days, 77 months, 4 growing seasons, and 3 water years of paired observations were removed because of wind filtering. Table B1 shows the accuracy statistics across daily, monthly, growing season, and water year. Overall, additional wind filtering resulted in minimal changes to the accuracy relative to the unfiltered data (Table 3). However, wind filtering also reduced the number of data points available for use in comparisons at the monthly, growing season, and water year timescales due to the relatively small number of sites available and the gap-filling constraints established in Volk et al (2023a) and used within this study. While small improvements in the overall relative error are seen for daily, monthly, and water year time aggregations, wind filtering can have a meaningful and positive impact on the accuracy metrics for individual sites. The changes in the overall relative error are within the uncertainty of both ground observations and models. Table B2 provides wind filtering flags and corresponding notes to justify the wind directions used.

Table B1. Accuracy metrics between modeled-measured ET for California cropland closed flux tower ET using paired data that was filtered for wind direction aggregated at daily, monthly, growing season, and water year temporal periods from OpenET Collection v2.0. See description of Table 3 for more details on how data are aggregated and how statistics are computed.

Aggregation	Mean measured ET	Statistic	Ensemble	DisALEXI	eeMETRIC	geeSEBAL	PT-JPL	SIMS	SSEBop	N sites	N data points
Daily	3.6 (mm)	Slope	0.95	0.96	0.96	0.87	0.96	0.94	1.08	27	3067
		MBE (mm)	0.11 (3.2%)	0.10 (2.9%)	0.17 (4.6%)	-0.21 (-5.9%)	0.34 (9.4%)	0.03 (0.9%)	0.42 (11.6%)	26	3064
		MAE (mm)	0.85 (23.7%)	0.95 (26.4%)	1.07 (29.8%)	1.17 (32.6%)	1.03 (28.8%)	0.90 (25.1%)	1.14 (31.8%)	26	3064
		RMSE (mm)	1.08 (29.9%)	1.19 (33.2%)	1.35 (37.5%)	1.48 (41.3%)	1.29 (36.0%)	1.15 (32.0%)	1.42 (39.5%)	26	3064
		R-squared	0.76	0.71	0.66	0.6	0.68	0.72	0.68	27	3067
Monthly	105 (mm)	Slope	0.96	0.97	0.95	0.88	0.99	0.97	1.04	22	633
		MBE (mm)	1.4 (1.3%)	1.8 (1.7%)	0.1 (0.1%)	-7.6 (-7.3%)	8.5 (8.1%)	1.5 (1.4%)	7.1 (6.7%)	21	632
		MAE (mm)	19.8 (18.8%)	22.7 (21.7%)	23.8 (22.7%)	27.3 (26.1%)	22.2 (21.2%)	21.9 (20.8%)	29.4 (28.1%)	21	632
		RMSE (mm)	24.4 (23.3%)	27.3 (26.1%)	29.9 (28.6%)	32.9 (31.4%)	27.7 (26.4%)	27.5 (26.3%)	34.5 (32.9%)	21	632
		R-squared	0.85	0.82	0.78	0.76	0.82	0.81	0.77	22	633
Growing season	699 (mm)	Slope	1.02	1.02	1.04	0.9	1.11	1.04	1.04	21	76
		MBE (mm)	68.7 (9.8%)	68.5 (9.8%)	75.0 (10.7%)	-10.1 (-1.4%)	138.3 (19.8%)	80.0 (11.4%)	92.8 (13.3%)	21	76
		MAE (mm)	128.4 (18.4%)	149.8 (21.4%)	155.8 (22.3%)	158.6 (22.7%)	163.5 (23.4%)	153.7 (22.0%)	172.2 (24.6%)	21	76
		RMSE (mm)	157.2 (22.5%)	166.8 (23.8%)	192.0 (27.5%)	174.9 (25.0%)	198.0 (28.3%)	192.7 (27.6%)	199.6 (28.5%)	21	76
		R-squared	0.82	0.79	0.76	0.72	0.8	0.74	0.72	21	76
Water year	1174 (mm)	Slope	0.92	0.9	0.95	0.78	1.01	0.99	0.92	8	25
		MBE (mm)	-92.3 (-7.9%)	-105.3 (-9.0%)	-57.3 (-4.9%)	-250.2 (-21.3%)	25.0 (2.1%)	-13.3 (-1.1%)	-88.7 (-7.6%)	8	25
		MAE (mm)	102.3 (8.7%)	157.0 (13.4%)	157.9 (13.4%)	250.2 (21.3%)	103.2 (8.8%)	125.2 (10.7%)	149.1 (12.7%)	8	25
		RMSE (mm)	109.5 (9.3%)	161.8 (13.8%)	168.8 (14.4%)	255.3 (21.7%)	109.0 (9.3%)	130.6 (11.1%)	157.5 (13.4%)	8	25
		R-squared	0.87	0.73	0.41	0.82	0.77	0.58	0.84	8	25

