2023-12-01

Evaluation of evapotranspiration estimates using an existing hybrid machine learning model in a natural and a managed dryland site

Katya Esquivel Herrera
University of Texas at El Paso

EVALUATION OF EVAPOTRANSPIRATION ESTIMATES USING AN EXISTING HYBRID MACHINE LEARNING MODEL IN A NATURAL AND A MANAGED DRYLAND SITE

KATYA ESQUIVEL HERRERA

Master's Program in Environmental Sciences

APPROVED:

Marguerite Mauritz, Ph.D., Chair

Lixin Jin, Ph.D.

Hugo Gutierrez Jurado, Ph.D.

Stephen L. Crites, Jr., Ph.D.

Dean of the Graduate School

Dedication

To Gibran, whose love and support kept me motivated when nothing else would.

Thank you for holding my hand through it all.

EVALUATION OF EVAPOTRANSPIRATION ESTIMATES USING AN EXISTING HYBRID MACHINE LEARNING MODEL IN A NATURAL AND A MANAGED DRYLAND SITE

by

KATYA ESQUIVEL HERRERA, B.S.

THESIS

Presented to the Faculty of the Graduate School of
The University of Texas at El Paso
in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE

Department of Earth, Environmental, and Resource Sciences
THE UNIVERSITY OF TEXAS AT EL PASO
December 2023

Acknowledgements

I would like to acknowledge and express my sincere gratitude to my advisor, Dr. Marguerite Mauritz, her guidance and patience were instrumental in completing this work. Thank you for always being willing to hear and discuss new ideas.

I am also grateful with Dr. Lixin Jin and Dr. Hugo Gutierrez for all their support, thoughtful feedback and interest in helping me become a better scientist.

I would like to thank the Consejo Nacional de Humanidades, Ciencia y Tecnologia in Mexico for their support through their foreign scholarships program.

This material is based upon work supported by the National Science Foundation - Division Of Chemical Bioengineering Environmental Transport Systems under award #2034312.

Table of Contents

List of Tables

List of Figures

Introduction

Definition and importance of evapotranspiration

One of the largest components of the hydrologic cycle is evapotranspiration (ET), which represents the sum of two processes: the evaporative loss of water from ground surfaces, plants, or water bodies, and the evaporative loss of water through stomata, also called transpiration (Irmak, 2008). The main environmental factors controlling ET as a whole are water availability, energy availability, and the capacity of the atmosphere to take up water (Alfieri et al., 2020) which translate into multiple meteorological variables such as soil water content, air temperature and relative humidity, just to name a few. Additional controls are exerted on the transpiration portion of ET, coming from the plant physiological responses, such as photosynthesis and stomatal conductance, to environmental factors like the availability of radiative energy, changes in vapor pressure deficit, carbon dioxide levels and temperature (Alfieri et al., 2020; Kannenberg et al., 2022).

Speaking strictly in terms of energy fluxes, evapotranspiration is the latent heat flux (LE) associated with the change of phase in water, leading to the transfer of energy from a surface to the atmosphere (Carter and Liang, 2019), with energy being provided by the solar radiation and the sensible heat flux, and vapor pressure deficit along with wind speed enhancing the diffusion of water vapor in the atmosphere (Granata, 2019). Therefore, ET is a unique climate variable that links the water cycle through evaporation, the energy cycle through latent heat flux, and the carbon cycle through transpiration-photosynthesis (Fisher et al., 2017; Zhao et al., 2019) while also being strongly correlated to factors such as root density, leaf area, and soil water retention (Biederman et al., 2017).

