Thomas J Ott[1], Sayantan Majumdar[1], Justin Huntington[2], Christopher Pearson[2], Matt Bromley[1], Blake A Minor[1], Charles Morton[2], Sachiko Sueki[1], Jordan P Beamer[3], and Richard Jasoni[1]

[1]Desert Research Institute
[2]Affiliation not available
[3]Oregon Water Resources Department

February 15, 2024

Abstract

Groundwater overdraft in western U.S. states has prompted water managers to start the development of groundwater management plans that include mandatory reporting of groundwater pumping (GP) to track water use. Most irrigation systems in the western U.S. are not equipped with irrigation water flow meters to record GP. Of those that do, performing quality assurance and quality control (QAQC) of the metered GP data is difficult due to the lack of reliable secondary GP estimates. We hypothesize that satellite (Landsat)-based actual evapotranspiration (ET) estimates from OpenET can be used to predict GP and aid in QAQC of the metered GP data. For this purpose, the objectives of this study are: 1) to pair OpenET estimates of consumptive use (Net ET, i.e., actual ET less effective precipitation) and metered annual GP data from Diamond Valley (DV), Nevada, and Harney Basin (HB), Oregon; 2) to evaluate linear regression and ensemble machine learning (ML) models (e.g., Random Forests) to establish the GP vs Net ET relationship; and 3) to compare GP estimates at the field- and basin-scales. Results from using a bootstrapping technique showed that the mean absolute errors (MAEs) for field-scale GP depth are 12% and 11% for DV and HB, respectively, and the corresponding root mean square errors (RMSEs) are 15% and 14%. Moreover, the regression models explained 50%-60% variance in GP depth and ~90% variance in GP volumes. The estimated average irrigation efficiency of 88% (92% and 83% for DV and HB, respectively) aligns with known center pivot system efficiencies. Additionally, OpenET proves to be useful for identifying discrepancies in the metered GP data, which are subsequently removed prior to model fitting. Results from this study illustrate the usefulness of satellite-based ET estimates for estimating GP, QAQC metered GP data and have the potential to help estimate historical GP.

Submitted to Agricultural Water Management.

Toward Sustainable Groundwater Management: Harnessing Remote Sensing and Climate Data to Estimate Field-Scale Groundwater Pumping

Thomas J. Ott [a*], Sayantan Majumdar [a*†], Justin L. Huntington [a†], Christopher Pearson [a], Matt Bromley [a], Blake A. Minor [a], Charles G. Morton [a], Sachiko Sueki [b], Jordan P. Beamer [c], Richard Jasoni [a]

[a] Desert Research Institute, Reno, NV, USA

[b] Desert Research Institute, Las Vegas, NV, USA

[c] Oregon Water Resources Department, Salem, OR, USA

Abstract

Groundwater overdraft in western U.S. states has prompted water managers to start the development of groundwater management plans that include mandatory reporting of groundwater pumping (GP) to track water use. Most irrigation systems in the western U.S. are not equipped with irrigation water flow meters to record GP. Of those that do, performing quality assurance and quality control (QAQC) of the metered GP data is difficult due to the lack of reliable secondary GP estimates. We hypothesize that satellite (Landsat)-based actual evapotranspiration (ET) estimates from OpenET can be used to predict GP and aid in QAQC of the metered GP data. For this purpose, the objectives of this study are: 1) to pair OpenET estimates of consumptive use (Net ET, i.e., actual ET less effective precipitation) and metered annual GP data from Diamond Valley (DV), Nevada, and Harney Basin (HB), Oregon; 2) to evaluate linear regression and ensemble machine learning (ML) models (e.g., Random Forests) to establish the GP vs Net ET relationship; and 3) to compare GP estimates at the field- and basin-scales. Results from using a bootstrapping technique showed that the

[*] These two authors have contributed equally.
[†] Corresponding author at: Division of Hydrologic Sciences, Desert Research Institute, 2215 Raggio Parkway, Reno, Nevada 89512-1095, USA
Email addresses: sayantan.majumdar@dri.edu (S. Majumdar), justin.huntington@dri.edu (J.L. Huntington)

mean absolute errors (MAEs) for field-scale GP depth are 12% and 11% for DV and HB, respectively, and the corresponding root mean square errors (RMSEs) are 15% and 14%. Moreover, the regression models explained 50%-60% variance in GP depth and ~90% variance in GP volumes. The estimated average irrigation efficiency of 88% (92% and 83% for DV and HB, respectively) aligns with known center pivot system efficiencies. Additionally, OpenET proves to be useful for identifying discrepancies in the metered GP data, which are subsequently removed prior to model fitting. Results from this study illustrate the usefulness of satellite-based ET estimates for estimating GP, QAQC metered GP data and have the potential to help estimate historical GP.

Keywords: groundwater pumping; remote sensing; evapotranspiration; irrigation; machine learning; consumptive use

1. INTRODUCTION

1.1 Background

In the western United States (U.S.), the combination of the already occurring and projected droughts (Meza et al., 2020), rising irrigation water demands, and population growth is expected to intensify groundwater consumption (Huntington et al., 2015; Ketchum et al., 2024). This intensified groundwater consumption has led to aquifer depletion (ADWR, 2018; Scanlon et al., 2012; Smith et al., 2017, 2023), land subsidence (Hasan et al., 2023; Herrera-García et al., 2021; Smith & Li, 2021; Smith & Majumdar, 2020), water contamination (Levy et al., 2021; Smith et al., 2018), and streamflow depletion (Ketchum et al., 2023; Zipper et al., 2022). Despite these pressing challenges, many groundwater basins in this region lack comprehensive monitoring of groundwater pumping (GP). Accurately assessing GP is imperative for implementing sustainable strategies to confront water security challenges.

Consequently, the development of reliable and efficient solutions for GP monitoring holds paramount importance in effectively addressing water management issues in this region.

New water management policies across the western U.S. states have begun to include mandatory reporting of GP. These new policies are being sparked by groundwater overdrafts in regions heavily dependent on groundwater. Understanding how much water is being withdrawn from aquifers allows water managers to manage groundwater resources more effectively. Most GP in U.S. western states is used for irrigated agriculture, e.g., in Nevada, California, and Oregon, about 70% to 90% of groundwater is used for irrigation (Dieter et al., 2018). Of the 256 designated hydrologic basins in Nevada, 96 are considered over-appropriated, and in some cases, by more than 300% (NDWR, 2021). Many of these over-appropriated basins are also pumping groundwater at rates above their perennial yield, causing groundwater levels to decline. Other western U.S. states are also experiencing over-appropriation and over-drafting in many of their hydrologic basins (Reilly et al., 2008; Zektser et al., 2005). In response, these regions are actively creating new groundwater management policies and laws to monitor GP further (Megdal et al., 2015). Examples of these new policies include the Sustainable Groundwater Management Act in California (SGMA), which seeks to balance basin water budgets (Owen et al., 2019), the Local Enhanced Management Areas (LEMA), which aims to develop enforceable and monitored water use reduction in Kansas (Butler et al., 2018), Active Management Areas (AMAs) and Irrigation Non-Expansion Areas (INAs) in Arizona (ADWR, 2023), and Critical Management Area (CMA) orders and other designations which strive to meter and report all non-domestic GP in dozens of groundwater basins in Nevada (http://water.nv.gov/StateEnginersOrdersList.aspx). As the business saying goes, “You can’t manage what you don’t measure.” This axiom is equally true for water resources. Recognizing this fact, developing programs for monitoring and reporting water use to state and local agencies is becoming a common trend with respect to water policy and regulation in the western U.S. (Deines et al., 2019; Megdal et al., 2015).

While metering all GP sounds like a simple solution for monitoring and reporting groundwater use, installing meters at all well heads or diversions is a costly process. Additionally, meter readings do not equate to the consumptive use of groundwater, which is the quantity ultimately needed for groundwater management. Perhaps more importantly, GP meter data have high uncertainty and are often erroneous (Fanning et al., 2001). Primary sources of uncertainty and error are due to the following variables: a large variation in the quality of meter type, poor meter installation, lack of meter calibration, unnoticed meter drift, meter failure or partial failure, erroneous recording of meter data, meter data input errors such as those commonly associated with self-reporting (Carroll et al., 2010; Little et al., 2016; Sheppard & Terveen, 2011), and lack of quality assurance and quality control (QAQC) procedures and guidelines. In addition, regarding the self-reporting aspect, the validity of the data could be compromised due to the potential of water users acting in bad faith. With large amounts of data being collected and reported either through online self-reporting systems or by state and local agency field surveys, coupled with high uncertainty and potential for errors, questions and concerns around the quality and validity of meter data will likely be a source of conflict for groundwater management in the future. Given these factors, it is important to have complementary, independent, and cost-effective approaches to data collection, which allow for direct GP estimates to be obtained, as well as the ability to assess GP when no records exist.

1.2 Previous Work

Numerous approaches have been developed for estimating GP and consumptive use. Here, we provide a brief overview of some of the more common as well as recently developed approaches.

In the past, electrical power records were one of the most common approaches for estimating GP and have been used by many studies (Burt et al., 1997; Frenzel, 1985; Said et al., 2005). This method requires information on pump efficiency, water lift height, and operating pressures (Frenzel, 1985). Error in these factors and their change through time leads to errors in GP estimates (Hurr & Litke, 1989). Obtaining power records for well pumps is difficult, especially for rural communities. The use of power records for assessing and estimating GP for the purpose of groundwater management is not feasible at large scales.

Current methods for estimating groundwater withdrawals commonly include surveying and organizing county-scale annual water use (Dieter et al., 2018; Martin et al., 2023), process-based models (ADWR, 2018; Ahamed et al., 2022; Brookfield et al., 2023; Dogrul et al., 2016; Faunt, 2009; Ruess et al., 2023) and the recent integrated remote sensing and machine learning-driven approaches (Majumdar et al., 2020, 2021, 2022, 2024). County-level estimates offer a comprehensive overview of water usage on a nationwide scale within the conterminous U.S. (CONUS), yet finer spatial or temporal details are lacking (Dieter et al., 2018; Martin et al., 2023).

Process-based models have shown success in specific regions; however, they often cannot effectively utilize the numerous remote sensing datasets accessible, such as field-scale evapotranspiration data— an obvious indicator of GP in areas with little to no surface water for irrigation (Bos et al., 2009; Melton et al., 2021). Also, process-based models are computationally intensive and require strict model calibration procedures. The Central Valley Hydrologic Model (CVHM) is one such example of a well-established hydrologic model developed by Faunt (2009), which adopts the MODFLOW FMP package (Schmid, 2004) for simulating water requirements. CVHM integrates land use, surface water supply, and water demand information using an ET model that factors in temperature, crop type, precipitation, and root depth. Following this, CVHM allocates the remaining water demand to GP.

Another method involves building lookup tables based on land use derived from remote sensing, modeled precipitation, and in-situ pumping data (Wilson, 2021). However, this approach overlooks the intricate interplays between climate, evaporative demand, land use, and soil composition. Machine learning-based solutions have the capacity to integrate a diverse range of datasets, including remote sensing data and model-generated datasets. Furthermore, these solutions can handle complex relationships among input datasets and have been proven to provide reliable estimates (Filippelli et al., 2022; Lamb et al., 2021; Majumdar et al., 2020, 2021, 2022, 2024; Wei et al., 2022).

For data-driven or machine learning-based methodologies, large amounts of quality data are required (Majumdar et al., 2022). Employing machine learning to estimate GP requires in-situ pumping measurements on expansive spatial and temporal scales (2002-2020), facilitating the validation of withdrawal quantities across extensive regions. As a result, generating and validating gridded annual GP estimates in Kansas (Majumdar et al., 2020) and Arizona (Majumdar et al., 2022) were feasible at 5 km and 2 km spatial resolutions, respectively.

However, GP from most aquifers is not measured; instead, only a small proportion of wells are metered (Foster et al., 2020). Additionally, in these areas, groundwater use is monitored to such a limited extent that validation of previously reported water use is often absent. In cases where validation is undertaken, skill metrics are comparatively lower than those of regional groundwater models (Wilson, 2021). Consequently, generating and validating gridded prediction rasters in regions with sparse in-situ GP measurements poses a challenge, underscoring the need for a thorough evaluation of model efficacy using cross-validation techniques (Hastie et al., 2001). Furthermore, in regions characterized by sparse datasets—a common scenario in the groundwater domain—cross-validating the total GP for each individual pixel, as carried out in data-rich regions like Kansas and Arizona, is not practical.

Instead, the cross-validation process must be conducted at the scale of individual fields with existing field boundaries, predictor attributes, and meter data (Majumdar et al., 2024).

In addition to these process-based and data-driven methods, there are deterministic approaches to estimating GP, which incorporate consumptive use (water transpired by the crop plus water evaporated from the soil surface) and net irrigation water requirement (NIWR) (water delivered to a system to meet irrigation requirement) (Allen & Robison, 2007; Bos et al., 2009; Huntington & Allen, 2009). These methods are based on reference ET and the single or dual crop coefficient approach (Allen et al., 1998).

Huntington and Allen (2009) used the dual crop coefficient method to estimate consumptive use and NIWR across basins in Nevada. The Nevada Department of Water Resources (NDWR) uses these numbers to estimate GP where meter data is absent (NDWR, 2022). The GP estimates are calculated by dividing NIWR by the irrigation efficiency factor (0.85 for pivot, 0.75 for wheel lines, and 0.60 for flood irrigation, Howell (2002)) and multiplying by the crop acreage. This approach has been used in many groundwater modeling and water budget studies throughout the western U.S. (Carroll et al., 2010; Huntington et al., 2022; Mefford & Prairie, 2022; NDWR, 1985; OWRD, 2015), and assumes the crop is well-watered, stress-free, and uniformly irrigated, which rarely occurs in all irrigated fields across a basin. Crop conditions and water use are highly variable in time and space due to the following factors: water availability, fallowing, partial irrigation, variation in soil type, crop stress, disease, and diverse farming practices. Using NIWR or similar potential ET-based approaches does not account for spatial variability in crop conditions but serves as an upper bound for estimating water use. Remote sensing of actual ET addresses many of these shortcomings through field-scale observations of actual field conditions.

1.3 Research Goals and Objectives

While the studies above are the first for estimating gridded (regional- or basin-scale) GP using remote sensing and data-driven or process-based approaches, they are not suitable for field-scale applications over large areas and periods. Moreover, field-scale GP volumes reported by Filippelli et al. (2022) in the Republican River Basin, Colorado aquifer impose artificial correlations between irrigated fields and GP volumes (i.e., a larger field will have higher GP than a smaller one). Filippelli et al. (2022) integrated remote sensing and machine learning techniques to estimate field-scale GP and was conducted in a data-rich setting (e.g., western Kansas has more than 90% metering, Foster et al. (2020)) like Majumdar et al. (2020, 2021, 2022). Hence, cross-validating the results and testing the spatial and temporal model generalizability with leave-one-area-out and leave-one-year-out strategies (these are based on leave-one-out cross-validation, Hastie et al. (2001); Pedregosa et al. (2011)) are not practical for data-scarce regions in the western U.S, where GP metering has recently begun (e.g., 2018 in Nevada and 2016 in Oregon).

At the time of this manuscript preparation, no study has been conducted comparing field-scale satellite-based ET estimates to GP depths over many fields and for multiple years and concurrently providing insights into irrigation efficiencies (Howell, 2002). The goal of this study was to conduct such a comparison by developing a regression model between the consumptive use or Net ET (actual ET less effective precipitation) and GP depth, which can be used to support QAQC of GP records and provide a means to estimate GP where meter data is unavailable. This regression will be derived using GP data from Diamond Valley (DV), Nevada, and Harney Basin (HB), Oregon, locations where good-quality GP data is available. Furthermore, we assess whether machine learning can improve the estimates obtained using linear regression and discuss the importance of developing carefully attributed irrigation data (digitized field boundaries and irrigation water source).

We hypothesize that 1) field-scale satellite-based ET estimates will be well-correlated with field-scale metered GP data, and 2) statistical relationships between field-scale satellite-based ET and GP data will be useful for QAQC of GP records and assessment of prior estimates, and 3) ET-based predictions of GP will compare reasonably well to metered GP at the field and basin scales. To test these hypotheses, this study: 1) employs the OpenET ensemble product to obtain field-scale actual ET, 2) links GP and ET values by delineating water rights place of use (POU) field boundaries, pairing GP with the POU and irrigated field boundaries, and pairing modeled ET with field boundaries 3) compares Net ET estimates with metered GP to identify potential outliers, 4) develops a regression model (GP as a function of Net ET) that assesses uncertainty using bootstrapping, and calculates the confidence and prediction intervals for the model, 5) compares predicted GP from the regression model with basin totals reported by NDWR and OWRD, 6) evaluates multiple machine learning model performance, and 7) compares the satellite-based OpenET ensemble mean with the individual ensemble members.