Table B2. Site ID, land cover, start date, end date, and wind direction filters (WF_LT: less than & WF_GT: greater than), whether additional wind flags are needed and justification for additional wind direction flagging.

Site ID	Land cover details	Start	End	WF_LT	WF_GT	WF_NEEDED	Notes
Almond_High	Almond	10/18/16	10/4/19			NO	Landcover change after Spring 2022 requires filtering if record extended
Almond_Low	Almond	10/11/16	10/4/19	225	90	YES	60 m to field border
Almond_Med	Almond	10/2/16	10/4/19			NO	Landcover change after 2022
BAR007	Vineyard	2/6/20	12/22/20	270	225	YES	80 m to grassland hills
BAR012	Vineyard	5/20/17	12/23/20			NO	
MandarinCitrus	Mandarin orange	11/23/20	9/13/23			NO	
MB_BRC_S1	Broccoli	7/26/23	9/28/23			NO	
MB_BRC_S2	Broccoli	7/27/24	10/15/24			NO	
MB_CEL	Celery	7/6/19	10/21/20			NO	80 m to other row crops
MB_LET_S1	Lettuce	4/20/23	6/8/23			NO	
MB_LET_S2	Lettuce	4/15/24	6/10/24			NO	
MB_Pch	Peach	4/7/12	12/12/15			NO	
MB_WG_ZAB	Winegrape	4/5/23	1/28/25			NO	
NavalCitrus	Navel orange	11/22/20	9/13/23	180	45	YES	100 m to non-orchard in S and solar panels to E by more than 170 m
RIP760	Vineyard	5/22/17	12/9/20	135	45	MAYBE	same and on the same irrigation schedule
Scheid	Winegrape	4/28/20	9/17/22	145	90	YES	20m from separate block or field with crop 60m from tree windbreak
SLM001	Vineyard	4/8/13	12/8/20	135	45	MAYBE	same crop with different row orientation. Coordinates innacurate and tower further N. 2022 is within the date range and when the neighboring field became different
SLM002	Vineyard	4/8/13	12/8/20	135	45	MAYBE	seems further N. 2022 is within the date range and when the neighboring field became different
US-Bi1	Alfalfa	1/27/16	4/30/24			NO	
US-Bi2	Corn	1/5/17	4/30/24	135	45	MAYBE	
US-Blo	Ponderosa pine plantation, mixed-evergreen forest	10/7/03	10/2/07	30	330	MAYBE	Clearing less than 100 m to the north.
US-CZ2	Oak/pine forest	1/28/10	9/25/16			NO	
US-CZ3	Pine/fir forest	8/2/08	9/25/16			NO	
US-Dmg	Restored tidal freshwater wetland	1/16/21	4/30/24	180	0	YES	Need to filter winds from E
US-DS1	Corn	2/1/18	12/30/19			MAYBE	Between 3 fields with potentially different irrigation
US-DS2	Corn	1/6/20	3/13/21			NO	
US-DS3	Rice	1/16/21	8/26/23			NO	
US-EDN	Permanent wetland	2/1/18	6/1/21	180	90	MAYBE	Open water to SE
US-Hsm	Diked marsh turned restored in wetland in Fall 2021	1/8/21	11/6/23	180	0	YES	Need to filter winds from E
US-RGF	Corn-wheat rotation	1/22/23	12/9/23			NO	
US-SCf	Oak/pine forest	1/19/06	12/27/14			NO	
US-SCg	Grassland	1/19/06	9/27/16	90	0	MAYBE	Shubs to NE. Need dynamic footprint
US-SCs	Coastal sage scrub	1/19/06	9/27/16			NO	Grass to N and shrubs to S. Need dynamic footprint
US-SCw	Pinyon/Juniper woodland	1/12/06	6/23/16			NO	
US-Snd	Peatland pasture	3/15/07	9/2/14	135	45	MAYBE	Field to E appears to be different crop
US-Sne	Restored wetland	1/27/16	12/7/20	180	0	YES	Need to filter winds from E
US-Snf	Pasture	2/9/18	7/8/20	315	180	YES	Short fetch.
US-SO2	Chaparral, severe fire in 2003	1/14/04	12/14/06			NO	
US-SO3	Chaparral, severe fire in 2003	1/14/04	12/29/06			NO	
US-SO4	Old-growth chaparral ecosystem	1/14/04	12/29/06			NO	
US-Srr	Brackish tidal marsh	1/5/14	9/26/17	180	0	YES	Need to filter winds from E
US-Tw2	Corn on peat soil	1/16/12	4/16/13			NO	Consider filtering N winds due to wetland in footprint.
US-Tw3	Alfalfa	1/2/13	6/1/18			NO	
US-Twt	rice	1/31/09	4/3/17	135	45	YES	
US-Var	C3 annual grasses	10/7/03	4/7/24			NO	
US-xTE	Red and white fir	2/3/18	8/12/23			NO	