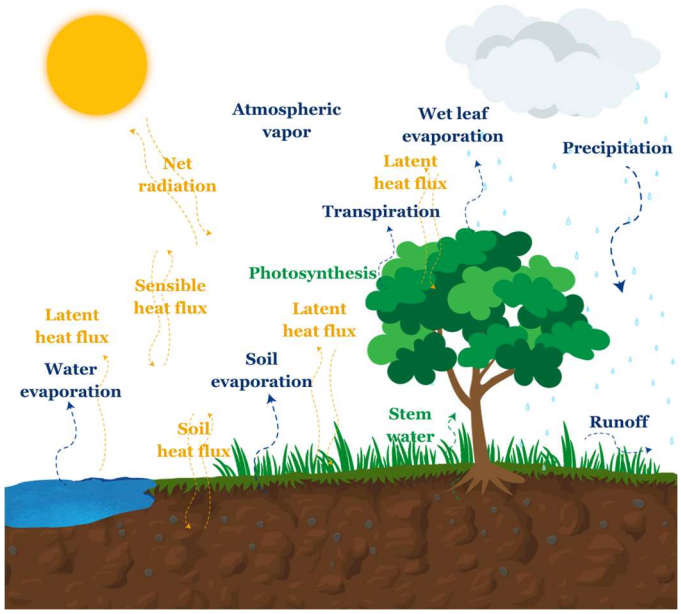

Figure 1. Schematic diagram of the interrelation between the water cycle (blue), the energy balance (yellow) and the carbon balance (green).

The interrelation of ET with multiple biogeochemical cycles makes it a better proxy for estimating ecosystems' water availability than precipitation. Evidence of this is provided in a study by Biederman et al. (2017) which showed that the correlation between annual anomalies in ET and ecosystem productivity was stronger than that between precipitation and productivity, where the anomalies were identified as years with values more than one standard deviation away from the mean. Therefore, quantification of ET is fundamental for understanding an ecosystem's ecohydrological dynamics.

In agroecosystems, ET quantification becomes especially relevant for improving water use efficiency and determining water allocation or irrigation management (Alfieri et al., 2020; Irmak, 2008). Accurate ET estimates are needed to optimize crop yield and improve the effectiveness of water resources management, especially as the demand for limited water resources grows (Alfieri et al., 2020). However, the motivation for getting accurate ET estimates goes beyond this. In 2012,

most drought monitoring tools, which look at drought indicators such as precipitation variation, temperature, soil moisture, water levels, and vegetation stress, failed to predict the magnitude and intensity of the U.S. Midwest mega-drought (Fisher et al., 2017). This failure has been attributed to these drought proxies not integrating enough information about the land-atmosphere coupling and the response of the vegetation, as the metrics that did capture the early warnings were based on ET, such as the Evaporative Stress Index (Fisher et al., 2017; Meng et al., 2014).

ET measurements and modeling

ET can be estimated directly with on-site measurements or indirectly using the energy balance residual. Direct or on-site measurements encompass techniques that consider the water inputs and outputs of the system (i.e., lysimeters) and are based on the principle of the conservation of mass and estimate ET by measuring changes in the amount of soil water. Micrometeorological techniques based on the Monin–Obukhov similarity theory and conservation of mass (i.e., eddy covariance) are also considered on-site measurements (Alfieri et al., 2020; Biederman et al., 2017; Fisher et al., 2017; Zhao et al., 2019). Indirect estimations on the other hand, rely on accounting for the remaining terms of the water and energy balance equations using meteorological data and other relevant climatic parameters to solve for equations like the Penman-Monteith, or on finding empirical relationships between variables to develop statistical models (Alfieri et al., 2020; Biederman et al., 2017; Fisher et al., 2017; Zhao et al., 2019).

Both types of methods present disadvantages. A key disadvantage of direct measurements is their spatial and temporal coverage being limited to a constrained footprint (Alfieri et al., 2020; Carter and Liang, 2019) which translates to sparse availability of these measurements. To fill these data gaps in measurements, indirect measurements are then used. A common example is the use of remote sensing products to create models for estimating ET indirectly, particularly at larger

spatial scales (Biederman et al., 2017; Elfarkh et al., 2022; Zhang et al., 2019). However, the use of these type of products leads to one of the primary disadvantages of indirect measurements, given their temporal discontinuity and coarse spatial resolution, the product models tend to either over or underestimate (Alfieri et al., 2020).