2. STUDY AREAS AND DATASETS

2.1 Study Areas

In this research, we focus on two study areas: Diamond Valley, Nevada, and Harney Basin, Oregon. DV is located in Central Nevada and is one of the only fully metered groundwater-dependent basins in Nevada and possibly the western U.S. In 2015, the Nevada State Engineer's Office designated the basin as a Critical Management Area, which

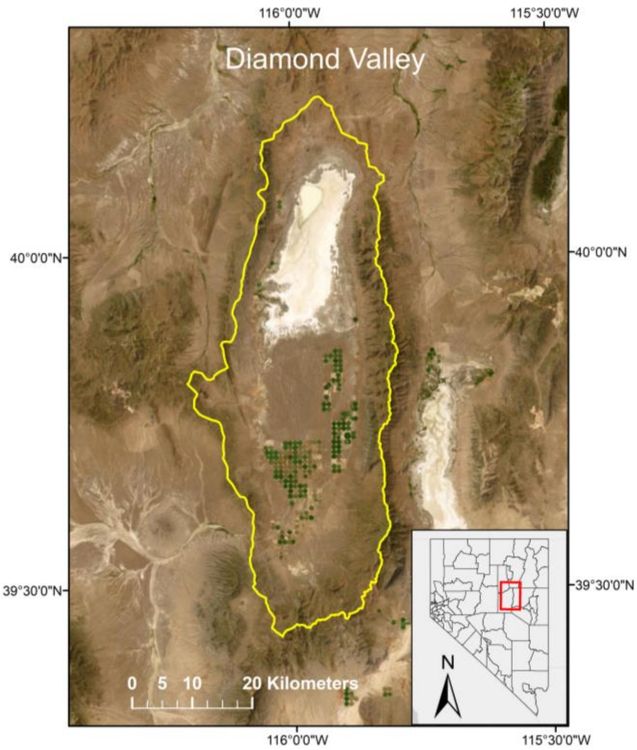


Figure 1. Satellite image of Diamond Valley (DV), Nevada showing the irrigated fields.

initiated the formation of a groundwater management plan (GMP). The plan involved all participating growers installing flow meters selected from the "Idaho Department of Water Resources List of Approved Closed Conduit Flow Meters" and reporting GP to the State of Nevada (Bugenig, 2017). The goal of the plan was to reduce GP by 55% within the next 35 years. The GMP was implemented in 2018 and continued into 2020 when the GMP was challenged and stuck down by Nevada District Courts. Despite the litigation, groundwater pumping was still reported by growers for 2020-2022. In mid 2022, the Nevada Supreme court upheld the GMP (Long, 2022). DV contains 10,500 hectares (26,000 acres) of irrigated agricultural land, and the estimated GP is 94 Mm^3/year (76,000 acre-ft/year) (Bugenig, 2017). The estimated basin perennial yield is 37 Mm^3/year (30,000 acre-ft/year) (Harrill 1968), causing groundwater levels to decline, with some areas experiencing nearly 25 m (80 ft) of decline over the last fifty years (Berger et al., 2016). Though Berger et al. (2016) estimated perennial yield to be 43 Mm^3/year (35,000 acre-ft/year) the Diamond Valley GMP used the perennial yield established by Harrill (1968) (Bugenig, 2017). The 30-year average precipitation in the valley is 230 mm/year, with approximately 60% occurring in the winter months. Warmest temperatures occur in July with an average high of 31˚C and lowest temperatures in December with an average low of -12˚C. The main agricultural crop is alfalfa or other grass hays primarily irrigated with center pivot systems.

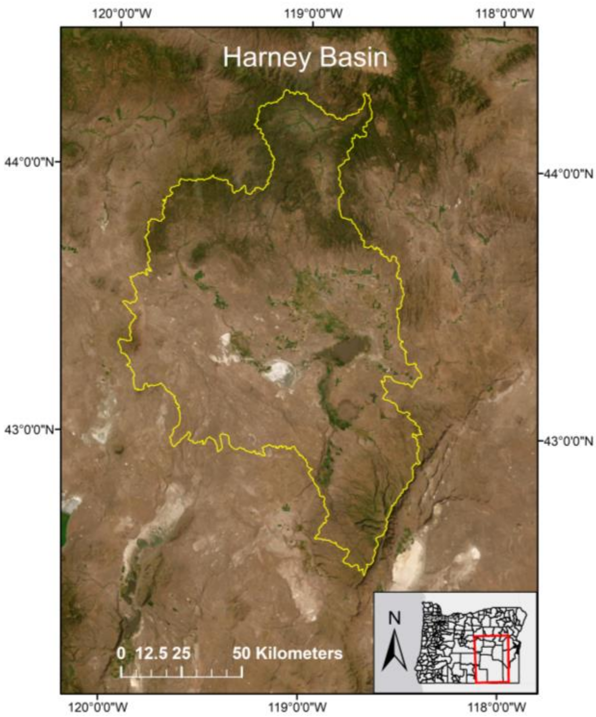


Figure 2. Satellite image of Harney Basin (HB), Oregon showing the irrigated fields.

Harney Basin is located in south-eastern Oregon, where irrigation is the primary user of groundwater, accounting for 95% of all groundwater use (Beamer & Hoskinson, 2021; Gingerich, Garcia, et al., 2022; Gingerich, Johnson, et al., 2022). The basin is semiarid in climate and receives an average of 230 mm to 300 mm of precipitation per year, with most occurring (80%) in the winter months (Beamer & Hoskinson, 2021).

Declining groundwater levels in recent years, likely caused by over-appropriation, have sparked concerns over the sustainability of the resource in HB. Currently, 38,777 hectares (95,821 acres) of permitted primary or supplementary groundwater rights exist in the Greater Harney Valley Area (GHVA) (Beamer & Hoskinson, 2021). The water rights for these permits exceed the estimated recharge for the basin, which is poorly defined. In 2016, the Oregon Water Resources Department (OWRD) co-developed a groundwater study plan with the U.S. Geological Survey (USGS) to facilitate an improved understanding of the groundwater resources and flow systems in HB (Garcia et al., 2021; Gingerich, Johnson, et al., 2022).

The primary crops irrigated in the HB region are alfalfa and grass hay, with May to September being the typical growing season. Additionally, limited quantities of spring and winter grains and mint are also produced in HB (Beamer & Hoskinson, 2021). The irrigation GP has nearly tripled during 1991–2018, increasing from about 62 Mm^3/year to 185 Mm^3/year, i.e., 51,000 acre-ft/year to 150,000 acre-ft/year (Gingerich, Garcia, et al., 2022). Therefore, it is essential to develop efficient and reliable field-scale GP estimation methods to support the GMPs in both DV, Nevada, and HB, Oregon.

2.2 Datasets

The key datasets in our study include Landsat actual evapotranspiration (ET) from OpenET (Melton et al., 2021; Volk et al., 2024), precipitation, and reference ET (ET_o) from gridMET

(Abatzoglou, 2013) and the irrigation data comprised of digitized field boundaries and irrigation water source (Huntington et al., 2018; Beamer & Hoskinson, 2021).

2.2.1 OpenET

OpenET provides actual ET measurements using data derived from various satellite-driven ET models while also computing a unified "ensemble value" derived from these models. The OpenET ensemble incorporates models that have been utilized by governmental bodies responsible for water use monitoring and management across the Western U.S. Some of these models are also widely adopted on an international scale. These models uniformly utilize Landsat satellite data to generate ET information at a spatial resolution of 30 m. Additional input factors include gridded meteorological variables such as solar radiation, air temperature, humidity, wind speed, and, in certain instances, precipitation data (Melton et al., 2021). Table 1 shows the current ET models used for generating the OpenET ensemble. With the exception of the vegetation index-based SIMS model, OpenET models are developed using either complete or simplified adaptations of the surface energy balance (SEB) methodology.

The SEB approach effectively factors in the energy required to convert liquid water within plants and soil into vapor, which is subsequently released into the atmosphere (Laipelt et al., 2021). For the monthly OpenET ensemble product, Volk et al. (2024) observed a strong correlation (R^2=0.9) with the flux tower ET (152 sites across the CONUS), low mean bias error (5.3 mm/month or 5.8%), and combined metrics, i.e., root mean square error (RMSE) and mean absolute error (MAE) of 20.4 mm/month (20.4%) and 15.9 mm/month (17.3%), respectively. With actual ET measurements at the field scale, the OpenET ensemble of ET data are the most important input datasets in our study.

Table 1. Existing ET models in the OpenET ensemble [reproduced from OpenET (2023)].

Model Acronym	Model Name	References
ALEXI/DisALEXI v 0.0.32	**A**tmosphere-**L**and **Ex**change **I**nverse / **Dis**aggregation of the **A**tmosphere-**L**and **Ex**change **I**nverse	Anderson et al. (2007, 2018)
eeMETRIC v 0.20.26	Google **E**arth **E**ngine implementation of the **M**apping **E**vapo**t**ranspiration at high **R**esolution with **I**nternalized **C**alibration model	Allen et al. (2005, 2007, 2011)
geeSEBAL v 0.2.2	**G**oogle **E**arth **E**ngine implementation of the **SEB A**lgorithm for **L**and	Bastiaanssen et al. (1998); Laipelt et al. (2021)
PT-JPL v 0.2.1	**P**riestley-**T**aylor **J**et **P**ropulsion **L**aboratory	Fisher et al. (2008)
SIMS v 0.1.0	**S**atellite **I**rrigation **M**anagement **S**upport	Melton et al. (2012); Pereira et al. (2020)
SSEBop v 0.2.6	**Op**erational **S**implified **SEB**	Senay (2018); Senay et al. (2013, 2022)

2.2.2 gridMET

The gridMET dataset (Abatzoglou, 2013) offers a comprehensive collection of daily surface measurements, including temperature, precipitation, winds, humidity, and radiation across the CONUS from 1979 at ~4 km spatial resolution. This dataset integrates the openly available ~4 km spatial data from the Parameter-elevation Relationships on Independent Slopes Model (PRISM) (Daly et al., 2008) with the high temporal-resolution data from the National Land

Data Assimilation System (NLDAS) (Xia et al., 2012). The validation metrics over the western U.S. indicate favorable results, with $\pm$5% precipitation bias (Abatzoglou, 2013). Gridded daily grass reference ET (ET_o) also displayed a strong correlation to daily ET_o measurements (median Pearson's correlation coefficient of 0.9), although it displayed a positive bias across most sites (median bias +0.5 mm).

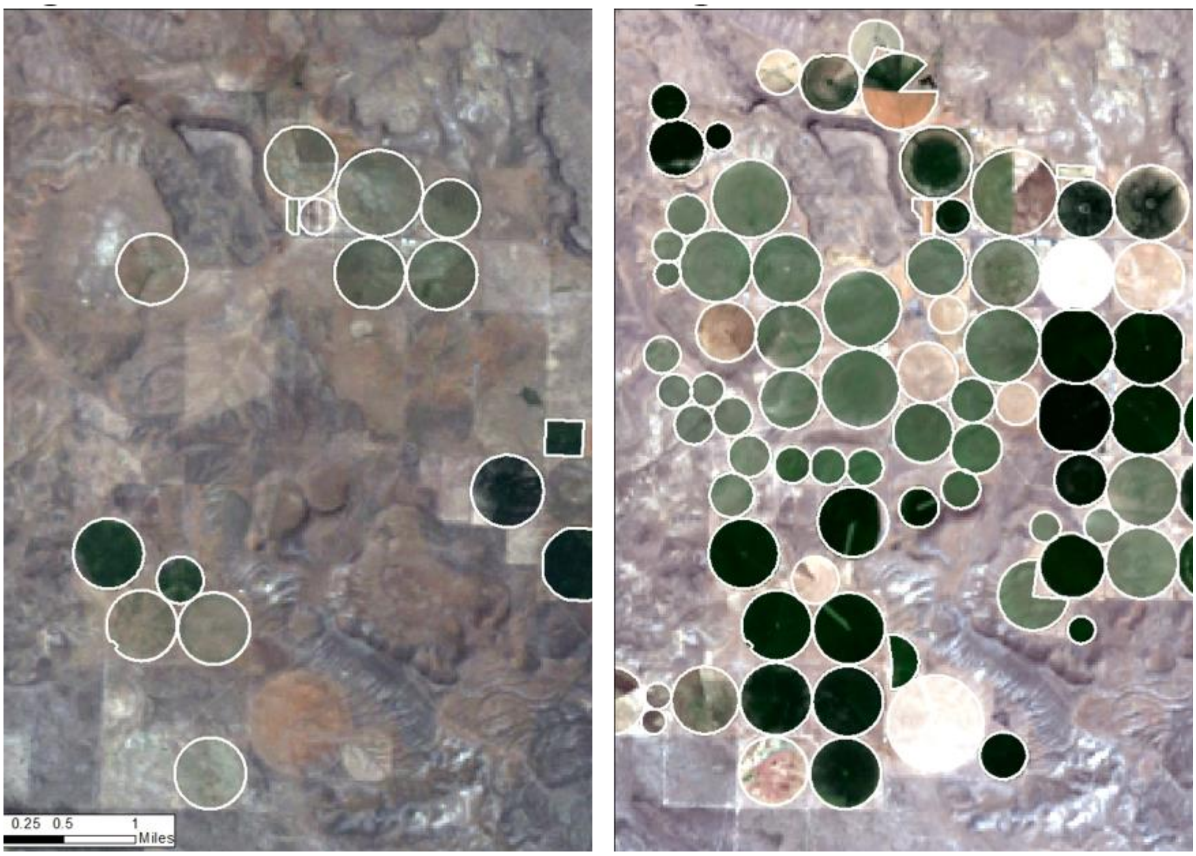
gridMET ET_o was bias corrected within OpenET based on agricultural weather station network datasets to account for mostly positive bias in gridMET ET_o as a result of evaporative cooling and boundary layer conditioning effects that occur within agricultural areas and not accounted for in gridMET and most other gridded meteorological datasets (Melton et al., 2021; Blankenau et al., 2020; Hobbins & Huntington, 2017; Volk et al., 2024). Since the gridMET product is operationally used in OpenET (Melton et al., 2021; Volk et al., 2024), we rely on the precipitation and bias-corrected gridMET ET_o data to calculate the effective precipitation and consumptive use, i.e., Net ET.

2.2.3 Irrigation Data

Irrigation data comprising carefully attributed irrigated field boundaries and water source type are critical for field-scale GP estimation. Here, we describe these datasets and their importance in performing field-scale assessments of GP and ET.

2.2.3.1 Digitized irrigation field boundaries

The field boundaries for each study year were derived from the U.S. Department of Agriculture (USDA) Common Land Unit (CLU) data representing the acreage in 2008 (USDA Farm Service Agency, 2017). These boundaries were manually adjusted using visualizations of high-resolution National Agricultural Imagery Program (NAIP) data (USDA, 2023) and mapped water rights POU boundaries obtained from state water agencies. In years without NAIP data, Landsat false color composites for the specific year were employed, along with NAIP and National Land Cover Database (NLCD) data (https://www.mrlc.gov/) from nearby years, to create a comprehensive annual irrigated field boundary dataset (Beamer & Hoskinson, 2021; Huntington et al., 2018).



(a) (b)

Figure 3. Digitized irrigated field boundaries in the GHVA portion of HB, Oregon (reproduced from Beamer and Hoskinson (2021)). The massive increase in irrigation and changes from square fields to circles between (a) August 1991 and (b) August 2016 showcases the need for maintaining and updating our irrigation field boundary dataset. The field boundary shape changes are due to switching irrigation systems, i.e., from flood or sprinkler-line systems to center pivots.

Changes in field boundaries were primarily observed when fields traditionally irrigated with flood or sprinkler-line systems were transformed into center-pivot irrigation or when new fields were brought into production. Figure 3 illustrates examples of mapped field boundaries in the central GHVA portion of the HB for 1991 and 2016. Each year, the individual polygons representing field boundaries were assigned a unique ID along with the start year of active irrigation, signifying the year when the field was initially identified in the imagery as actively irrigated (Beamer & Hoskinson, 2021). These carefully attributed and digitized irrigation field boundaries are used in OpenET platform as well as the ensemble product to generate field-scale actual ET estimates throughout the western U.S. (Melton et al., 2021).

2.2.3.2 Water source type

Irrigation is entirely groundwater dependent within the DV, Nevada study area (NDWR, 2020), while in the HB, Oregon irrigation is sourced from groundwater, surface water, or a combination of these two for irrigation (Beamer & Hoskinson, 2021; Garcia et al., 2021). Since the focus of our study is to estimate field-scale irrigation GP, having a water source type attribute to the digitized fields is essential to remove fields that are irrigated with surface water and/or a combination of surface water and groundwater.

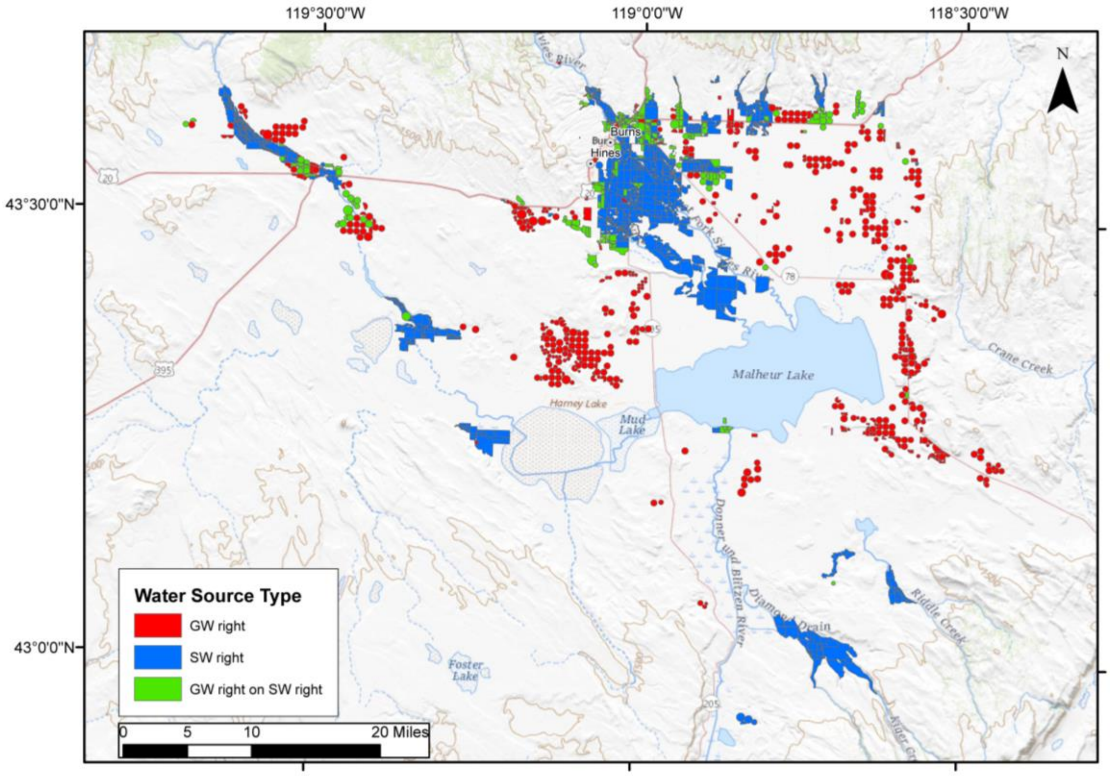


Figure 4. Mapped irrigated field boundaries in the GHVA, HB, Oregon for 2016 with the associated water source type (reproduced from Beamer and Hoskinson (2021)). GW right, SW right, and GW right on SW represent groundwater, surface water, and combination source types, respectively.