Appendix C: Site level details and statistics

Table C1. Site-level closure correction context for California eddy covariance sites based on paired monthly flux tower data. Mean closure is reported as unclosed ET divided by closed ET, expressed as a percentage. Mean Closed – Unclosed ET is reported in mm month^{-1} and summarizes the average magnitude of the energy-balance closure correction applied at each site.

Site ID	Land Cover Details	n months	Mean closure (%)	Mean Closed - Unclosed (mm/month)
US-Bi1	Alfalfa	92	78.4	18.5
US-Tw3	Alfalfa	60	85.9	12.8
Almond_High	Almond	35	79.8	21.7
Almond_Low	Almond	30	87.3	23.4
Almond_Med	Almond	34	77.3	26.1
US-Srr	Brackish tidal marsh	42	78.8	17.9
MB_BRC_S2	Broccoli	1	81.9	19.6
US-Var	C3 annual grasses	257	86.8	4.7
US-SO2	Chaparral, severe fire in 2003	45	97.8	0.1
US-SO3	Chaparral, severe fire in 2003	52	110.2	0.4
US-SCs	Coastal sage scrub	69	81.0	6.3
US-Bi2	Corn	84	72.8	21.4
US-DS1	Corn	17	94.4	2.7
US-DS2	Corn	9	96.8	-0.4
US-Tw2	Corn on peat soil	3	76.4	33.6
US-RGF	Corn-wheat rotation	4	115.7	-13.1
US-Hsm	Diked marsh turned restored in wetland in Fall 2021	29	92.1	3.9
US-SCg	Grassland	40	95.4	1.3
MandarinCitrus	Mandarin orange	22	82.5	17.2
NavalCitrus	Navel orange	20	87.9	7.9
US-CZ2	Oak/pine forest	26	83.9	10.0
US-SCf	Oak/pine forest	29	83.2	12.1
US-SO4	Old-growth chaparral ecosystem	4	85.3	7.4
US-Snf	Pasture	23	79.5	14.5
MB_Pch	Peach	42	100.7	-0.7
US-Snd	Peatland pasture	15	88.1	11.7
US-EDN	Permanent wetland	2	79.5	20.4
US-CZ3	Pine/fir forest	20	124.8	-11.0