The models used for obtaining indirect ET measurements can be grouped into two broad categories: Physics-based and statistical-based (Granata, 2019; Hu et al., 2021; Zhang et al., 2019; Zhao et al., 2019). Physics-based models use explicit physical representations and can conserve energy but need to empirically estimate some parameters like surface roughness length and surface resistance. These models include surface energy-balance models, vegetation index-land surface temperature triangle/trapezoidal models, and Penman-Monteith or Priestley-Taylor models. On the other hand, statistical-based models use fewer physical mechanisms and, instead, ET is directly estimated using statistical correlations of ET with meteorological variables (empirical and semi-empirical models) or forcing data using machine learning algorithms (ML-based).

The statistical models commonly derive ET as a physical energy variable, LE. However, deriving it this way leads to requiring additional measurements to adequately capture the abiotic and biotic controls (Fisher et al., 2017). These additional measurements include meteorological variables that control the transfer of water from the land to the atmosphere (i.e. solar radiation, humidity, air temperature, wind speed, and soil moisture), data about phenology and changes in vegetation cover to provide information about the dynamics and magnitudes of ET fluxes, as well as land surface temperature information which relates directly to the evaporative flux (Fisher et al., 2017; Zhao et al., 2019). Therefore, a fundamental part of the statistical models includes the use of ground-based observations that complement and validate the use of remotely sensed data.

Machine learning models for ET estimation

In recent years, remote sensing techniques combined with machine learning (ML) algorithms have become another approach for estimating ET, as ML models allow for the analysis of the complex climatic conditions near the surface as well as the dynamic variability of water and heat transfer processes which can reflect complex patterns not easily captured with traditional models (Amani and Shafizadeh-Moghadam, 2023; Dou and Yang, 2018; Granata, 2019). However, while these models are easier to apply, they can behave poorly outside the range of the data they were validated with and usually do not conserve the surface energy budget (Zhao et al., 2019). In addition, selecting the most suitable ML algorithm for estimating ET can vary greatly depending on factors such as the study area, data availability, and the approach taken for ET estimation. Recent studies (Carter and Liang, 2019; Chen et al., 2020; Dou and Yang, 2018; Granata, 2019; Hu et al., 2021; Zhao et al., 2019) have systematically compared a range of ML algorithms and different sets of predictors to estimate LE from global remote sensing data and ground-based flux tower data. Overall, satisfactory results have been obtained from different methods with only marginal differences in performance (Carter and Liang, 2019; Dou and Yang, 2018) but wide variation in algorithm efficiency (Carter and Liang, 2019; Granata, 2019). In the mentioned literature cases, the performance of models varied mainly between individual ecosystem types (Carter and Liang, 2019; Dou and Yang, 2018), but training the models with data from a single ecosystem led to worst performance compared to using data from all sites (Carter and Liang, 2019).

So far, the models described are based strictly on ML, or ‘pure models’. These present the disadvantage of lacking physical constraints, such as energy balance closure and very limited physical interpretability (Reichstein et al., 2019). To account for these limitations, a new approach

has been developed in the form of hybrid ML models which aim to integrate the theoretical foundations and constraints of the physical processes with the prediction power of traditional ML models. With these models, physical consistency can be preserved and predictive accuracy can be improved while maintaining interpretability, thus making it possible to take better advantage of the data adaptiveness that ML models provide (Reichstein et al., 2019; Zhao et al., 2019). A hybrid ML model (Figure 2) is achieved by improving parameterizations of ML using physical models. With this approach, the patterns that are not represented in the physical model can be identified and used to correct the model bias caused by dynamic variables and facilitate downscaling to finer spatial scales, which conceptually translates to making ML models more physically realistic (Reichstein et al., 2019). Multiple hybrid ML models for ET estimation have been developed in the past few years (Chen et al., 2022; Hu et al., 2021; Zhao et al., 2019) combining flux tower data along with different remote sensing products (i.e. optical vegetation indices). In each of these cases, the developed hybrid model is compared to an equivalent 'pure' model, with results predominantly showing that while pure models tend to have the best accuracy, hybrid models have at least similar or overall improved performance while keeping the physical constraints. The model presented by Zhao et al. (2019) is of particular interest as the performance between the pure and hybrid model remains similar, while the results also indicate that it has a high capability for predicting extreme events.