The annual field boundaries were associated with specific irrigation source types: groundwater irrigated (GW), surface water irrigated (SW), or a combination of groundwater and surface water (GW&SW). The initial stage in determining the irrigation source type involved overlaying the annual field boundaries with the OWRD-mapped dataset of water rights POU. The irrigated POU dataset for the Harney Basin was categorized into areas with exclusively groundwater rights, exclusively surface water rights, and areas with both surface water and groundwater rights where they intersected. For each year, only the POU polygons with priority dates for that year and all preceding years were incorporated into the analysis to depict irrigation development accurately. The chosen POU polygons were then transformed into a 30 m raster using the USGS 3D Elevation Program (3DEP) 30 m Digital Elevation Models (DEMs) (USGS, 2023). Cells within this raster were categorized into irrigation source types using integer values (1 = GW irrigated, 2 = SW irrigated, 3 = Combination). The resultant POU irrigation raster for the year 2016 is illustrated in Figure 4. Beamer and Hoskinson (2021) provide more details on this approach.

2.2.4 Additional Datasets

In addition to the datasets discussed, we rely on several other remote sensing and climate data that serve as predictors for the machine learning models (Section 4.3). These datasets are intricately related to the hydrologic and hydroclimatic processes driving GP. These include the gridMET (Abatzoglou, 2013) minimum and maximum air temperature, minimum and maximum relative humidity, vapor pressure deficit, grass reference ET, alfalfa reference ET, and wind velocity. Moreover, we use the Daymet v4 precipitation data (~1 km spatial resolution, Thornton et al., 2021), Landsat-8 32-day normalized difference vegetation index (NDVI) composite (30 m, courtesy of the USGS), NASA digital elevation model (NASADEM, 30 m, NASA JPL, 2020), conterminous U.S. (CONUS) 800-m soil properties that include the hydrologic soil group (HSG), soil depth, and the saturated hydraulic conductivity (Walkinshaw et al., 2022), and the OpenET ensemble mean actual ET as well as the individual model actual ET (Melton et al., 2021, Table 1).

3. METHODS

3.1 Matching Point of Diversions, Meter Readings, and Places of Use

One of the major challenges in this project is matching the metered GP data with the field-scale ET data. Groundwater applications in the State of Nevada are required to include a Point of Diversion (POD) (often a well) and POU, i.e., the maximum area over which water from the POD can be applied (https://www.leg.state.nv.us/nrs/nrs-533.html). Multiple applications can be filed for a single well, allowing multiple PODs and POUs for the same area (i.e., stacked water rights).

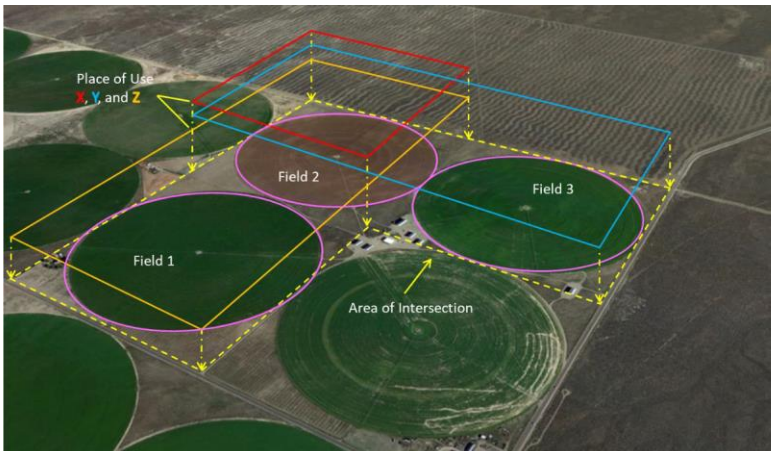


Figure 5. POUs X, Y, and Z paired with the OpenET fields 1, 2, and 3 using spatial intersection in Diamond Valley, Nevada. This is an illustration of a one-to-one mapping with one total pumping value being paired with one OpenET ensemble actual ET estimate. This process is replicated for the Harney Basin, Oregon.

Reported GP values for each well (can be several) are totaled and assigned to the senior most water right by the Nevada State Engineers Office (NSEO). These groups of POUs were joined into a single polygon representing the total area of application. However, since a POU typically extends beyond actual irrigated areas (e.g., quarter section POU with center pivot irrigated area within, see Figure 5), we cannot directly use the POU to estimate satellite-based ET. To better define irrigated areas, we relied on Geographic Information System (GIS) and Python software (Van Rossum & Drake, 2009) to spatially join the POU polygons with the irrigated area database

(Section 2.2.3) developed by the Desert Research Institute as part of the OpenET project (Huntington et al., 2018; Melton et al., 2021). More specifically, we use Geopandas (Jordahl et al., 2020) to spatially join the grouped POU data from the NSEO and the irrigated area polygons from OpenET. Figure 5 illustrates this process where POUs X, Y, and Z are grouped with fields 1, 2, and 3. This grouping is one-to-one where one total GP value is paired with one area of intersection (AoI) ID and, ultimately, one satellite-based ET estimate for the irrigated area. This one-to-one pairing process was replicated for the HB, wherein the OWRD-provided POU groupings were matched to irrigated area polygons from OpenET, like Beamer and Hoskinson (2021).

3.2 Effective Precipitation and Consumptive Use

Effective precipitation is the amount of total precipitation on the cropped area that is available to meet the potential ET requirements in that area (Bos et al., 2009). Typically, it is computed by subtracting losses due to runoff and deep percolation beyond the rootzone of the crops from the total precipitation (Allen et al., 1998). Numerous methods exist for estimating effective precipitation, ranging from simple approaches (e.g., a ratio of reference ET and precipitation) to more intricate methods (e.g., involving detailed soil water balance and crop modeling) (Dastane, 1974; Kumar et al., 2017). Many empirical techniques are tailored to specific conditions, and their accuracy and applicability beyond those specific conditions are often limited unless they account for the factors influencing infiltration, runoff, and deep percolation (Feddes et al., 1988; Huntington et al., 2015, 2022; Stamm, 1967; USDA SCS, 1993).

Patwardhan et al. (1990) demonstrated the superior accuracy of the daily soil water balance method in estimating effective precipitation. This method considers soil moisture and plant available water, considering the water-holding capacity and root depths specific to crop areas within each model cell. The runoff from precipitation is computed using the USDA Natural Resources Conservation Service (NRCS) curve number (CN) method (USDA NRCS, 2004). CN values are scaled between dry and wet conditions based on antecedent soil water content, employing Hawkins et al. (1985) expressions.

In this study, we compute the basin-scale effective precipitation fraction ($P_{e_{fr}}$) using the ET-Demands model (Allen & Robison, 2007; Huntington et al., 2015, 2022; USBR, 2019). This model incorporates various factors, including daily gridMET precipitation data (Abatzoglou, 2013), antecedent soil moisture before a precipitation event, and deep percolation and surface runoff from precipitation. The ET-Demands model utilizes daily weather information, including reference evapotranspiration (ET_o), in conjunction with crop-specific growth curves. Widely applied, it has been used to assess historical and future irrigation water demands for specific USBR irrigation projects (USBR, 2016) and to estimate historical and future irrigation water requirements for the USBR's WaterSMART Basin Studies Program (USBR, 2023). We used ET-Demands to generate the basin-scale irrigated lands $P_{e_{fr}}$ and then computed the field-scale effective precipitation, $P_{e_{field}}$ (more details in Equation 1).

The Net ET or consumptive use is defined in Equation 1 as the actual ET (ET_a) minus effective precipitation ($P_{e_{field}}$). Here, we must subtract the portion of precipitation that is considered 'effective' or contributes to ET, i.e., $P_{e_{field}}$, from the total ET_a because it includes ET derived from precipitation (P_{field}). These Net ET estimates are foundational not only for estimating GP but also for assessing irrigation application rates and irrigation water requirements (Huntington et al., 2022).

$$Net\ ET = ET_a - P_{e_{field}} \tag{1}$$

where,

$Net\ ET$: Field-scale consumptive use.

ET_a: Total annual (January 1 to December 31) field-scale actual ET from the OpenET ensemble product (Melton et al., 2021; Volk et al., 2024).

$P_{e_{field}} = P_{e_{fr}} * \ P_{field}$, the field-scale effective precipitation.

$P_{e_{fr}}$: Basin-scale effective precipitation fraction for irrigated fields from ET-Demands (USBR, 2019).

P_{field}: Total annual gridMET precipitation (originally at ~4 km spatial resolution, Abatzoglou (2013)) aggregated at the field scale using spatial reductions available through the Google Earth Engine Python API (Gorelick et al., 2017).

3.3 Estimating Field-Scale Groundwater Pumping

We use the least-squares linear regression (Equation 2) to develop individual DV and HB-specific regression models between GP depth and Net ET. To make the model independent of area, we consider GP depths rather than GP volumes, i.e., dividing the reported pumping volumes from NDWR and OWRD by the respective irrigated field areas in the associated AoI (Figure 5).

$$\boldsymbol{GP = \ \widehat{\beta}_0 + \widehat{\beta}_1 * \ Net\ ET + \epsilon} \tag{2}$$

where,

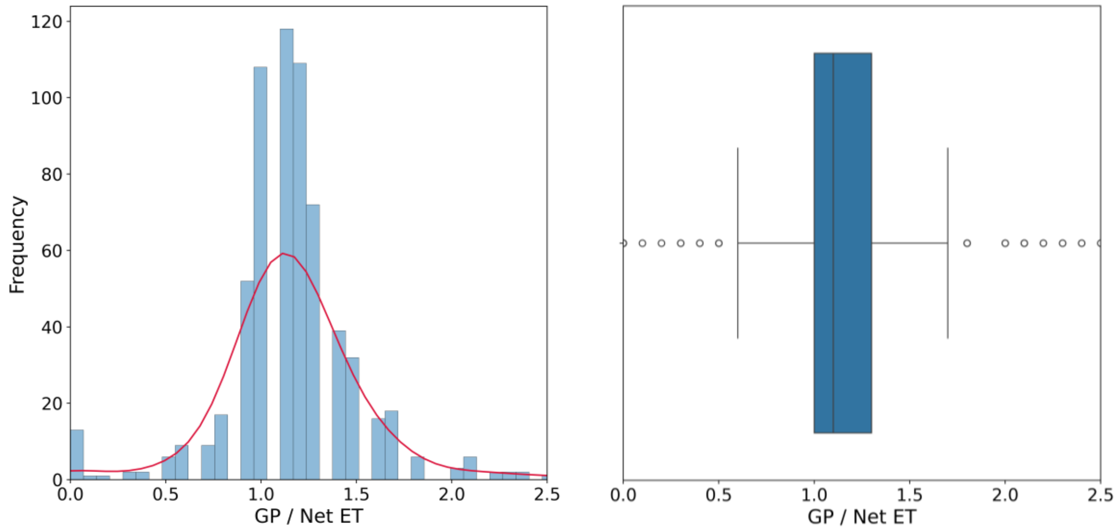
$\boldsymbol{GP}$: Metered annual groundwater pumping depth.

$\boldsymbol{Net\ ET}$: Field-scale consumptive use (Equation 1).

$\boldsymbol{\widehat{\beta}_0}$, $\boldsymbol{\widehat{\beta}_1}$: Regression coefficients.

$\boldsymbol{\epsilon}$: Random error associated with estimating $\boldsymbol{GP}$.

Prior to fitting the regression model, we remove fields with GP / Net ET ratios lying outside the (0.5, 1.5) interval. We obtain this interval based on histogram analysis (Figure 6 (a)) and boxplot-derived lower and upper limits (based on the interquartile range, Figure 6 (b), Hastie et al. (2001)) of the GP / Net ET ratios for DV, Nevada. Essentially, we remove fields where the reported metered GP data are below 50% or above 150% of the consumptive use.



(a) (b)

Figure 6. The (a) histogram and (b) boxplot distributions of the GP and Net ET ratios over DV, Nevada. The red line in (a) denotes the kernel density estimate (Hastie et al., 2001) of GP / Net ET, and the x-axis is cutoff at GP / Net ET = 2.5.

These discrepancies are typically caused due to flowmeter issues and changes in the flowmeters. In addition to applying the same (0.5, 1.5) interval, we remove ten fields in HB, Oregon, with purely surface water rights and combined groundwater and surface water rights based on the water source type data (Figure 4). Furthermore, for both DV and HB, we only consider fields where GP > 0. Overall, we discard 19% and 67% of the original DV and HB metered GP data, respectively, which essentially showcases the necessity for relying on field-scale ET to directly estimate GP as developing robust metering infrastructure is not trivial.

As for model evaluation, we employ bootstrapping (Hastie et al., 2001) to estimate the confidence intervals of the regression model. The nonparametric approach of bootstrapping provides a means of estimating confidence intervals and standard errors for the regression coefficients when relatively little data is available. The least-squares regression provides an estimate of the model parameters, these are not the true values of model parameters since the entire population is unknown and thus would be different if other data were used.

Here, we take a random sample with replacement using all data points, perform the least-squares regression (Equation 2) without fitting the intercept (i.e., $\hat{\beta}_0 = 0$), and estimate the regression coefficient $\hat{\beta}_1$. This process is repeated 1000 times after which we report the coefficient of determination (R^2), RMSE, MAE, and the coefficient of variation (CV, i.e., standard deviation of the predictions divided by the mean of the predictions). Note that we deliberately set $\hat{\beta}_0 = 0$ because, theoretically, GP is the ratio of the Net ET and irrigation efficiency (Section 3.1.4) (Howell, 2002).

We then compute the confidence interval (CI) and prediction interval (PI) using the bootstrap percentile interval method, which assigns the lower and upper 95% CI and PI values to the 2.5th and 97.5th percentile of the resulting bootstrap distributions. Moreover, we compare the predicted GP to the actual GP (both depth and volumes) at the field scale, perform basin-scale GP assessments, and analyze the observed GP residuals to test for normality (Sections 4.1 and 4.2). Additionally, we evaluate the performance of the linear regression model against ensemble machine learning algorithms, such as Random Forests (RF) (Breiman, 2001), Gradient Boosting Trees (GBT) (Friedman, 2001), and Extremely Randomized Trees (ERT) (Geurts et al., 2006) available through the scikit-learn (Pedregosa et al., 2011) and LightGBM (Ke et al., 2017) Python APIs (Section 4.3).

3.4 Calculating Irrigation Efficiency

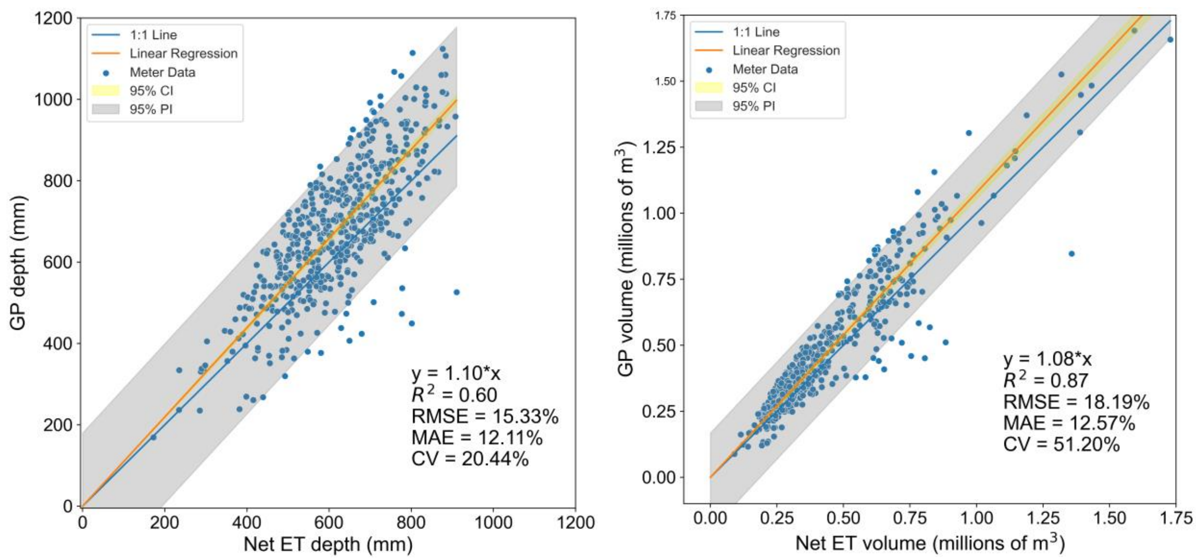
The irrigation efficiency (IE ∈ [0, 1]), as defined in Equation 3, is the ratio of the Net ET and GP (Howell, 2002). Conversely, GP can be obtained by dividing the Net ET by the IE. Therefore, the inverse of the slope of the fitted regression in Equation 2, i.e., $\hat{\beta}_1^{-1}$, gives us the IE (since $\hat{\beta}_0 = 0$). Here, we use the terms 'irrigation efficiency' and 'application efficiency' interchangeably (Howell, 2002).

$$\boldsymbol{IE} = \frac{\boldsymbol{Net\ ET}}{\boldsymbol{GP}} \tag{3}$$

4. RESULTS AND DISCUSSION

4.1 Field-scale GP estimates in DV, Nevada

We observe a good agreement (R^2 = 0.6, RMSE = 15.33%, MAE = 12.11%, CV=20.44%) between the metered GP depths and the predicted GP depths at the field scale using linear regression (Equation 2) over DV (Figure 7 (a)).

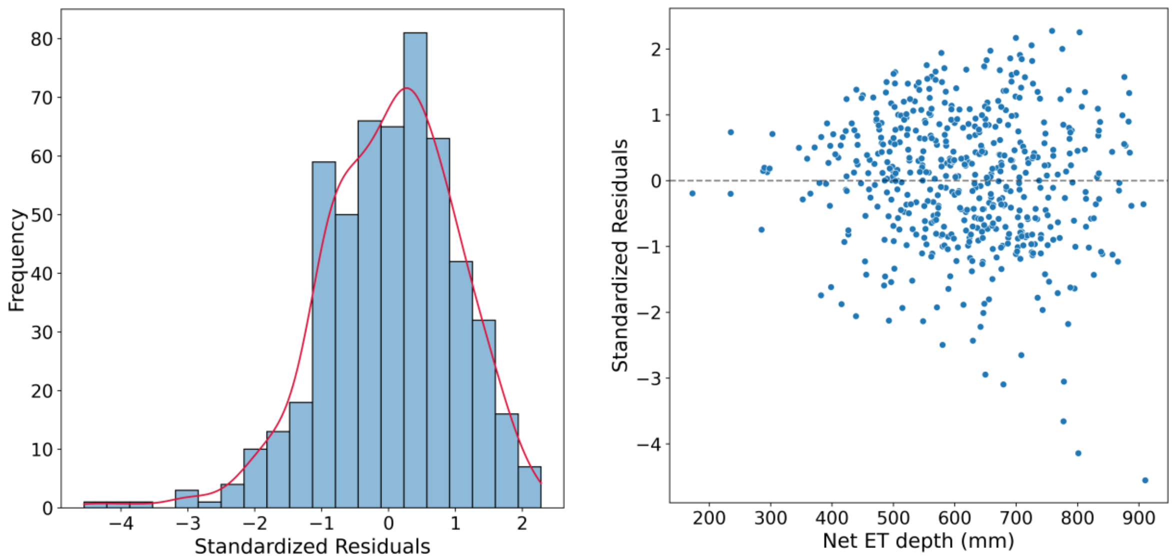


(a) (b)

Figure 7. Scatter plots of the fitted (a) GP depth and (b) GP volume over DV, with the Net ET depth and Net ET volume as the corresponding predictors. There are a total of 533 samples after the outlier removal process (Section 3.3). The 95% CI and PI are obtained using bootstrapping. Here, y and x denote the response (GP) and the predictor (Net ET) variables, respectively. The RMSE and MAE percentages are obtained by dividing the RMSE and MAE by the mean of the actual metered GP depth/volume.