US-SCw	Pinyon/Juniper woodland	60	89.6	2.0
US-Blo	Ponderosa pine plantation, mixed-evergreen forest	32	96.0	4.9
US-Dmg	Restored tidal freshwater wetland	32	78.9	19.7
US-Sne	Restored wetland	55	80.5	18.6
US-DS3	Rice	4	110.3	-10.2
BAR007	Vineyard	1	86.1	13.8
BAR012	Vineyard	17	81.3	15.2
RIP760	Vineyard	18	82.3	30.7
SLM001	Vineyard	37	92.7	8.6
SLM002	Vineyard	28	89.7	11.3
MB_WG_ZAB	Winegrape	10	82.3	12.1
Scheid	Winegrape	14	84.0	9.2
US-Twt	rice	96	89.8	11.1

Table C2. Site level daily accuracy statistics between modeled and measured ET for California eddy covariance sites using OpenET Collection v2.0 paired daily data. Statistics are reported for the OpenET ensemble value relative to closed flux tower ET. For each site, n is the number of paired daily observations, Slope is the regression slope forced through the origin, and r^2 is Pearson correlation squared. Mean bias error (MBE), mean absolute error (MAE), and root mean square error (RMSE) are reported in mm d^{-1}, with the corresponding normalized values shown in parentheses as percentages of the mean measured closed flux ET for that site. Site metadata and land-cover classifications are provided in Table 2.

Site ID	Land Cover Details	n	Mean Closed ET (mm)	Slope	r^2	MBE	MAE	RMSE
Almond_High	Almond	197	4.5	0.92	0.81	-0.1 (-2.7%)	0.9 (19.7%)	1.1 (25.7%)
Almond_Low	Almond	184	5.0	0.80	0.86	-0.8 (-16.2%)	1.1 (22.1%)	1.5 (29.1%)
Almond_Med	Almond	195	4.9	0.85	0.84	-0.5 (-9.6%)	1.1 (21.9%)	1.3 (27.6%)
BAR007	Vineyard	28	1.6	0.83	0.62	-0.1 (-6.9%)	0.6 (34.6%)	0.8 (46.2%)
BAR012	Vineyard	113	2.5	0.97	0.77	0.0 (1.6%)	0.5 (20.9%)	0.7 (28.1%)
MandarinCitrus	Mandarin orange	77	3.6	1.09	0.78	0.7 (18.6%)	0.9 (23.8%)	1.1 (30.2%)

MB_BRC_S1	Broccoli	11	3.4	1.07	0.50	0.5 (16.2%)	0.8 (25.1%)	1.0 (30.8%)
MB_BRC_S2	Broccoli	15	3.2	1.08	0.07	0.6 (17.6%)	1.3 (39.6%)	1.5 (45.9%)
MB_LET_S1	Lettuce	3	3.3	1.06	0.80	0.5 (15.7%)	1.0 (29.7%)	1.0 (30.3%)
MB_LET_S2	Lettuce	7	3.4	1.00	0.54	0.1 (2.8%)	0.8 (23.3%)	0.9 (26.8%)
MB_Pch	Peach	102	4.3	0.85	0.84	-0.4 (-9.6%)	0.9 (20.5%)	1.2 (27.1%)
MB_WG_ZAB	Winegrape	49	2.5	0.98	0.48	0.1 (2.1%)	0.6 (24.0%)	0.8 (29.9%)
NavalCitrus	Navel orange	145	2.3	1.45	0.73	1.2 (50.1%)	1.2 (53.7%)	1.5 (65.4%)
RIP760	Vineyard	159	4.6	0.93	0.80	-0.1 (-1.2%)	0.8 (16.5%)	1.0 (20.7%)
Scheid	Winegrape	58	1.8	1.17	0.51	0.4 (22.4%)	0.6 (33.6%)	0.8 (46.2%)
SLM001	Vineyard	340	2.9	1.07	0.82	0.5 (16.0%)	0.7 (23.9%)	0.9 (29.9%)
SLM002	Vineyard	292	3.0	1.16	0.74	0.7 (22.5%)	0.8 (26.7%)	1.0 (34.3%)
US-Bi1	Alfalfa	336	3.5	0.90	0.86	-0.3 (-8.6%)	0.6 (18.0%)	0.8 (24.1%)
US-Bi2	Corn	276	3.1	0.94	0.79	-0.1 (-3.0%)	0.8 (25.6%)	1.0 (33.7%)
US-Blo	Ponderosa pine plantation, mixed-evergreen forest	138	3.1	1.06	0.73	0.5 (16.6%)	0.8 (25.9%)	1.0 (33.1%)
US-CZ2	Oak/pine forest	110	1.9	1.61	0.44	2.0 (106.5%)	2.0 (107.2%)	2.3 (119.2%)
US-CZ3	Pine/fir forest	130	1.6	1.92	0.53	2.1 (129.1%)	2.1 (129.5%)	2.3 (139.9%)
US-Dmg	Restored tidal freshwater wetland	87	4.6	0.89	0.82	-0.3 (-7.1%)	0.8 (18.6%)	1.1 (23.7%)
US-DS1	Corn	57	2.9	1.23	0.88	0.6 (19.7%)	1.0 (33.9%)	1.1 (40.1%)