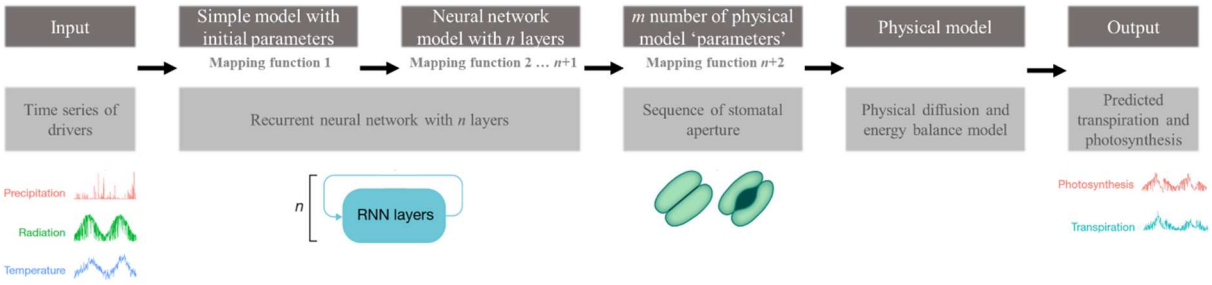

Figure 2. Diagram of hybrid machine learning modelling by addition of a physical layer to a multilayer neural network (dark grey) with an example of how it translates to an ecological problem like prediction of transpiration and photosynthesis (light grey) (modified from Reichstein et al., 2019).

Water limitation in drylands and the importance of ET estimation

Drylands are characterized by their limited water availability and, as a result of this, they are highly vulnerable to changes in species abundance, soils, or ecosystem processes in response to alterations in environmental conditions such as those driven by climate change (Bestelmeyer et al., 2015; Reynolds et al., 2007; Smith et al., 2019). Given their large extent of over 41% of the terrestrial surface, 39% of the global population is dependent on dryland ecosystems and their resources despite their vulnerability to climate and land use change (Plaza et al., 2018; Reynolds et al., 2007). Approximately 11% of the global dryland area is used as cropland and 30% as pasture, meaning that 50% and 74% of global croplands and pastures are located in drylands (Plaza et al., 2018). Aside from the importance of drylands for the human population and economy, recent studies have shown that semi-arid ecosystems dominate the trend and interannual variability of the land carbon dioxide sink (Ahlström et al., 2015).

Water is a primary limiting factor in drylands, making the spatial and temporal availability of water a key driver of every vegetation related process, leading to a strong connection between ecosystem ecology and hydrology in these regions (Smith et al., 2019). Although water is far from being the only control on dryland processes (Austin, 2011), it still plays a critical role in its

biogeochemical cycles. Dryland systems generally go through brief but extreme precipitation events leading to a short window of water availability and pulses of biological activity, this is traditionally called pulse dynamics (Knapp et al., 2008). Many ecological processes in drylands are driven by the size, frequency, and intensity of these water pulses, which can be highly variable between years (Collins et al., 2014; Petrie et al., 2019). Global climate change models predict that many arid regions will get warmer and drier in the future, intensifying the already shifting precipitation regime (Collins et al., 2014). Considering this and the significant role croplands and pastures have in drylands at a global level, there is a marked need to improve water use efficiency in these regions, which involves improving irrigation practices in agriculture (Ramirez-Valle, 2022).