Additionally, we achieve R^2 = 0.87, RMSE = 18.19%, MAE = 12.57%, and CV = 51.2% when the predicted and metered depths are converted to the volume space by multiplying the irrigated field boundary areas (Figure 7 (b)). This substantial increase in the R^2 can be attributed to the artificial correlations imposed by multiplying the field areas, i.e., a larger field will have higher GP volume than a smaller one, which is also evident from the ~31% increase in the CV (the predicted volumes have a higher variability than the depths because of the varying field areas). Nevertheless, we report the error metrics in both the depth and volume space to demonstrate the effectiveness of our approach.

The slopes of 1.1 and 1.08 in Figures 7 (a)-(b) indicate that the average IE for DV is about 92%, which aligns with typical center pivot system efficiencies (Howell, 2002). Accordingly, the standardized GP depth residuals, calculated as observed GP depth minus predicted GP depth, approximately follow a normal distribution (skewness = -0.59, kurtosis = 1.15) and mostly lie in the [-2, 2] interval (Figure 8 (a)). There are also no observable systematic patterns in the standardized GP depth residual vs. the Net ET depth scatter plot (Figure 8 (b)). Moreover, the basin-scale comparison (Figure 9) of the metered and predicted annual total GP volumes further showcases the reliability of our approach.



(a) (b)

Figure 8. Residual analysis for the fitted linear regression using the DV meter data showing the (a) standardized residual histogram and (b) scatter plot of the standardized residuals vs. Net ET depth. The residuals are calculated as observed GP depth minus the predicted GP depth. The red line in (a) denotes the kernel density estimate like before in Figure 6 (a).

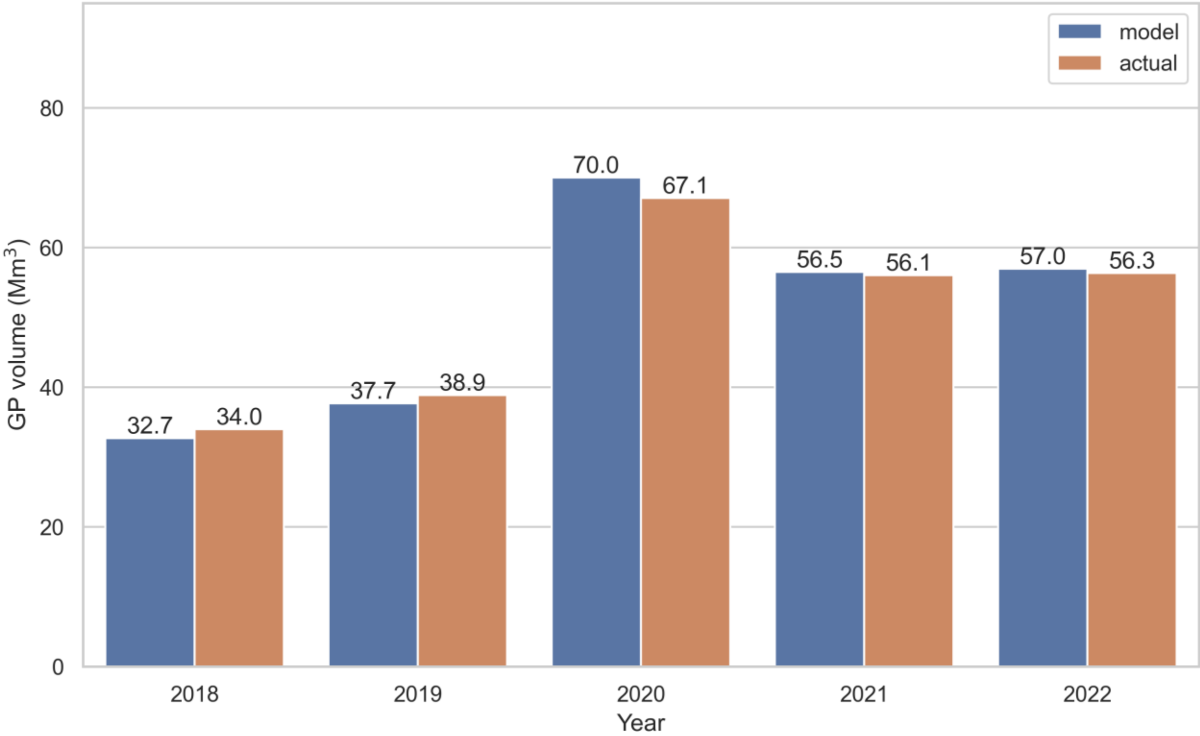
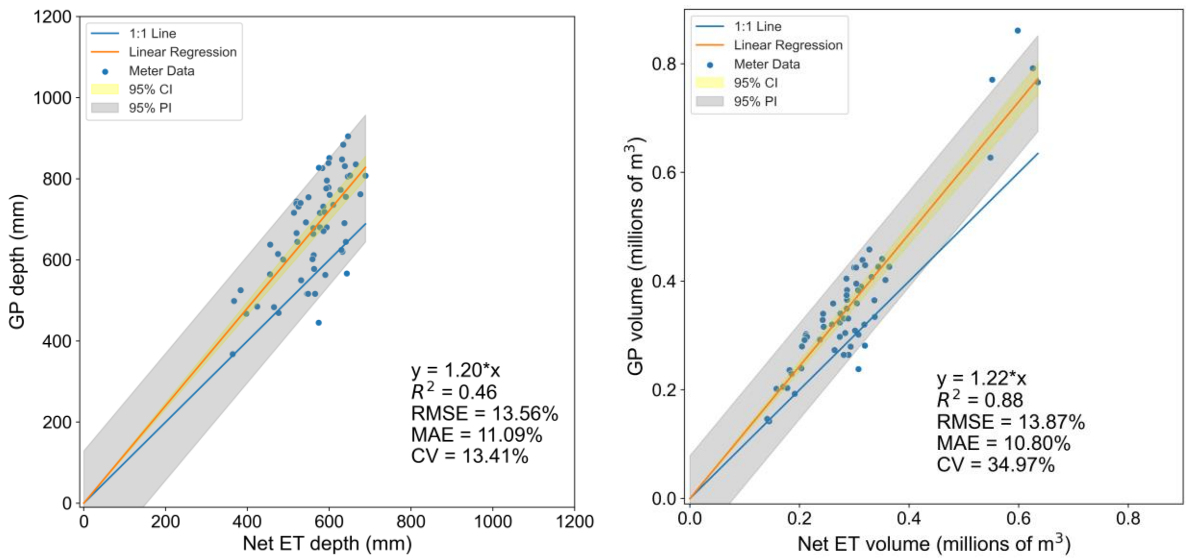


Figure 9. Comparison of the basin-scale total annual GP volumes in DV, Nevada. Note that the actual GP volumes are computed using the field data which are kept after the outlier removal process (Section 3.3).

4.2 Field-scale GP estimates in HB, Oregon

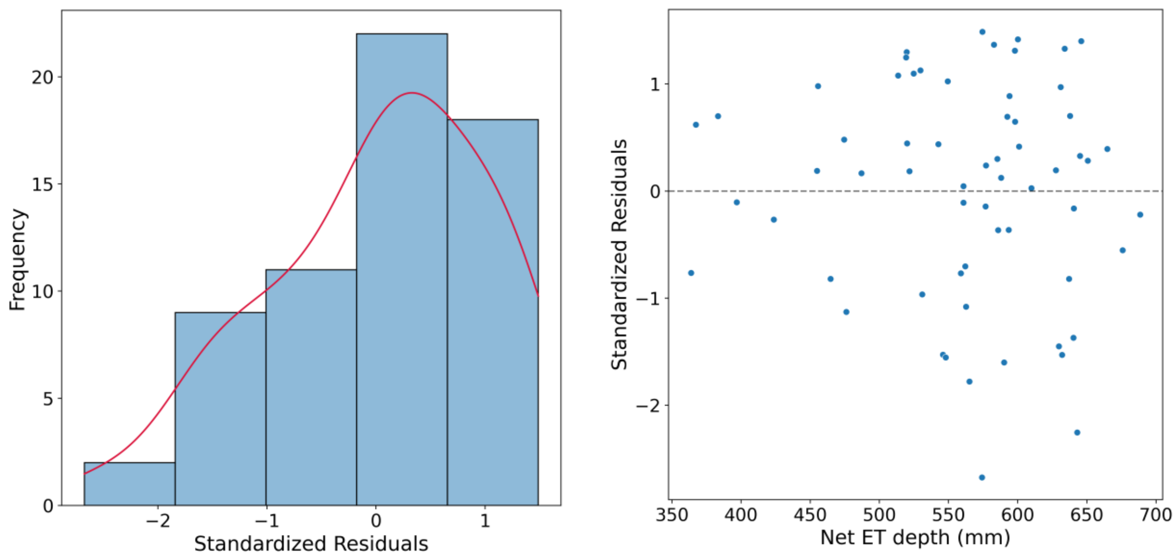
For HB, Oregon, we observe a satisfactory agreement (R^2 = 0.46, RMSE = 13.56%, MAE = 11.09%, and CV = 13.41%) between the metered GP depths and the predicted GP depths at the field scale using linear regression (Figure 10 (a)). Additionally, we obtain R^2 = 0.88, RMSE = 13.87%, MAE = 10.8%, and CV = 34.97% considering the GP volumes (Figure 10 (b)). These substantial increases in the R^2 and CV are due to the artificial correlations imposed by the field areas, and the variability of the field areas, respectively.

The slopes of 1.2 and 1.22 in Figures 10 (a)-(b) imply an average IE of 83%, which aligns with typical center pivot system efficiencies (Howell, 2002). However, the HB IE is about 9% less than that of DV. The standardized GP depth residuals approximately follow a normal distribution (skewness = -0.56, kurtosis = -0.39) and mostly lie in the [-2, 1] interval (Figure 11 (a)). Like DV, there are no observable systematic patterns in the standardized GP depth residual vs. the Net ET depth scatter plot (Figure 11 (b)). Additionally, the basin-scale comparison (Figure 12) of the metered and predicted annual total GP volumes shows good agreement and again demonstrates the reliability of our approach.



(a) (b)

Figure 10. Scatter plots of the fitted (a) GP depth and (b) GP volumes over HB, with the Net ET depth and Net ET volume as the corresponding predictors. There are a total of 62 samples after the outlier removal process (Section 3.3).



(a) (b)

Figure 11. Residual analysis for the fitted linear regression using the HB meter data showing the (a) standardized residual histogram and (b) scatter plot of the standardized residuals vs. Net ET depth. The residuals are calculated as observed GP depth minus the predicted GP depth. The red line in (a) denotes the kernel density estimate like before in Figure 6 (a).

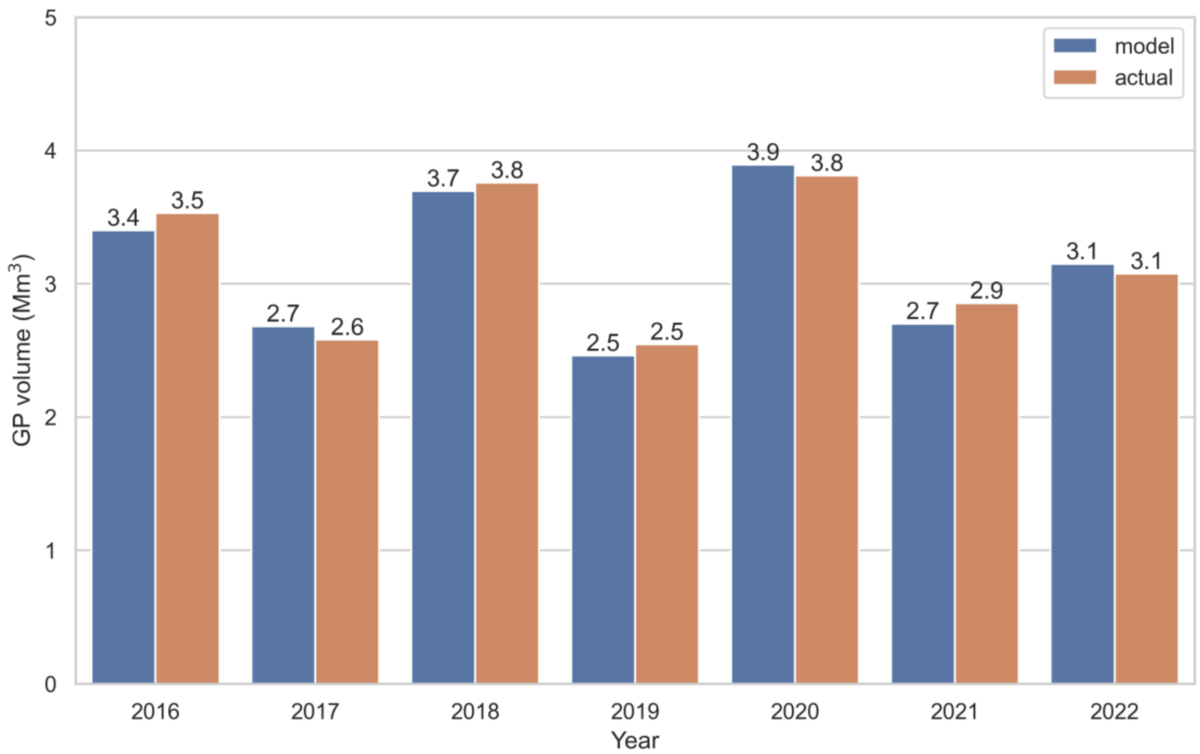


Figure 12. Comparison of the basin-scale total annual GP volumes in HB, Oregon. Note that the actual GP volumes are computed using the field data which are kept after the outlier removal process (Section 3.3).

4.3 Comparison with Ensemble Machine Learning

Here, we only compare the linear regression and ensemble ML model performances for predicting GP depth over DV, Nevada, consisting of 533 valid samples (2018-2022). Since there are only 62 valid samples in HB, Oregon (2016-2022), developing ML models is unreasonable.

We perform a random 70%-30% training and test data split to assess the model performances through five-fold cross-validation (Hastie et al., 2001). The training, validation, and test metrics are shown in Table 2, where the validation data are automatically generated using the five-fold cross-validation technique, i.e., 20% of the training data are used to tune the hyperparameters of each ML model (Supplementary Table 1). We use the OpenET ensemble product to calculate Net ET like the linear regression model and include all other actual ET models (Table 1) as input predictors along with additional predictors described in Section 2.2.4. Overall, there are 28 predictors in our ML models listed in Supplementary Table 2.

Table 2. The training, validation, and test error metrics (rounded to two decimal places) for the ensemble ML models. The ERT model shows the best performance across all metrics for the test data and has the least overfitting, i.e., training, validation, and test error metrics are closer to each other compared to the other models (GBT has the highest overfitting).

Data	**Metrics**	**Ensemble ML models**		
		ERT	GBT	RF
Training	R^2	0.73	0.95	0.82
	RMSE (%)	12.21	3.39	10.07
	MAE (%)	9.51	2.54	7.89
	CV (%)	16.61	22.75	19.0
Validation	R^2	0.56	0.53	0.58
	RMSE (%)	15.84	16.42	15.5
	MAE (%)	12.29	12.75	12.12
	CV (%)	15.42	20.29	17.63
Test	R^2	0.63	0.62	0.63
	RMSE (%)	14.82	14.94	14.91
	MAE (%)	11.46	11.68	11.59
	CV (%)	17.43	18.21	17.78

We find that the ERT model gives the best prediction performance with test R^2 = 0.63, RMSE = 14.82%, MAE = 11.46%, and CV = 17.43%, which is marginally better than the DV, Nevada linear regression model (R^2 = 0.6, RMSE = 15.33%, MAE = 12.11%, and CV = 20.44%, Section 4.1). The corresponding permutation importance (Breiman, 2001) plots of the top five features or predictors for the training and test data are shown in Supplementary Figures 1 and 2, respectively. These show that the Net ET, field-scale actual ET, air temperature, relative humidity, soil depth, effective precipitation, and NDVI constitute the key predictors across the three ML models, with Net ET, being the most important one as removing it from the predictor set substantially decreases the model performance, with an average 12%-15% increase in training (including validation) RMSE, and 9%-11% increase in test RMSE.

Thus, the linear regression and ensemble ML model results strongly support the three hypotheses of our study (Section 1.3), 1) field-scale satellite-based ET estimates are well-correlated with field-scale metered GP data, 2) statistical relationships between field-scale satellite-based ET and GP data are useful for QAQC of GP records and assessment of prior estimates, and 3) ET-based predictions of GP compare reasonably well to metered GP at the field and basin scales.

4.4 OpenET Ensemble vs. Individual ET Models

Here, we compare the performance of the individual OpenET models (Table 1) with that of the OpenET ensemble in predicting GP depths using both linear regression (DV and HB) and ML (only DV) methods.