US-DS2	Corn	41	2.8	1.19	0.89	0.6 (20.0%)	0.9 (32.1%)	1.1 (40.6%)
US-DS3	Rice	46	4.2	0.90	0.86	-0.4 (-9.5%)	0.8 (18.8%)	1.0 (24.5%)
US-EDN	Permanent wetland	44	3.9	0.81	0.72	-0.7 (-17.9%)	0.9 (22.0%)	1.1 (27.7%)
US-Hsm	Diked marsh turned restored in wetland in Fall 2021	82	3.1	1.16	0.73	1.0 (31.7%)	1.2 (38.4%)	1.5 (47.8%)
US-RGF	Corn-wheat rotation	42	2.9	1.16	0.88	0.5 (16.9%)	0.7 (23.6%)	0.9 (31.1%)
US-SCf	Oak/pine forest	163	1.7	1.80	0.48	1.8 (107.1%)	1.9 (109.5%)	2.2 (125.4%)
US-SCg	Grassland	331	1.0	1.09	0.10	0.6 (56.0%)	1.0 (98.6%)	1.3 (128.2%)
US-SCs	Coastal sage scrub	383	1.1	1.28	0.07	1.0 (87.0%)	1.4 (120.3%)	1.7 (147.3%)
US-SCw	Pinyon/Juniper woodland	387	0.6	0.83	0.40	0.0 (6.8%)	0.4 (56.6%)	0.5 (74.6%)
US-Snd	Peatland pasture	113	3.1	1.39	0.41	1.4 (44.7%)	1.4 (47.4%)	1.7 (54.3%)
US-Sne	Restored wetland	142	4.6	1.02	0.82	0.2 (5.0%)	0.8 (16.8%)	1.0 (21.5%)
US-Snf	Pasture	60	3.1	0.87	0.72	-0.4 (-12.1%)	0.7 (21.5%)	0.9 (27.9%)
US-SO2	Chaparral, severe fire in 2003	86	1.2	0.56	0.32	-0.5 (-43.2%)	0.7 (55.5%)	0.8 (71.5%)