Gaps in knowledge of dryland ecosystem dynamics remain, despite the increased development of data networks in the last years, especially when compared to more mesic ecosystems (Biederman et al., 2017). The main reason for this is the drylands' highly dynamic structure and function, the intermittent and unpredictable water availability and the chronic underrepresentation of long-term, continuous field measurements (Smith et al., 2019). In addition to this, the lack of long-term measurements of water exchange is further accentuated in drylands compared to more mesic ecosystems, leading to a heavy reliance on models for estimating ET in them (Biederman et al., 2017). However, the scale at which these models are created, along with the use of datasets with low representativeness of dryland ecosystems, leads to poorly constrained models and inaccurate estimates of ET (Biederman et al., 2017; Zhang et al., 2019), unless the model is explicitly constrained for the conditions of the site, which limits its application in areas with less available data (Dhungel et al., 2023; Elfarkh et al., 2022). For example, a study evaluated widely used remote sensing-based estimates of gross primary productivity and ET across dryland

sites of the United States Southwest and Mexico, results showed that they captured only 31% and 29% of the interannual variability of site measured GPP and ET (Biederman et al., 2017). This mismatch has been primarily attributed to specific challenges with using drylands remote sensed data due to the difficulty of capturing their land surface phenology. Namely, low vegetation coverage leads to high signal-to-ground noise ratios, the presence of evergreen vegetation shows as low seasonal vegetation index variability, phenology is highly dependent on climatic conditions leading to highly irregular growing seasons, and the overall presentation of high spatial heterogeneity (Smith et al., 2019; Taylor et al., 2021).

A more recent effort to improve model prediction of carbon and water fluxes in drylands was the one made by Dannenberg et al. (2023), who developed an artificial neural networks model to predict dryland ecosystem fluxes by combining optical vegetation indices, multitemporal thermal observations, and microwave soil moisture and temperature retrievals. This model was able to improve the ET predictions for 23 out of 26 sites with varying degree of improvement (R^2 between 0.4 and 0.9), but still underestimated the magnitude of interannual variability in ET similar to MODIS models. The model used in this case is considered a 'pure' ML model, where there is no physical constraints, which limits the ability to preserve the physical connections between the predicted variables such as GPP and ET. This study also mentions that the model performance was remarkably worse for sites with taller vegetation in the case of GPP and NEE predictions, although it is not mentioned if this is the case as well for ET predictions.

Therefore, the challenges in quantifying ET using remote sensed data might be further accentuated for agricultural dryland sites, where management introduces abrupt changes in land use, physical soil properties, and water availability. The agricultural fields along the Rio Grande in Southwest Texas and Southern New Mexico present an example of challenges that might be

present in a dryland agricultural area. In this region, natural sites along the river valley have been converted into managed agricultural sites for over 100 years, which are flood-irrigated using diverted river water and pumped groundwater (Miyamoto et al., 1995; Ortiz et al., 2022). The distribution of these agricultural sites is often near urban areas, adding to the effect of land use changes and anthropogenic inputs that might be absent in less managed sites.

At the time this work was developed and to the best of the author's knowledge, no hybrid machine learning models has been developed or tested for drylands. Therefore, the focus of this thesis will be to investigate the predictive accuracy of Zhao et al. (2019) hybrid model for two dryland sites. This hybrid model will be described in detail on the methodology section.

Driving questions and goals

Considering the challenges present in quantifying ET on drylands, the aim of this project is to evaluate how accurately an established globally trained hybrid ML model integrating remote sensing data and flux tower data can predict the evapotranspiration of drylands. For this purpose, the model developed by Zhao et al. (2019) will be tested using ground measurements of two sites located within drylands, one natural ecosystem and one agricultural field. By applying a hybrid ML model using the parameters described in Zhao et al. (2019) and evaluating it against these sites, the present work seeks to address the following questions:

- Is the generalizability of a hybrid ML model trained on global flux tower data enough for accurately estimating drylands' ET?
- Does the applicability of the model change between a natural and an irrigated dryland system?
- Finally, how high is the potential of the model for informing management decisions?