4.4.1 ET comparison through Linear Regression

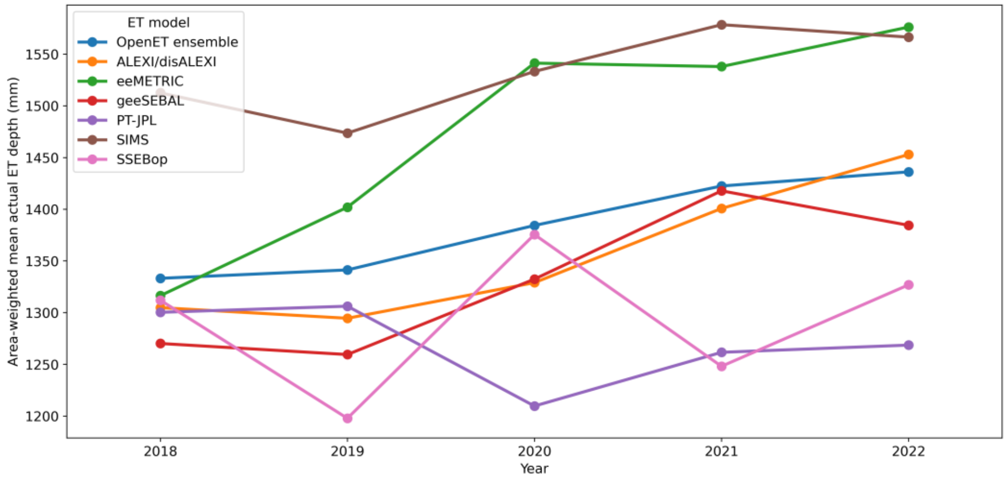
In DV, Nevada, the OpenET ensemble produces the best error metrics in estimating the GP depths (Table 3, Figure 7 (a)). For each model (Supplementary Figures 3 (a)-(f)), we use the same (0.5, 1.5) interval for removing the outliers based on the GP / Net ET ratios, where the

Net ET is calculated using the OpenET ensemble and the ET-Demands-derived effective precipitation (Section 3.2).

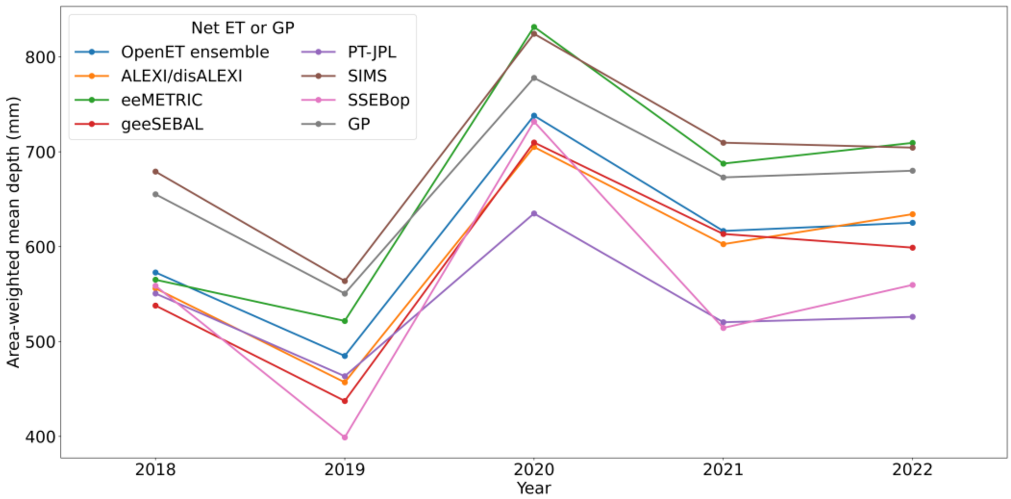
Table 3. Comparison of the linear regression model metrics (GP depths, DV, Nevada) and slopes for different field-scale ET models used to calculate the Net ET. The metrics and slopes are rounded to two decimal places.

ET Model	**GP depth metrics**				**Slope**
	R^2	RMSE (%)	MAE (%)	CV (%)	
OpenET ensemble	0.6	15.33	12.11	20.44	1.1
ALEXI/DisALEXI	0.44	18.15	14.8	24.29	1.11
eeMETRIC	0.55	16.3	12.78	22.58	0.99
geeSEBAL	0.49	17.27	13.62	22.21	1.19
PT-JPL	0.51	16.99	13.16	17.47	1.25
SIMS	0.36	19.34	13.61	21.15	0.95
SSEBop	0.4	18.76	14.78	27.61	1.18

The performance of individual ET models was assessed over the same 533 samples as in Section 4.1. Although selecting the outliers based on the individual ET model-specific GP / Net ET ratios and adjusting the intervals from histogram and boxplot analyses would have improved the corresponding metrics, using the same OpenET ensemble-derived GP / Net ET ratios make the comparison more consistent. Table 3 shows that eeMETRIC and PT-JPL are the most skillful models after the ensemble mean, with SIMS having the least skill. However, the slopes for eeMETRIC and SIMS are close to 1, implying that the consumptive use equals pumping, i.e., Net ET = GP, which is not practical and could be due to both these ET models being biased high.



(a)



(b)

Figure 13. Comparisons of the area-weighted mean annual (a) ET depths and (b) Net ET and metered GP depths for each ET model in DV, Nevada. Note that the area-weighted means in (a) and (b) are computed after the outlier removal process described in Section 3.3 (Figure 6).

To investigate this issue, we compare the area-weighted mean annual actual ET depths (Figure 13 (a)) and the area-weighted mean annual Net ET depths with the metered GP depths for each ET model (Figure 13 (b)). We observe that eeMETRIC is biased high between 2020 and 2022, whereas SIMS is biased high across all years. The consistent high bias in SIMS is expected because it assumes well-irrigated crop conditions, and therefore, exhibits a positive bias particularly for deficit irrigated crops and croplands with short-term or intermittent crop water stress (OpenET, 2023; Volk et al., 2024). Both SSEBop and PT-

JPL vary substantially, but Net ET predictions from these two models are consistently lower than the GP, which is similar to ALEXI/DisALEXI, geeSEBAL. However, it is expected that most of these models are biased low due to model limitations associated with advection, aridity, and sharp contrasts between irrigated and non-irrigated arid landscapes (OpenET, 2023; Volk et al., 2024). The OpenET ensemble value is the average across all models after up to two outliers are identified and removed following the median absolute deviation (MAD) from the median approach (Hampel, 1974; Leys et al., 2013). The calculation of an ensemble mean is a useful and common technique for combining model predictions that each have positive or negative biases and random errors and is especially useful for water management where single values are commonly required (Thompson et al., 1977; Kirtman et al., 2014; Bai et al., 2021). Notably the OpenET ensemble mean had the highest skill, with a slope value that follows our conceptual model and aligns with published irrigation efficiencies associated with high efficiency center pivot irrigation systems (Howell, 2002).

Moreover, from Table 4, and Supplementary Figures 4 and 5, we observe that the OpenET ensemble is also consistent in HB, Oregon, with similar GP depth R^2, RMSE, and MAE metrics like the ones based on the SSEBop Net ET (which performs slightly better), and leads to the best GP depth precision in terms of CV. Thus, relying on the OpenET ensemble leads to a more consistent approach because of these high and low bias issues with the individual models (Volk et al., 2024).

Table 4. Comparison of the linear regression model metrics (GP depths, HB, Oregon) and slopes for different field-scale ET models used to calculate the Net ET. The metrics and slopes are rounded to two decimal places.

ET Model	**GP depth metrics**				**Slope**
	R^2	RMSE (%)	MAE (%)	CV (%)	
OpenET ensemble	0.46	13.56	11.09	13.41	1.2

ALEXI/DisALEXI	0.11	17.3	14.38	15.08	1.3
eeMETRIC	0.33	15.06	12.14	15.16	1.19
geeSEBAL	0.17	16.78	13.83	16.68	1.11
PT-JPL	0.3	15.38	12.46	14.19	1.32
SIMS	-1.6	29.64	16.37	30.78	1.07
SSEBop	0.48	13.31	10.56	17.92	1.23

4.4.2 ET comparison through Machine Learning

To compare the ML model performances corresponding to each ET model in DV, we did not use the full 28 predictors as we did in Section 4.3. Instead, we used the ET model-specific Net ET and the actual ET and removed other ET predictors in each case. Therefore, the ML models in Table 5 rely on 22 predictors (see Supplementary Table 2 for more details). The training, validation, and test data are generated in the same way as in Section 4.3., i.e., 70%-30% training and test data split, followed by the automatic validation data generation (20% from the training data) using the five-fold cross-validation technique.

Table 5. Comparison of the ML model metrics *(*GP depths, DV, Nevada*)* for different field-scale ET models used to calculate the Net ET. For each of the ET models, the metrics (rounded to two decimal places) are only reported for the test data obtained using the best ML model in terms of the RMSE and overfitting.

ET Model	Best ML model	GP depth metrics			
		R^2	RMSE (%)	MAE (%)	CV (%)
OpenET ensemble	ERT	0.62	14.96	11.47	17.06
ALEXI/DisALEXI	RF	0.59	15.62	12.23	15.93
eeMETRIC	RF	0.62	15.06	11.6	17.96
geeSEBAL	GBT	0.61	15.16	11.92	17.92

PT-JPL	GBT	0.61	15.25	11.65	18.53
SIMS	GBT	0.6	15.39	11.82	18.61
SSEBop	ERT	0.59	15.52	12.13	16.43

From Table 5, we find that the OpenET ensemble leads to the best performance metrics in terms of R^2, RMSE, and MAE. Although the ML models appear to be more robust to changes in the ET models compared to the linear regression, these results are only for a single test data. Ideally, these comparisons should be repeated over thousands of model iterations and train-test partitions for more reliable reporting of these metrics. Nevertheless, the OpenET ensemble product demonstrates consistent results across different statistical and ML modeling paradigms and the two study areas (DV, Nevada and HB, Oregon).

5. CONCLUSIONS

This is the first study to predict field-scale groundwater pumping and concurrently provide estimates of irrigation efficiencies using integrated remote sensing, irrigation, and climate data in a statistical learning framework. We used statistical (linear regression and bootstrapping) and ensemble machine learning (Random Forests, Gradient Boosting Trees, and Extremely Randomized Trees) approaches to predict field-scale groundwater pumping in Diamond Valley, Nevada, and Harney Basin, Oregon. We relied on several remote sensing, irrigation, and climate datasets for modeling. The primary datasets include OpenET (Melton et al., 2021; Volk et al., 2024) ensemble-derived field-scale actual evapotranspiration, ET-Demands (USBR, 2023) and gridMET (Abatzoglou, 2013)-derived effective precipitation, and carefully attributed field boundaries and water source type data (Huntington et al., 2018; Beamer & Hoskinson, 2021). Moreover, we ingested multiple temporally static (elevation, soil depth, saturated hydraulic conductivity, hydrologic soil group) and dynamic geospatial datasets (reference evapotranspiration, relative humidity, air temperature, NDVI, and others) as additional predictors to the machine learning models.

The linear regression and machine learning model results demonstrate that the OpenET ensemble product leads to more consistent results compared to the individual ET models across the two study areas and simultaneously aids in quality assurance and quality control of the reported pumping data. More specifically, the mean absolute errors for field-scale groundwater pumping depth are 12% and 11% for Diamond Valley and Harney Basin, respectively, and the corresponding root mean square errors are 15% and 14%. The regression models can explain 50%-60% variance in the pumping depths and ~90% variance in the pumping volumes. Furthermore, the estimated average irrigation efficiency of 88% (92% and 83% for Diamond Valley and Harney Basin, respectively) aligns with known center pivot system efficiencies (Howell, 2002).

Regarding the limitations of our approach, the primary bottleneck is the amount of pre-processing time involved in linking the points of diversions (wells) to the places of use (fields). Matching the wells to the fields is an extremely tedious yet critical task as it directly influences the model performance. Other limitations include data scarcity in both the study areas, particularly the Harney Basin, where there is mixed water use, i.e., fields with both groundwater and surface water rights, and hence, a few fields had to be discarded completely because of this issue.

Still, our data-driven approach provides a more systematic way of estimating groundwater pumping than conventional methods based on water right duties, potential crop ET, low-quality meter readings, or assumed values. As part of future work, we aim to incorporate climate model projection data to generate hindcasts and future projections of groundwater pumping at regional or basin scales. The broader goal of our study is to present water resource and user communities with valuable insights into water use and budgets, supporting the implementation of field-scale management strategies across both metered and unmetered groundwater basins in Nevada, Oregon, and other states in the western U.S. Essentially, this work is an advancement toward improved field-scale evaluations of groundwater pumping, consumptive use, and irrigation efficiencies, thereby contributing to more efficient and sustainable water management solutions.

CRediT AUTHORSHIP CONTRIBUTION STATEMENT

Thomas J. Ott[‡]: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Visualization, Writing – Original Draft; **Sayantan Majumdar[‡]:** Conceptualization, Methodology, Software, Investigation, Formal analysis, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing; **Justin L. Huntington:** Funding acquisition, Supervision, Project administration, Conceptualization, Methodology, Investigation, Writing – Review & Editing; **Christopher Pearson:** Data Curation; **Matt Bromley:** Data Curation; **Blake A. Minor:** Data Curation; **Charles G. Morton:** Data Curation; **Sachiko Sueki:** Data Curation, Writing – Review & Editing; **Jordan P. Beamer:** Data Curation, Visualization, Writing – Review & Editing; **Richard Jasoni:** Supervision, Writing – Review & Editing.

ACKNOWLEDGMENTS

We would like to express our gratitude for the support received from the State of Nevada / U.S. Department of the Treasury (grant number 27042), United States Geological Survey (USGS) NASA Landsat Science Team (grant number 140G0118C0007), USGS Water Resources Research Institute (grant G22AC00584-00), Desert Research Institute Maki Endowment, and Windward Fund. Our appreciation extends to the open-source software and

[‡] These two authors have equal contribution.

data communities for their generosity in making their resources available to the public. We also acknowledge the invaluable contribution of the OpenET consortium and respective funding partners, Google Earth Engine, Nevada Division of Water Resources, and the Oregon Water Resources Department for providing the necessary computational resources and datasets pertaining to ET, groundwater pumping, places of use, and other shapefiles used in this research. We also extend our gratitude to Dr. Richard G. Niswonger (USGS) for his helpful suggestions on handling discrepancies in the metered pumping data and to Marty Plaskett and Mark Moyle (Diamond Valley farmers) for their interest and support in our efforts. Furthermore, we are thankful to our colleagues and families for their unwavering motivation and support throughout this endeavor. It is important to note that any opinions, findings, conclusions, or recommendations presented in this manuscript are solely those of the authors and do not necessarily reflect the viewpoints of the funding agencies.

DATA AVAILABILITY

The project codes, pumping, and irrigation data are publicly available at https://github.com/montimaj/OpenET-GW. All the remote sensing and climate data are publicly available and were automatically downloaded using the Google Earth Engine Python API.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, *33*(1), 121–131. https://doi.org/10.1002/joc.3413

ADWR. (2018). *Groundwater Flow Model of the Willcox Basin*. Arizona Department of Water Resources. https://new.azwater.gov/sites/default/files/Willcox_Report_2018.pdf

ADWR. (2023). *Annual Report 2022*. https://new.azwater.gov/sites/default/files/media/ANNUALREPORT_WEB_1.pdf

Ahamed, A., Knight, R., Alam, S., Pauloo, R., & Melton, F. (2022). Assessing the utility of remote sensing data to accurately estimate changes in groundwater storage. *Science of The Total Environment*, *807*, 150635. https://doi.org/10.1016/j.scitotenv.2021.150635

Allen, R. G., Irmak, A., Trezza, R., Hendrickx, J. M. H., Bastiaanssen, W., & Kjaersgaard, J. (2011). Satellite-based ET estimation in agriculture using SEBAL and METRIC. *Hydrological Processes*, *25*(26), 4011–4027. https://doi.org/10.1002/hyp.8408

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). *Crop Evapotranspiration - Guidelines for Computing Crop Water Requirements - FAO Irrigation and drainage paper 56*. FAO - Food and Agriculture Organization of the United Nations. http://www.fao.org/3/X0490E/x0490e00.htm

Allen, R. G., & Robison, C. W. (2007). *Evapotranspiration and consumptive irrigation water requirements for Idaho. Idaho Waters Digital Library, Digital Initiatives, University of Idaho Library*. https://www.lib.uidaho.edu/digital/iwdl/docs/iwdl-200703.html

Allen, R. G., Tasumi, M., Morse, A., & Trezza, R. (2005). A Landsat-based energy balance and evapotranspiration model in Western US water rights regulation and planning. *Irrigation and Drainage Systems*, *19*(3–4), 251–268. https://doi.org/10.1007/s10795-005-5187-z

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*(4), 380–394. https://doi.org/10.1061/(ASCE)0733-9437(2007)133:4(380)

Anderson, M., Gao, F., Knipper, K., Hain, C., Dulaney, W., Baldocchi, D., Eichelmann, E., Hemes, K., Yang, Y., Medellin-Azuara, J., & Kustas, W. (2018). Field-Scale Assessment of Land and Water Use Change over the California Delta Using Remote Sensing. *Remote Sensing*, *10*(6), 889. https://doi.org/10.3390/rs10060889

Anderson, M., Norman, J. M., Mecikalski, J. R., Otkin, J. A., & Kustas, W. P. (2007). A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. *Journal of Geophysical Research: Atmospheres*, *112*(D10). https://doi.org/10.1029/2006JD007506

Bai, Y., Zhang, S., Bhattarai, N., Mallick, K., Liu, Q., Tang, L., Im, J., Guo, L., & Zhang, J. (2021). On the use of machine learning based ensemble approaches to improve evapotranspiration estimates from croplands across a wide environmental gradient. Agricultural and Forest Meteorology, 298–299, 108308. https://doi.org/10.1016/j.agrformet.2020.108308

Bastiaanssen, W. G. M., Menenti, M., Feddes, R. A., & Holtslag, A. A. M. (1998). A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. *Journal of Hydrology*, *212–213*, 198–212. https://doi.org/10.1016/S0022-1694(98)00253-4

Beamer, J., & Hoskinson, M. (2021). *Historical Irrigation Water Use and Groundwater Pumpage Estimates in the Harney Basin, Oregon, 1991-2018 (Open File Report No. 2021-02)*. https://www.oregon.gov/owrd/wrdreports/OWRD_OFR_2021-02_Harney_Basin_METRIC_Irrigation_Use_Report.pdf

Berger, D. L., Mayers, C. J., Garcia, C. A., Buto, S. G., & Huntington, J. M. (2016). *Budgets and chemical characterization of groundwater for the Diamond Valley flow system, central Nevada, 2011–12: U.S. Geological Survey Scientific Investigations Report 2016–5055*. https://doi.org/10.3133/sir20165055