US-SO3	Chaparral, severe fire in 2003	67	1.0	0.81	0.53	-0.2 (-16.4%)	0.4 (42.1%)	0.6 (58.7%)
US-SO4	Old-growth chaparral ecosystem	64	1.2	1.45	0.50	1.0 (86.5%)	1.1 (91.4%)	1.2 (104.0%)
US-Srr	Brackish tidal marsh	110	3.9	0.98	0.67	-0.1 (-1.6%)	0.8 (21.4%)	1.0 (25.9%)
US-Tw2	Corn on peat soil	12	3.9	0.86	0.63	-0.4 (-10.2%)	0.8 (20.1%)	1.1 (28.9%)
US-Tw3	Alfalfa	161	3.3	0.92	0.85	-0.3 (-7.5%)	0.6 (18.7%)	0.8 (23.7%)
US-Twt	rice	232	4.0	0.95	0.89	-0.0 (-0.7%)	0.7 (17.7%)	0.9 (23.1%)
US-Var	C3 annual grasses	1155	1.0	1.03	0.55	0.5 (51.1%)	0.8 (77.3%)	1.0 (100.1%)
US-xTE	Red and white fir	14	2.4	1.29	0.46	1.2 (48.1%)	1.3 (53.5%)	1.6 (65.6%)

Table C3. Site-level monthly accuracy statistics between modeled and measured ET for California eddy covariance sites using OpenET Collection v2.0 paired monthly data. Statistics are reported for the OpenET ensemble value relative to closed flux tower ET. For each site, n is the number of paired monthly observations, Slope is the regression slope forced through the origin, and r^2 is Pearson correlation squared. Mean bias error (MBE), mean absolute error (MAE), and root mean square error (RMSE) are reported in mm month^{-1}, with the corresponding normalized values shown in parentheses as percentages of the mean measured closed flux ET for that site. Site metadata and land-cover classifications are provided in Table 2.

Site ID	Land Cover Details	n	Mean Closed ET (mm)	Slope	r^2	MBE	MAE	RMSE
Almond_High	Almond	35	120.9	0.93	0.91	-5.4 (-4.4%)	18.7 (15.4%)	24.2 (20.0%)

Almond_Low	Almond	30	147.0	0.81	0.95	-25.6 (-17.4%)	27.8 (18.9%)	35.5 (24.1%)
Almond_Med	Almond	34	130.6	0.87	0.91	-11.7 (-8.9%)	21.8 (16.7%)	28.8 (22.0%)
BAR007	Vineyard	1	99.3	0.71		-29.2 (-29.4%)	29.2 (29.4%)	29.2 (29.4%)
BAR012	Vineyard	17	79.0	0.97	0.92	-0.6 (-0.8%)	9.5 (12.0%)	11.0 (13.9%)
MandarinCitrus	Mandarin orange	22	102.6	1.08	0.88	16.2 (15.8%)	19.1 (18.6%)	25.8 (25.2%)
MB_BRC_S2	Broccoli	1	108.4	1.13		14.1 (13.0%)	14.1 (13.0%)	14.1 (13.0%)
MB_Pch	Peach	42	114.2	0.88	0.91	-9.6 (-8.4%)	20.0 (17.5%)	26.8 (23.5%)
MB_WG_ZAB	Winegrape	9	77.2	1.09	0.78	7.1 (9.2%)	8.8 (11.4%)	13.3 (17.2%)
NavalCitrus	Navel orange	20	67.3	1.42	0.92	27.9 (41.4%)	28.2 (41.9%)	35.6 (52.9%)
RIP760	Vineyard	18	161.9	0.93	0.67	-4.7 (-2.9%)	22.9 (14.1%)	28.8 (17.8%)
Scheid	Winegrape	14	56.9	1.17	0.83	9.5 (16.7%)	11.8 (20.7%)	15.8 (27.8%)
SLM001	Vineyard	37	99.8	1.09	0.86	16.0 (16.0%)	19.7 (19.7%)	23.6 (23.6%)
SLM002	Vineyard	28	108.7	1.19	0.71	24.7 (22.7%)	25.0 (23.0%)	30.3 (27.9%)
US-Bi1	Alfalfa	92	88.2	0.92	0.92	-9.2 (-10.4%)	13.7 (15.6%)	17.3 (19.6%)
US-Bi2	Corn	84	80.8	0.93	0.85	-5.1 (-6.3%)	18.0 (22.2%)	23.0 (28.4%)
US-Blo	Ponderosa pine plantation, mixed-evergreen forest	17	94.3	1.13	0.78	19.5 (20.7%)	22.5 (23.9%)	28.4 (30.1%)
US-CZ2	Oak/pine forest	26	57.3	1.66	0.58	64.5 (112.6%)	64.5 (112.6%)	69.1 (120.7%)