I hypothesize that the generalizability of the hybrid ML model is likely to have limited accuracy for estimating dryland's ET, with applicability varying by system, since the assumptions it is based on may not be entirely valid for dryland systems. This is expected to lead to a low potential for informing management decisions. The factors expected to contribute to the limited accuracy along with the ecosystem variability include the short distance changes in land use for the agricultural site and the highly dynamic structure and function of the natural site. While the agricultural site is located within a dryland region, the physical processes present on it due to management might make it more closely related with mesic ecosystems, which the model is expected to be properly trained for. However, short distance changes in land use are likely to affect the remotely sensed products, considering the more widely available ones have pixel sizes between

250 and 500 meters. At that resolution, the presence of roads, buildings, and other structures will be mixed, in some cases, with the signal of the agricultural land. The natural site, in contrast, will have greater homogeneity within pixels, as land use is more homogenous and far from anthropogenic structures of significant size. Nevertheless, the response to pulse events in the natural site might be challenging to capture by the remotely sensed products due to the data acquisition frequency. In addition to this, the natural site's characteristics such as water scarcity, vegetation adaptations and soil-water dynamics might challenge the assumptions present in the existent model. Therefore, the performance of the model is expected to be limited for both systems but at different levels. In the case of the agricultural site, the remotely sensed data is expected to present the primary limitation, while for the natural site both the remotely sensed and the meteorological data are expected to present limitations in the context of the model. The potential for informing management decisions is expected to be limited by the lack of inclusion of more spatially variable data and the coarse spatial scale of the remotely sensed data used.

Study areas

For this project, ET at two dryland sites will be evaluated with the hybrid machine learning model developed by Zhao et al. (2019). The chosen sites represent two extremes of dryland-located sites dynamics: a natural site US-Jo1 (Jornada Experimental Range Bajada Site) and an irrigated orchard US-PeA (5R Pecan Farm Above Canopy). Both sites have a running eddy covariance system with an additional set of biometeorological sensors.

Jornada Experimental Range Bajada site

The Jornada Experimental Range is located in the Jornada del Muerto Basin, in Southern New Mexico, at the northern end of the Chihuahuan desert. This experimental range is part of the National Science Foundation's long-term ecological research network, and it covers 78266 ha and is located 20 km from Las Cruces, New Mexico (Greenland and Anderson, 1997). It was established in 1912, after a series of droughts combined with overgrazing led to a transition from native grass to shrubs, and aims to study the dynamics of this arid ecosystem in response to the impacts of climate change, land-use practices and disturbances (Greenland and Anderson, 1997; Tweedie, 2023). The Bajada site flux tower is located in a representative shrubland on the piedmont slope to monitor the CO_2 and H_2O dynamics (Tweedie, 2023). The vegetation in this site is predominantly *Larrea tridentata* (Creosote) and winter-deciduous *Prosopis glandulosa* (Honey mesquite), with a lower presence of *Flourensia cernua* (tarbush) and the grasses *Muhlenbergia porteri* (Bush Muhly) and *Dasyochloa pulchella* (Fluff Grass) (Greenland and Anderson, 1997; Tweedie, 2023). Local fauna includes stray domestic cattle, free-ranging introduced Oryx, and other native herbivores (Tweedie, 2023). The mean annual temperature is 15.7°C, with a mean minimum of 13.3°C, and mean maximum of 36°C. The annual precipitation average in the region is 247 mm, with over half of it occurring in brief, local, but intense thundershowers from July to

September and most of it being consumed as actual evapotranspiration (Greenland and Anderson, 1997). The flux and biometeorological data collected from this site covers from 2010 to 2020, from the Ameriflux database (Tweedie, 2023).