Blankenau, P. A., Kilic, A., & Allen, R. (2020). An evaluation of gridded weather data sets for the purpose of estimating reference evapotranspiration in the United States. Agricultural Water Management, 242, 106376. https://doi.org/10.1016/j.agwat.2020.106376

Bos, M. G., Kselik, R. A. L., Allen, R. G., & Molden, D. (2009). *Water Requirements for Irrigation and the Environment*. Springer Netherlands. https://doi.org/10.1007/978-1-4020-8948-0

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brookfield, A. E., Zipper, S., Kendall, A. D., Ajami, H., & Deines, J. M. (2023). Estimating Groundwater Pumping for Irrigation: A Method Comparison. *Groundwater*. https://doi.org/10.1111/gwat.13336

Bugenig, D. C. (2017). *Appendix I– Groundwater Flow Modeling Report Supporting Banking Depreciation (DIAMOND VALLEY GROUNDWATER MANAGEMENT PLAN)*. http://water.nv.gov/documents/Final%20DV%20GMP%20for%20Petition.pdf

Burt, C. M., Clemmens, A. J., Strelkoff, T. S., Solomon, K. H., Bliesner, R. D., Hardy, L. A., Howell, T. A., & Eisenhauer, D. E. (1997). Irrigation Performance Measures: Efficiency and Uniformity. *Journal of Irrigation and Drainage Engineering*, *123*(6), 423–442. https://doi.org/10.1061/(ASCE)0733-9437(1997)123:6(423)

Butler, J. J., Whittemore, D. O., Wilson, B. B., & Bohling, G. C. (2018). Sustainability of aquifers supporting irrigated agriculture: a case study of the High Plains aquifer in Kansas. *Water International*, *43*(6), 815–828. https://doi.org/10.1080/02508060.2018.1515566

Carroll, R. W. H., Pohll, G., McGraw, D., Garner, C., Knust, A., Boyle, D., Minor, T., Bassett, S., & Pohlmann, K. (2010). Mason Valley Groundwater Model: Linking Surface Water and Groundwater in the Walker River Basin, Nevada [1]. *JAWRA Journal of the American Water Resources Association*, *46*(3), 554–573. https://doi.org/10.1111/j.1752-1688.2010.00434.x

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *28*(15), 2031–2064. https://doi.org/10.1002/joc.1688

Dastane, N. G. (1974). *Effective Rainfall in Irrigated Agriculture (Volume 25 of Irrigation and drainage paper)*. Food and Agriculture Organization of the United Nations. https://www.fao.org/3/X5560E/X5560E00.htm

Deines, J. M., Kendall, A. D., Butler, J. J., & Hyndman, D. W. (2019). Quantifying irrigation adaptation strategies in response to stakeholder-driven groundwater management in the US High Plains Aquifer. *Environmental Research Letters*, *14*(4), 044014. https://doi.org/10.1088/1748-9326/aafe39

Dieter, C. A., Maupin, M. A., Caldwell, R. R., Harris, M. A., Ivahnenko, T. I., Lovelace, J. K., Barber, N. L., & Linsey, K. S. (2018). *Estimated use of water in the United States in 2015: U.S. Geological Survey Circular 1441*. https://doi.org/10.3133/cir1441

Dogrul, E. C., Brush, C., & Kadir, T. (2016). Groundwater Modeling in Support of Water Resources Management and Planning under Complex Climate, Regulatory, and Economic Stresses. *Water*, *8*(12), 592. https://doi.org/10.3390/w8120592

Fanning, J. L., Schwarz, G. E., & Lewis, W. C. (2001). *A field and statistical modeling study to estimate irrigation water use at Benchmark Farms study sites in southwestern Georgia, 1995-96 (U.S. Geological Survey Water-Resources Investigations Report 2000-4292)*. https://doi.org/10.3133/wri20004292

Faunt, C. C. (2009). Groundwater availability of the Central Valley Aquifer, California. In C. C. Faunt (Ed.), *U.S. Geological Survey Professional Paper 1766*. https://doi.org/10.3133/pp1766

Feddes, R. A., Kabat, P., Van Bakel, P. J. T., Bronswijk, J. J. B., & Halbertsma, J. (1988). Modelling soil water dynamics in the unsaturated zone — State of the art. *Journal of Hydrology*, *100*(1–3), 69–111. https://doi.org/10.1016/0022-1694(88)90182-5

Filippelli, S. K., Sloggy, M. R., Vogeler, J. C., Manning, D. T., Goemans, C., & Senay, G. B. (2022). Remote sensing of field-scale irrigation withdrawals in the central Ogallala aquifer region. *Agricultural Water Management*, *271*, 107764. https://doi.org/10.1016/j.agwat.2022.107764

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*(3), 901–919. https://doi.org/10.1016/j.rse.2007.06.025

Foster, T., Mieno, T., & Brozović, N. (2020). Satellite-Based Monitoring of Irrigation Water Use: Assessing Measurement Errors and Their Implications for Agricultural Water Management Policy. *Water Resources Research*, *56*(11). https://doi.org/10.1029/2020WR028378

Frenzel, S. A. (1985). Comparison of Methods for Estimating Ground-Water Pumpage for Irrigation. *Ground Water*, *23*(2), 220–226. https://doi.org/10.1111/j.1745-6584.1985.tb02795.x

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://jerryfriedman.su.domains/ftp/trebst.pdf

Garcia, C. A., Corson-Dosch, N. T., Beamer, J. P., Gingerich, S. B., Grondin, G. H., Overstreet, B. T., Haynes, J. V., & Hoskinson, M. D. (2021). *Hydrologic budget of the Harney Basin groundwater system, southeastern Oregon (ver. 1.1, November 2022): U.S. Geological Survey Scientific Investigations Report 2021–5128, 144 p.* https://doi.org/10.3133/sir20215128

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Gingerich, S. B., Garcia, C. A., & Johnson, H. M. (2022). *Groundwater resources of the Harney Basin, southeastern Oregon: U.S. Geological Survey Fact Sheet*. https://doi.org/10.3133/fs20223052

Gingerich, S. B., Johnson, H. M., Boschmann, D. E., Grondin, G. H., & Garcia, C. A. (2022). *Groundwater resources of the Harney Basin, Oregon: U.S. Geological Survey Scientific Investigations Report 2021–5103, 118 p.* https://doi.org/10.3133/sir20215103

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. https://doi.org/10.1016/j.rse.2017.06.031

Hampel, F.R. (1974). The Influence Curve and Its Role in Robust Estimation. Journal of the American Statistical Association, 69(346), 383. https://doi.org/10.2307/2285666

Harrill, J.R. (1968). Hydrologic response to irrigation pumping in Diamond Valley, Eureka and Elko Counties, Nevada, 1950-65, with a section on surface water by R.D. Lamke: Nevada Department of Conservation and Natural Resources, Water Resources Bulletin 35, 85 p.

Hasan, M. F., Smith, R., Vajedian, S., Pommerenke, R., & Majumdar, S. (2023). Global land subsidence mapping reveals widespread loss of aquifer storage capacity. *Nature Communications*, *14*(1), 6180. https://doi.org/10.1038/s41467-023-41933-z

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer New York.

Hawkins, R. H., Hjelmfelt, A. T., & Zevenbergen, A. W. (1985). Runoff Probability, Storm Depth, and Curve Numbers. *Journal of Irrigation and Drainage Engineering*, *111*(4), 330–340. https://doi.org/10.1061/(ASCE)0733-9437(1985)111:4(330)

Herrera-García, G., Ezquerro, P., Tomás, R., Béjar-Pizarro, M., López-Vinielles, J., Rossi, M., Mateos, R. M., Carreón-Freyre, D., Lambert, J., Teatini, P., Cabral-Cano, E., Erkens, G., Galloway, D., Hung, W.-C., Kakar, N., Sneed, M., Tosi, L., Wang, H., & Ye, S. (2021). Mapping the global threat of land subsidence. *Science*, *371*(6524), 34 LP – 36. https://doi.org/10.1126/science.abb8549

Hobbins, M. T., & Huntington, J. L. (2017). Evapotranspiration and Evaporative Demand, Chapter 44. In V. P. Singh (Ed.), Handbook of Applied Hydrology (Second Edition). McGraw-Hill Education, New York.

Howell, T. (2002). Irrigation Efficiency. In R. Lal (Ed.), *Encyclopedia of Soil Science* (1st ed.). Marcel Dekker, Inc. https://www.researchgate.net/publication/43256707_Irrigation_Efficiency

Huntington, J., & Allen, R. G. (2009). Evapotranspiration and Net Irrigation Water Requirements for Nevada. *World Environmental and Water Resources Congress 2009*, 1–15. https://doi.org/10.1061/41036(342)420

Huntington, J., Bromley, M., Morton, C. G., & Minor, T. (2018). *Remote Sensing Estimates of Evapotranspiration from Irrigated Agriculture, Northwestern Nevada and Northeastern California (DRI Publication No. 41275 prepared for the U.S. Bureau of Reclamation)*. https://s3-us-west-2.amazonaws.com/webfiles.dri.edu/Huntington/Huntington_et_al_2018_-_DRI_41275.pdf

Huntington, J., Gangopadhyay, S., Spears, M., Allen, R. G., King, D., Morton, C., Harrison, A., McEvoy, D., Joros, A., & Pruitt, T. (2015). *West-Wide Climate Risk Assessments: Irrigation Demand and Reservoir Evaporation Projections (Technical Memorandum No. 68-68210-2014-01)* (U.S. Bureau of Reclamation, Ed.). U.S. Bureau of Reclamation. https://www.usbr.gov/watersmart/baseline/docs/irrigationdemand/irrigationdemands.pdf

Huntington, J., Pearson, C., Minor, B., Volk, J., Morton, C., Melton, F., & Allen, R. (2022). *Upper Colorado River Basin OpenET Intercomparison Summary: Prepared for U.S. Bureau of Reclamation*. https://doi.org/10.13140/RG.2.2.21605.88808

Hurr, R. T., & Litke, D. W. (1989). *Estimating pumping time and ground-water withdrawals using energy- consumption data (U.S. Geological Survey Water-Resources Investigations Report 89-4107)*. https://doi.org/10.3133/wri894107

Jordahl, K., Bossche, J. Van den, Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., … Leblanc, F. (2020). *geopandas/geopandas: v0.8.1 (v0.8.1). Zenodo.* https://doi.org/10.5281/ZENODO.3946761

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In I. Guyon, U. V Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

Ketchum, D., Hoylman, Z. H., Huntington, J., Brinkerhoff, D., & Jencso, K. G. (2023). Irrigation intensification impacts sustainability of streamflow in the Western United States. Communications Earth & Environment, 4(1), 479. https://doi.org/10.1038/s43247-023-01152-2

Ketchum, D., Hoylman, Z. H., Brinkerhoff, D., Huntington, J., Maneta, M. P., Kimball, J., & Jencso, K. (2024). Irrigation response to drought in the Western United States, 1987–2021. JAWRA Journal of the American Water Resources Association. https://doi.org/10.1111/1752-1688.13190

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., van den Dool, H., Saha, S., Mendez, M. P., Becker, E., Peng, P., Tripp, P., Huang, J., DeWitt, D. G., Tippett, M. K., Barnston, A. G., Li, S., Rosati, A., Schubert, S. D., … Wood, E. F. (2014). The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. Bulletin of the American Meteorological Society, 95(4), 585–601. https://doi.org/10.1175/BAMS-D-12-00050.1

Kumar, R., Kumar, M., Farooq, Z., & Dadhich, S. M. (2017). Evaluation of models and approaches for effective rainfall in irrigated agriculture - An overview. *Journal of Soil and Water Conservation*, *16*(1), 32. https://doi.org/10.5958/2455-7145.2017.00012.1

Laipelt, L., Henrique Bloedow Kayser, R., Santos Fleischmann, A., 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. https://doi.org/10.1016/j.isprsjprs.2021.05.018

Lamb, S. E., Haacker, E. M. K., & Smidt, S. J. (2021). Influence of Irrigation Drivers Using Boosted Regression Trees: Kansas High Plains. *Water Resources Research*, *57*(5). https://doi.org/10.1029/2020WR028867

Levy, Z. F., Jurgens, B. C., Burow, K. R., Voss, S. A., Faulkner, K. E., Arroyo-Lopez, J. A., & Fram, M. S. (2021). Critical Aquifer Overdraft Accelerates Degradation of Groundwater Quality in California's Central Valley During Drought. *Geophysical Research Letters*, *48*(17). https://doi.org/10.1029/2021GL094398

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(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Little, K. E., Hayashi, M., & Liang, S. (2016). Community-Based Groundwater Monitoring Network Using a Citizen-Science Approach. *Groundwater*, *54*(3), 317–324. https://doi.org/10.1111/gwat.12336

Long, A. (2022). *Diamond Nat. Res. Prot. and Conservation Ass'n vs. Diamond Valley Ranch, 138 Nev. Adv. Op. 43 (June 16, 2022)*. Nevada Supreme Court Summaries. 1520. https://scholars.law.unlv.edu/nvscs/1520

Majumdar, S., Smith, R., Butler, J. J., & Lakshmi, V. (2020). Groundwater withdrawal prediction using integrated multitemporal remote sensing data sets and machine learning. *Water Resources Research*, *56*(11), e2020WR028059. https://doi.org/10.1029/2020WR028059

Majumdar, S., Smith, R., Conway, B. D., Butler, J. J., Lakshmi, V., & Dagli, C. H. (2021). Estimating Local-Scale Groundwater Withdrawals Using Integrated Remote Sensing Products and Deep Learning. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 4304–4307. https://doi.org/10.1109/IGARSS47720.2021.9554784

Majumdar, S., Smith, R., Conway, B. D., & Lakshmi, V. (2022). Advancing Remote Sensing and Machine Learning-Driven Frameworks for Groundwater Withdrawal Estimation in Arizona: Linking Land Subsidence to Groundwater Withdrawals. *Hydrological Processes*, *36*(11), e14757. https://doi.org/10.1002/hyp.14757

Majumdar, S., Smith, R. G., Hasan, M. F., Wilson, J. L., White, V. E., Bristow, E. L., Rigby, J. R., Kress, W. H., & Painter, J. A. (2024). Improving crop-specific groundwater use estimation in the Mississippi Alluvial Plain: Implications for integrated remote sensing and machine learning approaches in data-scarce regions. Journal of Hydrology: Regional Studies, 52, 101674. https://doi.org/10.1016/j.ejrh.2024.101674

Martin, D. J., Regan, R. S., Haynes, J. V., Read, A. L., Henson, W. R., Stewart, J. S., Brandt, J. T., & Niswonger, R. G. (2023). *Irrigation water use reanalysis for the 2000-20 period by HUC12, month, and year for the conterminous United States: U.S. Geological Survey data release*. https://doi.org/10.5066/P9YWR0OJ

Mefford, B., & Prairie, J. (2022). *Assessing Agricultural Consumptive Use in the Upper Colorado River Basin - Phase III Report. U.S. Bureau of Reclamation and the Upper Colorado River Commission*. http://www.ucrcommission.com/reports-studies/

Megdal, S. B., Gerlak, A. K., Varady, R. G., & Huang, L.-Y. (2015). Groundwater Governance in the United States: Common Priorities and Challenges. *Groundwater*, *53*(5), 677–684. https://doi.org/10.1111/gwat.12294

Melton, F., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., Erickson, T., Allen, R., Anderson, M., Fisher, J. B., Kilic, A., Senay, G. B., Volk, J., Hain, C., Johnson, L., Ruhoff, A., Blankenau, P., Bromley, M., Carrara, W., … Anderson, R. G. (2021). OpenET: Filling a Critical Data Gap in Water Management for the Western United States. *JAWRA Journal of the American Water Resources Association*. https://doi.org/10.1111/1752-1688.12956

Melton, F., 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., Votava, P., Trout, T. J., Temesgen, B., Frame, K., Sheffner, E. J., & Nemani, R. R. (2012). Satellite Irrigation Management Support With the Terrestrial Observation and Prediction System: A Framework for Integration of Satellite and Surface Observations to Support Improvements in Agricultural Water Resource Management. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*(6), 1709–1721. https://doi.org/10.1109/JSTARS.2012.2214474

Meza, I., Siebert, S., Döll, P., Kusche, J., Herbert, C., Eyshi Rezaei, E., Nouri, H., Gerdener, H., Popat, E., Frischen, J., Naumann, G., Vogt, J. V., Walz, Y., Sebesvari, Z., & Hagenlocher, M. (2020). Global-scale drought risk assessment for agricultural systems. *Natural Hazards and Earth System Sciences*, *20*(2), 695–712. https://doi.org/10.5194/nhess-20-695-2020

NASA JPL (2020). *NASADEM Merged DEM Global 1 arc second V001 [Data set]*. NASA EOSDIS Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MEaSUREs/NASADEM/NASADEM_HGT.001

NDWR. (1985). *AMARGOSA VALLEY BASIN #230 1985 GROUNDWATER PUMPAGE INVENTORY*. http://water.nv.gov/Pumpage%20Inventories/230%20-%20Amargosa%20Valley/230%20-%201985%20-%20Amargosa%20Valley.pdf

NDWR. (2020). *Diamond Valley Groundwater Management Plan*. http://water.nv.gov/Diamond%20Valley%20GMP/Diamond%20Valley%20GMP/Diamond%20Valley%20GMP--Final.pdf

NDWR. (2021). *GROUNDWATER APPROPRIATIONS & COMMITMENTS*. https://www.leg.state.nv.us/App/NELIS/REL/82nd2023/ExhibitDocument/OpenExhibitDocument?exhibitId=66244&fileDownloadName=DivisionofWaterResourcesAnswerstoCommitteeQuestions.pdf

NDWR. (2022). *AMARGOSA DESERT HYDROGRAPHIC BASIN 14-230 GROUNDWATER PUMPAGE INVENTORY, WATER YEAR 2021*. http://water.nv.gov/Pumpage%20Inventories/230%20-%20Amargosa%20Valley/230%20-%202021%20-%20Amargosa%20Valley.pdf

OpenET. (2023). *OpenET Methodologies*. https://openetdata.org/methodologies/

Owen, D., Cantor, A., Nylen, N. G., Harter, T., & Kiparsky, M. (2019). California groundwater management, science-policy interfaces, and the legacies of artificial legal distinctions. *Environmental Research Letters*, *14*(4), 045016. https://doi.org/10.1088/1748-9326/ab0751