US-CZ3	Pine/fir forest	21	51.8	2.25	0.76	76.7 (147.9%)	76.7 (147.9%)	77.4 (149.4%)
US-Dmg	Restored tidal freshwater wetland	32	111.8	0.90	0.93	-8.3 (-7.4%)	16.6 (14.9%)	22.1 (19.8%)
US-DS1	Corn	17	67.7	1.19	0.92	5.8 (8.5%)	24.0 (35.5%)	27.6 (40.8%)
US-DS2	Corn	9	59.5	1.27	0.92	16.8 (28.3%)	20.0 (33.6%)	26.5 (44.6%)
US-DS3	Rice	4	99.4	0.88	0.89	-13.4 (-13.4%)	13.6 (13.7%)	16.3 (16.4%)
US-EDN	Permanent wetland	2	117.3	0.83	1.00	-22.9 (-19.5%)	22.9 (19.5%)	24.4 (20.8%)
US-Hsm	Diked marsh turned restored in wetland in Fall 2021	29	90.5	1.07	0.85	15.3 (16.9%)	19.7 (21.7%)	28.0 (31.0%)
US-RGF	Corn-wheat rotation	4	81.9	1.07	0.98	6.8 (8.3%)	6.8 (8.3%)	7.7 (9.5%)
US-SCf	Oak/pine forest	29	53.8	1.95	0.57	63.8 (118.6%)	63.8 (118.6%)	68.5 (127.4%)
US-SCg	Grassland	40	24.4	1.45	0.28	17.1 (69.9%)	20.4 (83.3%)	29.9 (122.4%)
US-SCs	Coastal sage scrub	68	29.5	1.77	0.14	34.9 (118.3%)	38.8 (131.6%)	50.2 (170.4%)
US-SCw	Pinyon/Juniper woodland	60	17.3	0.99	0.32	3.3 (19.3%)	9.1 (52.4%)	11.5 (66.3%)
US-Snd	Peatland pasture	15	98.5	1.47	0.66	50.1 (50.9%)	50.1 (50.9%)	51.5 (52.3%)

US-Sne	Restored wetland	55	114.0	1.05	0.94	5.7 (5.0%)	12.9 (11.4%)	18.2 (16.0%)
US-Snf	Pasture	23	77.6	0.88	0.94	-11.6 (-14.9%)	13.0 (16.8%)	15.1 (19.5%)
US-SO2	Chaparral, severe fire in 2003	8	41.1	0.63	0.47	-14.0 (-34.2%)	18.8 (45.8%)	20.7 (50.4%)
US-SO3	Chaparral, severe fire in 2003	3	18.3	0.77	0.96	-5.7 (-30.9%)	5.7 (30.9%)	6.4 (34.7%)
US-SO4	Old-growth chaparral ecosystem	4	50.7	1.65	0.12	37.3 (73.5%)	37.3 (73.5%)	39.5 (77.7%)
US-Srr	Brackish tidal marsh	42	99.9	1.00	0.86	-5.3 (-5.3%)	20.9 (20.9%)	24.7 (24.7%)
US-Tw2	Corn on peat soil	3	130.7	0.89	0.87	-11.9 (-9.1%)	19.0 (14.6%)	21.7 (16.6%)
US-Tw3	Alfalfa	60	86.0	0.93	0.93	-8.4 (-9.8%)	12.8 (14.9%)	16.5 (19.2%)
US-Twt	rice	95	98.6	0.97	0.95	0.6 (0.6%)	12.0 (12.1%)	16.3 (16.5%)
US-Var	C3 annual grasses	219	31.0	1.00	0.47	10.6 (34.3%)	20.6 (66.7%)	26.5 (85.7%)