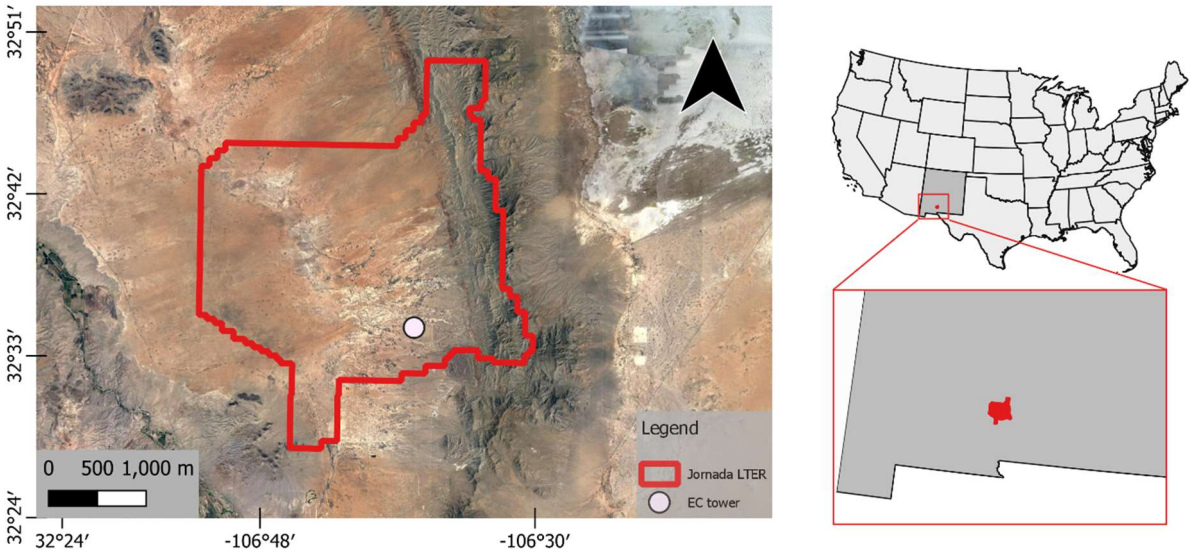

Figure 3. Site location map for Jornada LTER Bajada site eddy covariance tower.

5R Pecan Farm Above Canopy site

This site is located on the 5R Farm pecan orchard in Tornillo, TX, USA, which is part of the Trans-Pecos irrigation district. The orchard covers over 400 ha with trees over 20 years old, on average, of the Western Schley variety, planted in square patterns of 9 m by 9 m, and with soils ranging from loam to silty clay loam (Gutschick and Sheng, 2013). Irrigation water use, on average, corresponds to 127 mm per irrigation, with between 12 and 14 irrigation events per season and an additional irrigation event before the growing season to leach out excess salts in the soil (Palmate et al., 2022), surface water availability is highly variable (Gutschick and Sheng, 2013) leading to the use of groundwater to supplement surface water. The mean annual temperature is 17°C, with a mean minimum of 9°C, and mean maximum of 27°C. The annual precipitation

average in the region ranges between 203 and 304 mm. The flux and biometeorological data collected from this site covers from 2022 to 2023, with gaps in the data due to instrument malfunctioning.