OWRD. (2015). *Oregon Statewide Long-Term Water Demand Forecast*. http://www.oregon.gov/OWRD/

Patwardhan, A. S., Nieber, J. L., & Johns, E. L. (1990). Effective Rainfall Estimation Methods. *Journal of Irrigation and Drainage Engineering*, *116*(2), 182–193. https://doi.org/10.1061/(ASCE)0733-9437(1990)116:2(182)

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn:Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

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. https://doi.org/10.1016/j.agwat.2020.106197

Reilly, T. E., Dennehy, K. F., Alley, W. M., & Cunningham, W. L. (2008). *Ground-water availability in the United States (USGS Publications Warehouse)*. https://doi.org/10.3133/cir1323

Ruess, P. J., Konar, M., Wanders, N., & Bierkens, M. (2023). Irrigation by Crop in the Continental United States From 2008 to 2020. *Water Resources Research*, *59*(2). https://doi.org/10.1029/2022WR032804

Said, A., Stevens, D. K., & Sehlke, G. (2005). ESTIMATING WATER BUDGET IN A REGIONAL AQUIFER USING HSPF-MODFLOW INTEGRATED MODEL. *Journal of the American Water Resources Association*, *41*(1), 55–66. https://doi.org/10.1111/j.1752-1688.2005.tb03717.x

Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., & McMahon, P. B. (2012). Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. *Proceedings of the National Academy of Sciences*, *109*(24), 9320–9325. https://doi.org/10.1073/pnas.1200311109

Schmid, W. (2004). *A farm package for MODFLOW-2000 : simulation of irrigation demand and conjunctively managed surface-water and ground-water supply (PhD Dissertation)* [Department of Hydrology and Water Resources, The University of Arizona]. https://repository.arizona.edu/handle/10150/191271

Senay, G. B. (2018). Satellite Psychrometric Formulation of the Operational Simplified Surface Energy Balance (SSEBop) Model for Quantifying and Mapping Evapotranspiration. *Applied Engineering in Agriculture*, *34*(3), 555–566. https://doi.org/10.13031/aea.12614

Senay, G. B., Bohms, S., Singh, R. K., Gowda, P. H., Velpuri, N. M., Alemu, H., & Verdin, J. P. (2013). Operational Evapotranspiration Mapping Using Remote Sensing and Weather Datasets: A New Parameterization for the SSEB Approach. *JAWRA Journal of the American Water Resources Association*, *49*(3), 577–591. https://doi.org/10.1111/jawr.12057

Senay, G. B., Friedrichs, M., Morton, C., Parrish, G. E. L., Schauer, M., Khand, K., Kagone, S., Boiko, O., & Huntington, J. (2022). Mapping actual evapotranspiration using Landsat for the conterminous United States: Google Earth Engine implementation and assessment of the SSEBop model. *Remote Sensing of Environment*, *275*, 113011. https://doi.org/10.1016/j.rse.2022.113011

Sheppard, S. A., & Terveen, L. (2011). Quality is a verb. *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, 29–38. https://doi.org/10.1145/2038558.2038565

Smith, R., Knight, R., Chen, J., Reeves, J. A., Zebker, H. A., Farr, T., & Liu, Z. (2017). Estimating the permanent loss of groundwater storage in the southern San Joaquin Valley, California. *Water Resources Research*, *53*(3), 2133–2148. https://doi.org/10.1002/2016WR019861

Smith, R., Knight, R., & Fendorf, S. (2018). Overpumping leads to California groundwater arsenic threat. *Nature Communications*, *9*(1), 2089. https://doi.org/10.1038/s41467-018-04475-3

Smith, R., & Li, J. (2021). Modeling elastic and inelastic pumping-induced deformation with incomplete water level records in Parowan Valley, Utah. *Journal of Hydrology*, *601*, 126654. https://doi.org/10.1016/j.jhydrol.2021.126654

Smith, R., Li, J., Grote, K., & Butler, J. (2023). Estimating Aquifer System Storage Loss With Water Levels, Pumping and InSAR Data in the Parowan Valley, Utah. *Water Resources Research*, *59*(4). https://doi.org/10.1029/2022WR034095

Smith, R., & Majumdar, S. (2020). Groundwater storage loss associated with land subsidence in Western United States mapped using machine learning. *Water Resources Research*, *56*(7), e2019WR026621. https://doi.org/10.1029/2019WR026621

Stamm, G. G. (1967). *Problems and Procedures in Determining Water Supply Requirements for Irrigation Projects* (pp. 769–785). https://doi.org/10.2134/agronmonogr11.c45

Thompson, P.D. (1977). How to improve accuracy by combining independent forecasts. Mon. Weather Rev. 105, 228–229.

Thornton, P. E., Shrestha, R., Thornton, M., Kao, S.-C., Wei, Y., & Wilson, B. E. (2021). Gridded daily weather data for North America with comprehensive uncertainty quantification. *Scientific Data*, 8(1), 190. https://doi.org/10.1038/s41597-021-00973-0

USBR. (2016). *Historical and Future Irrigation Water Requirements for Select Reclamation Project Areas*. https://www.usbr.gov/watersmart/baseline/docs/historicalandfutureirrigationwaterrequirements.pdf

USBR. (2019). *ET-Demands*. https://github.com/usbr/et-demands

USBR. (2023). *WaterSMART Basin Study*. https://www.usbr.gov/watersmart/bsp/docs/BasinStudy_FactSheet_2023.pdf

USDA. (2023). *National Agriculture Imagery Program (NAIP)*. https://naip-usdaonline.hub.arcgis.com/

USDA Farm Service Agency. (2017). *Common Land Unit (CLU)*. https://www.fsa.usda.gov/Assets/USDA-FSA-Public/usdafiles/APFO/support-documents/pdfs/clu_infosheet_2017_Final.pdf

USDA NRCS. (2004). Chapter 9 Hydrologic Soil-Cover Complexes. In *Part 630 Hydrology National Engineering Handbook*. https://directives.sc.egov.usda.gov/OpenNonWebContent.aspx?content=17758.wba

USDA SCS. (1993). Chapter 2 Irrigation Water Requirements. In *Part 623 National Engineering Handbook*. USDA Soil Conservation Service. https://www.wcc.nrcs.usda.gov/ftpref/wntsc/waterMgt/irrigation/NEH15/ch2.pdf

USGS. (2023). *1 Arc-second Digital Elevation Models (DEMs) - USGS National Map 3DEP Downloadable Data Collection: U.S. Geological Survey.* https://www.sciencebase.gov/catalog/item/4f70aa71e4b058caae3f8de1

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace. https://dl.acm.org/doi/book/10.5555/1593511

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., Morton, C., Ott, T., Johnson, L., Comini de Andrade, B., Carrara, W., Doherty, C. T., Dunkerly, C., Friedrichs, M., Guzman, A., … Yang, Y. (2024). Assessing the accuracy of OpenET satellite-based evapotranspiration data to support water resource and land management applications. Nature Water. https://doi.org/10.1038/s44221-023-00181-7

Walkinshaw, M., O'Geen, A. T., & Beaudette, D. E. (2022). "Soil Properties." California Soil Resource Lab. casoilresource.lawr.ucdavis.edu/soil-properties/.

Wei, S., Xu, T., Niu, G.-Y., & Zeng, R. (2022). Estimating Irrigation Water Consumption Using Machine Learning and Remote Sensing Data in Kansas High Plains. *Remote Sensing*, *14*(13), 3004. https://doi.org/10.3390/rs14133004

Wilson, J. L. (2021). Aquaculture and Irrigation Water-Use Model (AIWUM) version 1.0—An agricultural water-use model developed for the Mississippi Alluvial Plain, 1999–2017. In *Scientific Investigations Report*. U.S. Geological Survey. https://doi.org/10.3133/sir20215011

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., & Mocko, D. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, *117*(D3). https://doi.org/10.1029/2011JD016048

Zektser, S., Loaiciga, H. A., & Wolf, J. T. (2005). Environmental impacts of groundwater overdraft: selected case studies in the southwestern United States. Environmental Geology, 47(3), 396–404. https://doi.org/10.1007/s00254-004-1164-3

Zipper, S. C., Farmer, W. H., Brookfield, A., Ajami, H., Reeves, H. W., Wardropper, C., Hammond, J. C., Gleeson, T., & Deines, J. M. (2022). Quantifying Streamflow Depletion from Groundwater Pumping: A Practical Review of Past and Emerging Approaches for Water Management. JAWRA Journal of the American Water Resources Association, 58(2), 289–312. https://doi.org/10.1111/1752-1688.12998

Supplementary Information for

Toward Sustainable Groundwater Management: Harnessing Remote Sensing and Climate Data to Estimate Field-Scale Groundwater Pumping

Thomas J. Ott [a*], Sayantan Majumdar [a*†], Justin L. Huntington [a†], Christopher Pearson [a], Matt Bromley [a], Blake A. Minor [a], Charles G. Morton [a], Sachiko Sueki [b], Jordan P. Beamer [c], Richard Jasoni [a]

[a] Desert Research Institute, Reno, NV, USA
[b] Desert Research Institute, Las Vegas, NV, USA
[c] Oregon Water Resources Department, Salem, OR, USA

This supplementary information file has five figures and two tables referenced in the main manuscript.

Supplementary Table 1. Ensemble machine learning (ML) models and the hyperparameters tuned in a randomized grid search with five-fold cross-validation. The random seed value is set to 1234 throughout, and the root mean square error (RMSE) is used as the objective function across these models. ERT and RF are available from scikit-learn, and GBT is available from LightGBM.

Model	Hyperparameter values	Tuned hyperparameters
Extremely Randomized Trees (ERT)	'n_estimators': [300, 400, 500, 800] 'max_features': [5, 6, 7, 10, 12, 20, 30, None] 'max_depth': [8, 15, 20, 6, 10, None] 'min_samples_leaf': [1, 2] 'max_samples': [None, 0.9] 'max_leaf_nodes': [16, 20, 31, 32, 63, 127, 15, 255, 7, None] 'min_samples_split': [2, 3, 4, 0.01] *Fixed parameters*: bootstrap=True	'n_estimators': 800 'max_features': None 'max_depth': 10 'min_samples_leaf': 2 'max_samples': None 'max_leaf_nodes': 127 'min_samples_split': 2
Gradient Boosting Machine (GBT)	'n_estimators': [300, 400, 500, 800] 'max_depth': [8, 15, 20, 6, 10, -1] 'learning_rate': [0.01, 0.005, 0.05, 0.1] 'subsample': [1, 0.9, 0.8] 'colsample_bytree': [1, 0.9] 'colsample_bynode': [1, 0.9] 'path_smooth': [0, 0.1, 0.2] 'num_leaves': [16, 20, 31, 32, 63, 127, 15, 255, 7] 'min_child_samples': [30, 40, 10, 20] *Fixed parameters*: tree_learner='feature', deterministic=True, force_row_wise=True	'n_estimators': 800 'max_depth': 8 'learning_rate': 0.01 'subsample': 0.8 'colsample_bytree': 0.9 'colsample_bynode': 1 'path_smooth': 0.2 'num_leaves': 7 'min_child_samples': 20
Random Forests (RF)	Same as ERT	'n_estimators': 500 'max_features': 20 'max_depth': 6 'min_samples_leaf': 2 'max_samples': None 'max_leaf_nodes': 16 'min_samples_split': 4

* These two authors have contributed equally.
† Corresponding author at: Division of Hydrologic Sciences, Desert Research Institute, 2215 Raggio Parkway, Reno, Nevada 89512-1095, USA
Email addresses: sayantan.majumdar@dri.edu (S. Majumdar), justin.huntington@dri.edu (J.L. Huntington)

Supplementary Table 2. Description of the 28 predictors used in the full machine learning models to estimate groundwater pumping depths in Diamond Valley, Nevada[‡]. The data references are in the main manuscript.

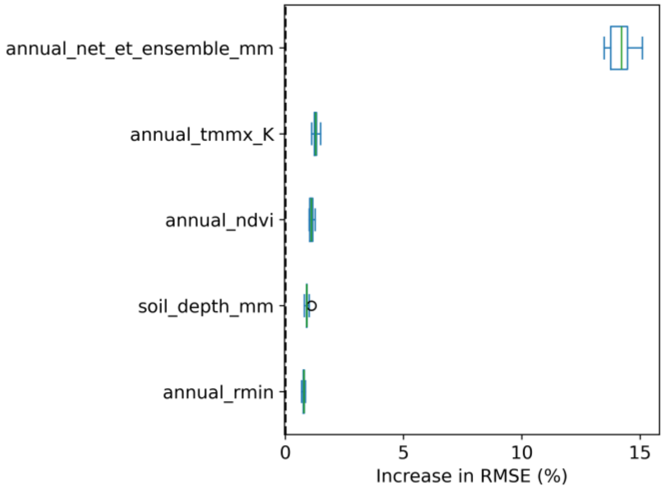
Predictor name	Description	Operations
annual_net_et_ensemble_mm[§]	OpenET ensemble-based $Net\ ET$ in mm	annual_et_ensemble_mm - annual_gridmet_precip_eff_mm
annual_et_eemetric_mm	eeMETRIC actual ET in mm	Temporal sum (calendar year) and zonal mean
annual_et_ssebop_mm	SSEBop actual ET in mm	Temporal sum (calendar year) and zonal mean
annual_et_geesebal_mm	geeSEBAL actual ET in mm	Temporal sum (calendar year) and zonal mean
annual_et_ensemble_mm	OpenET ensemble actual ET in mm	Temporal sum (calendar year) and zonal mean
annual_daymet_precip_eff_mm	Daymet v4 effective precipitation in mm	annual_daymet_precip_mm * eff_factor
annual_daymet_precip_mm	Daymet v4 precipitation in mm	Temporal sum (calendar year) and zonal mean
annual_et_disalexi_mm	ALEXI/DisALEXI actual ET in mm	Temporal sum (calendar year) and zonal mean
annual_gridmet_precip_mm	gridMET precipitation in mm	Temporal sum (calendar year) and zonal mean
annual_gridmet_precip_eff_mm	gridMET effective precipitation in mm	annual_gridmet_precip_mm * eff_factor
annual_et_pt_jpl_mm	PT-JPL actual ET in mm	Temporal sum (calendar year) and zonal mean
annual_et_sims_mm	SIMS actual ET in mm	Temporal sum (calendar year) and zonal mean
annual_ndvi	Landsat-8 32-day composite NDVI	Temporal max (calendar year) and zonal mean
annual_rmin	gridMET minimum relative humidity %	Temporal median (calendar year) and zonal mean

[‡] Here, the temporal operations are performed for each year between 2018 and 2022, and the zonal operations are performed for each field. If a well waters multiple fields, then we sum up the corresponding actual ET, reference ET, Net ET, precipitation, effective precipitation, effective precipitation factor, and vapor pressure deficit for those fields, average the NDVI, minimum relative humidity, maximum relative humidity, minimum air temperature, maximum air temperature, soil depth, saturated hydraulic conductivity, and wind velocity, and take the mode of the hydrologic soil groups for those fields.

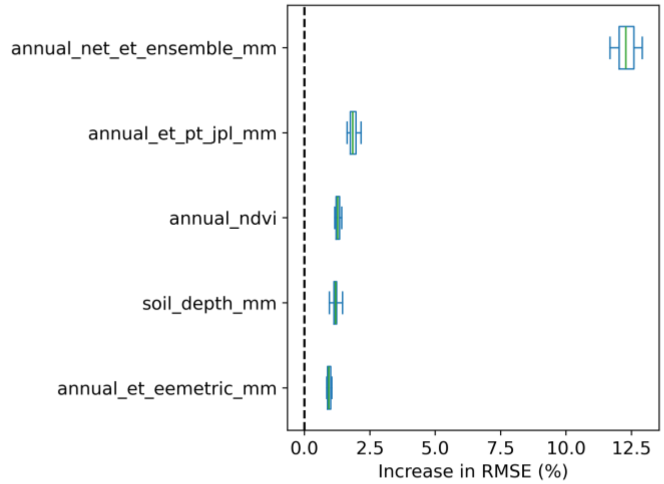
[§] For the ML models used to compare the ET model performances (Table 5, main manuscript), we replace the annual_net_et_ensemble_mm with the corresponding Net ET (e.g., annual_net_et_eemetric_mm) and only keep the corresponding actual ET predictor, e.g, annual_et_eemetric_mm. Other ET predictors are removed to negate the correlation effects. All the remaining predictors are kept as in the full ML model. Therefore, we end up with 22 predictors for each of the models in Table 5 of the main manuscript.

Supplementary Table 2 (Contd.). Description of the 28 predictors used in the full machine learning models to estimate groundwater pumping depths in Diamond Valley, Nevada. Here, the temporal operations are performed for each year between 2018 and 2022, and the zonal operations are performed for each field.

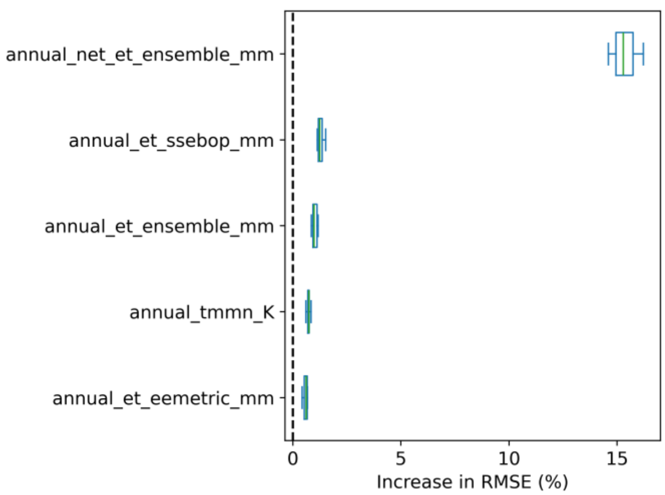
Predictor name	Description	Operations
annual_rmax	gridMET maximum relative humidity %	Temporal median (calendar year) and zonal mean
ksat_mean_micromps	Saturated hydraulic conductivity in μm/s	Zonal mean
soil_depth_mm	Soil depth in mm	Zonal mean
annual_vpd_kPa	gridMET vapor pressure deficit in kPa	Temporal sum (calendar year) and zonal mean
annual_tmmn_K	gridMET minimum air temperature (K)	Temporal median (calendar year) and zonal mean
annual_tmmx_K	gridMET maximum air temperature (K)	Temporal median (calendar year) and zonal mean
eff_factor	ET-Demands-derived basin-scale effective precipitation factor	
elevation_m	NASADEM elevation in m	Zonal mean
annual_vs_mps	gridMET wind velocity in m/s	Temporal mean (calendar year) and zonal mean
annual_etr_mm	gridMET alfalfa reference ET in mm	Temporal sum (calendar year) and zonal mean
annual_eto_mm	gridMET grass reference ET in mm	Temporal sum (calendar year) and zonal mean
HSG_1	Hydrologic soil group 1 (A)	Zonal mode
HSG_3	Hydrologic soil group 3 (B)	Zonal mode
HSG_5	Hydrologic soil group 5 (C)	Zonal mode



(a)

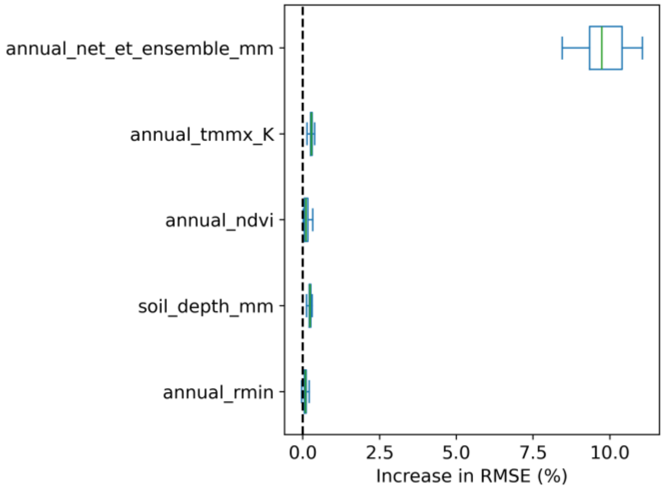


(b)

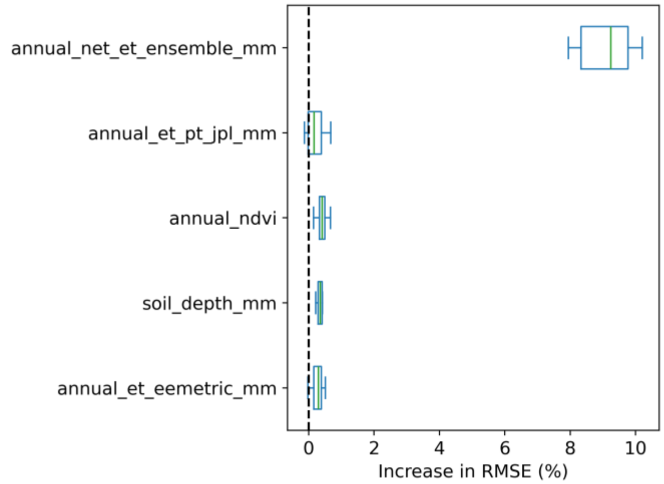


(c)

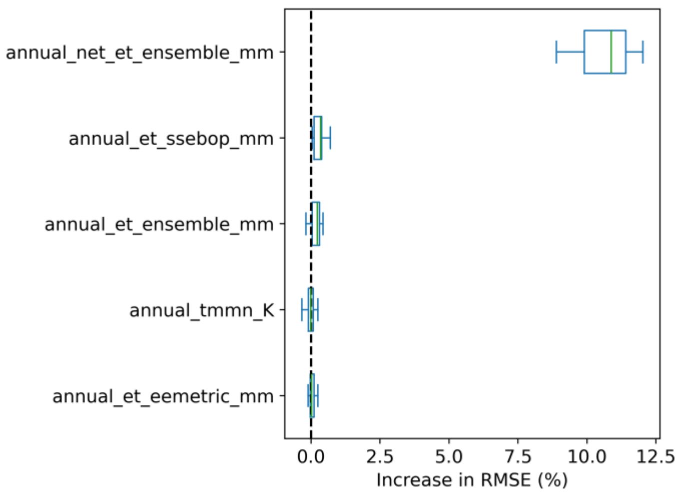
Supplementary Figure 1. Permutation importance plots showing the top five features for the training data (including validation) for (a) ERT, (b) GBM, and (c) RF.



(a)

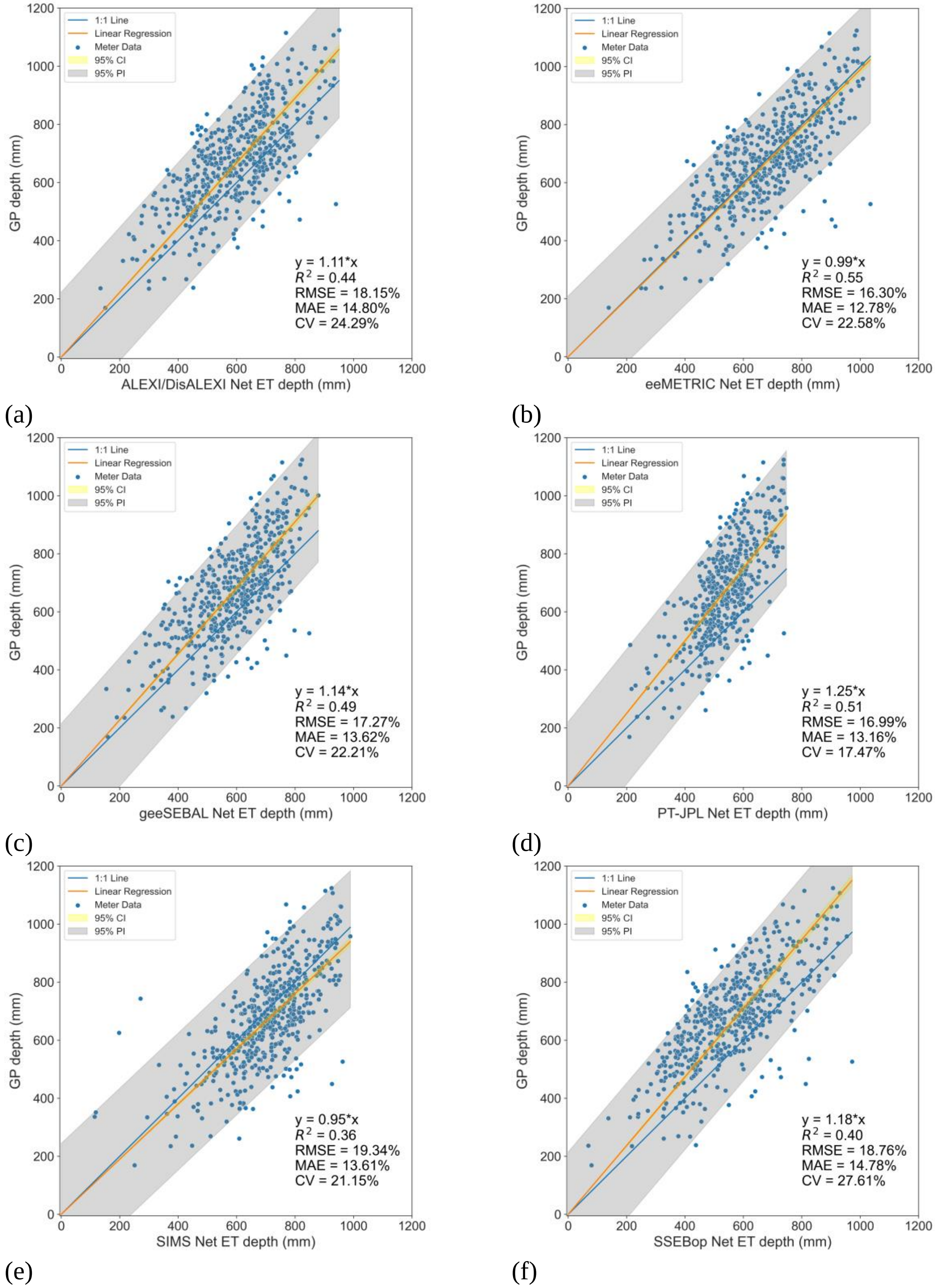


(b)

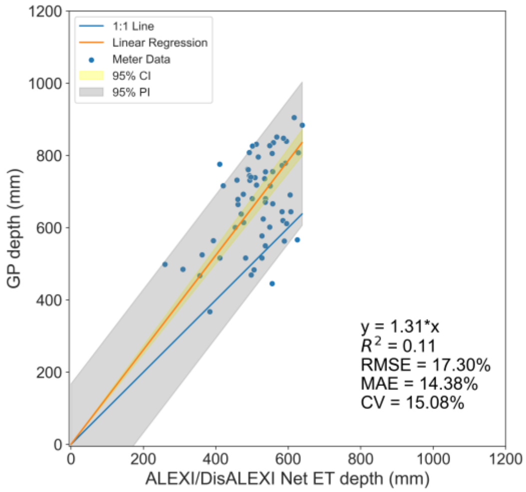


(c)

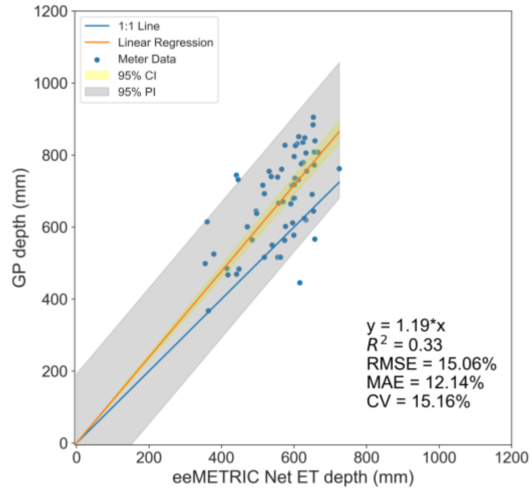
Supplementary Figure 2. Permutation importance plots showing the top five features for the test data for (a) ERT, (b) GBM, and (c) RF.



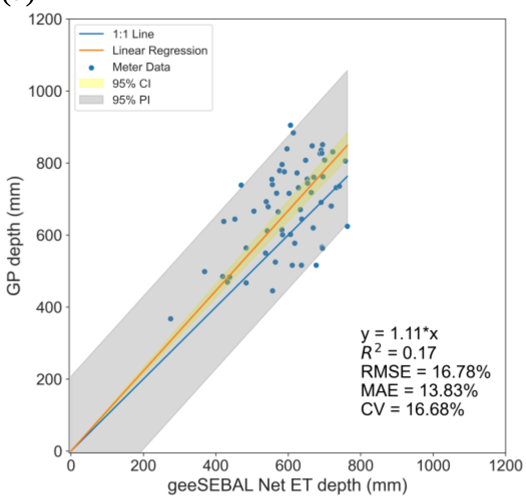
Supplementary Figure 3. Scatter plots of the linear regression models for (a) ALEXI/DisALEXI, (b) eeMETRIC, (c) geeSEBAL, (d) PT-JPL, (e) SIMS, and (f) SSEBop in DV, Nevada. The symbols and labels are defined in the main manuscript. The scatter plot of the OpenET ensemble is shown in Figure 7 (a) of the main manuscript.



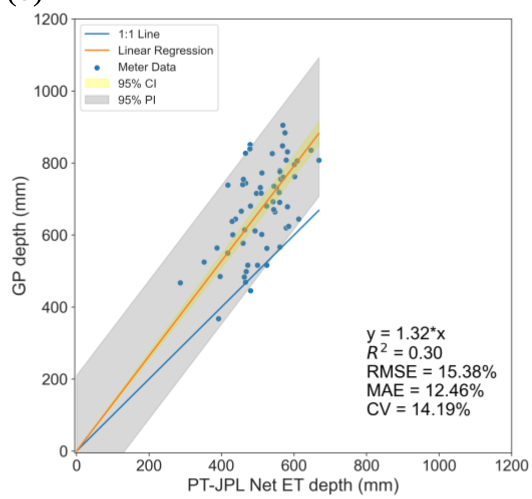
(a)



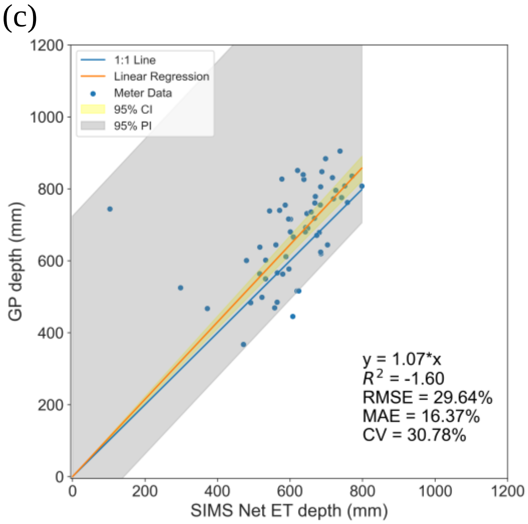
(b)



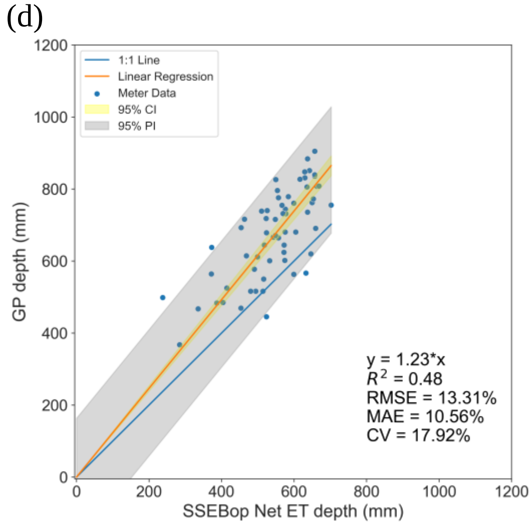
(c)



(d)

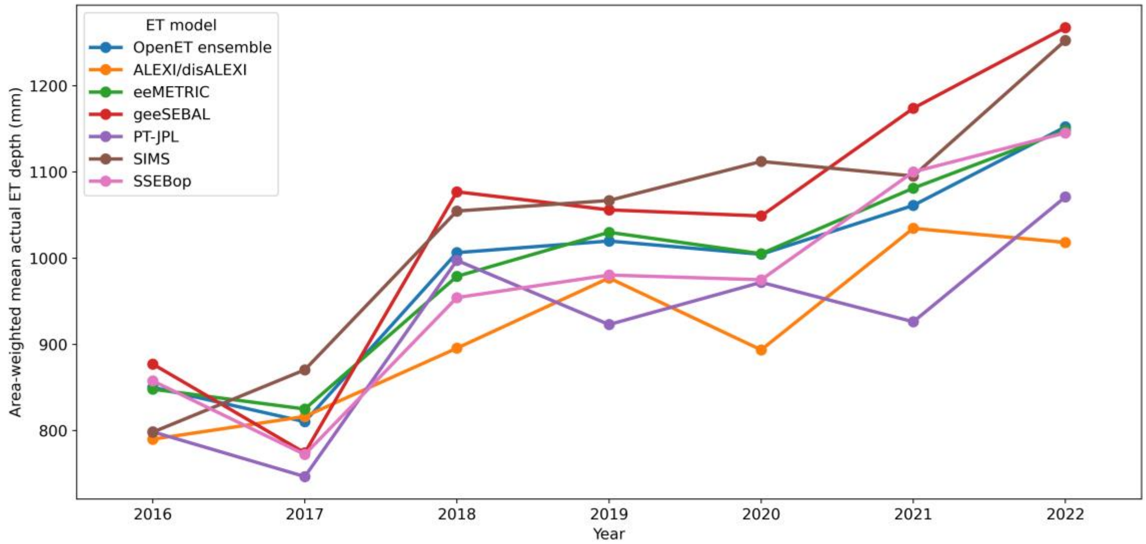


(e)

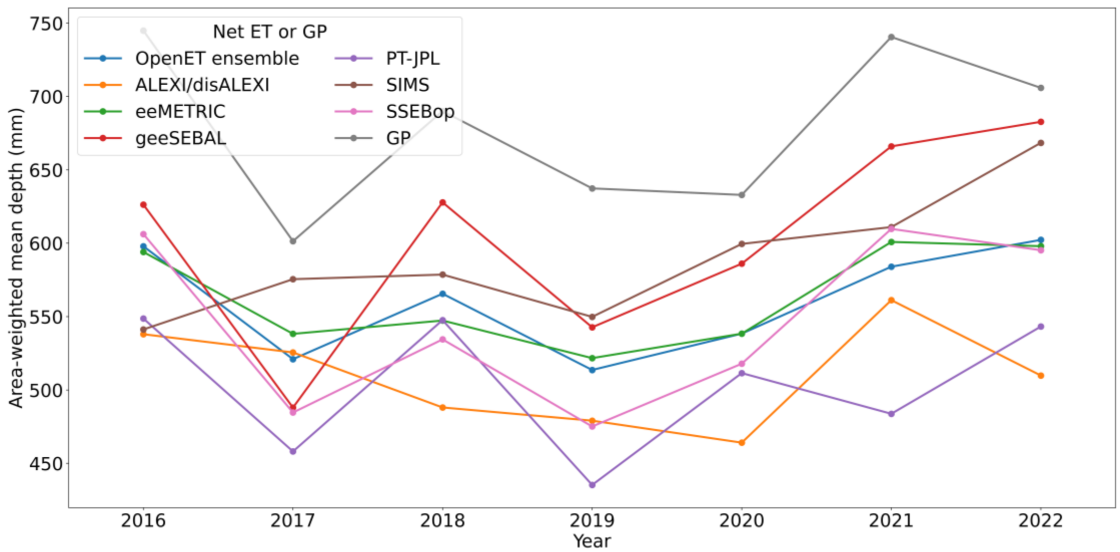


(f)

Supplementary Figure 4. Scatter plots of the linear regression models for (a) ALEXI/DisALEXI, (b) eeMETRIC, (c) geeSEBAL, (d) PT-JPL, (e) SIMS, and (f) SSEBop in HB, Oregon. The symbols and labels are defined in the main manuscript. The scatter plot of the OpenET ensemble is shown in Figure 10 (a) of the main manuscript.



(a)



(b)

Supplementary Figure 5. Comparisons of the area-weighted mean annual (a) ET depths and (b) Net ET and metered GP depths for each ET model in HB, Oregon. Note that these area-weighted means in (a) and (b) are calculated after the outlier removal process described in Section 3.3 (Figure 6) of the main manuscript.