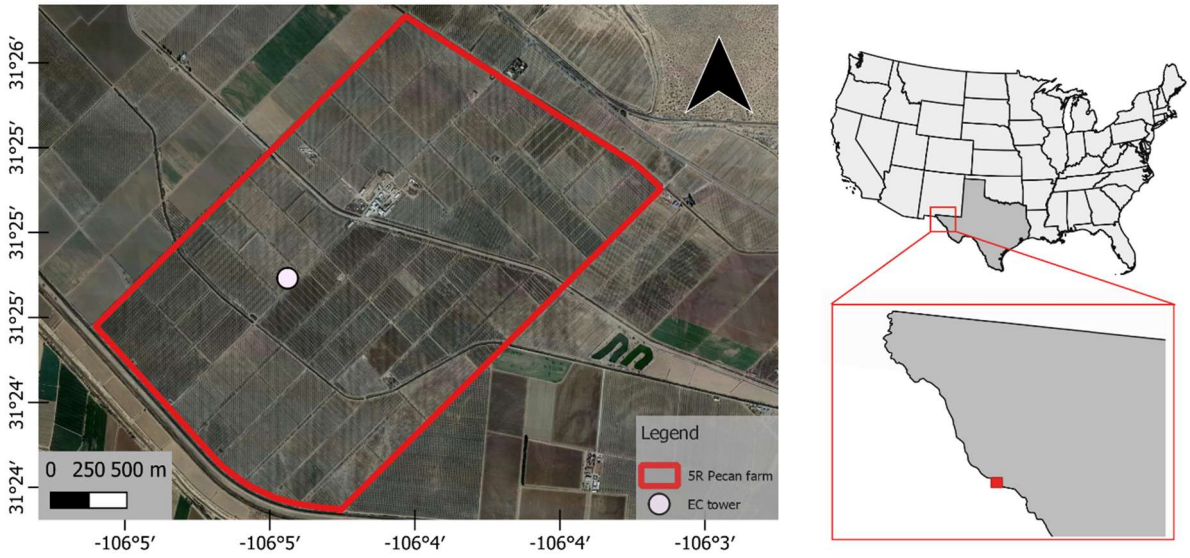

Figure 4. Site location map for 5R Pecan Farm eddy covariance tower.

Methodology

Hybrid machine learning models for estimating ET

Hybrid machine learning models represent a new tool for generating accurate predictions of variables, like ET, while preserving physical constraints, such as the energy balance. Zhao et al. (2019) developed one such of these models, an Artificial Neural Network hybrid model for predicting ET with 13 variables, integrating both flux tower data and remotely sensed data. In Zhao et al. (2019) two models were developed, a pure machine learning model and a hybrid one that integrated a modified Penman-Monteith equation to predict LE. In this hybrid model, the surface resistance (r_s) was considered a sub-model in the physical equation for predicting LE, as it is the main unknown term. Therefore, instead of LE, r_s was the prediction target for the purpose of ensuring energy conservation and constraining ET as a down gradient of vapor pressure. The results showed nine key predictors (fraction of photosynthetically active radiation, soil water content, temperature, carbon dioxide concentration, wind speed, relative humidity, canopy height, photosynthetically active radiation, and plant function type) for both models. While the pure ML provided a better fit for the data (R^2 values of 0.81 and 0.78, respectively), energy was not conserved, and, overall, the hybrid model had a lower mean absolute percent error.

Overall, the hybrid model was systematically better at predicting extremes in the dataset, suggesting that these types of models can generalize better than pure ML models. Additionally, the prediction of r_s was used to systematically understand its dependence on the predictor variables, leading to identifying features such as the role of carbon dioxide effects on stomatal conductance at the ecosystem scale. However, despite evaluating multiple plant function types, the study did not consider dryland systems directly. Given this gap and the apparent versatility of the hybrid model, it will be replicated in this project with the purpose of evaluating its accuracy for drylands.

Flux tower data

The flux tower data will come from three sources: the FLUXNET 2015 1 Tier 1 dataset (Pastorello et al., 2020), the US-Jo1 site (Tweedie, 2023), and the US-PeA site (Mauritz, 2023). The entirety of the FLUXNET dataset (excluding US-Jo1) will be used for training and validating the model, replicating the approach of Zhao et al (2019), while the two dryland site data will be used individually for applying the model. The FLUXNET and US-Jo1 data are available to download on the FLUXNET and AMERIFLUX websites, respectively. The US-PeA dataset is currently being collected and has yet to be submitted to Ameriflux. The variables of interest from these datasets include soil water content, vapor pressure deficit, plant function type, air temperature, carbon dioxide concentration, wind speed, atmospheric pressure, relative humidity, net radiation, soil heat flux, and photosynthetically active radiation (PAR).

Data pre-processing

Following Zhao et al. (2019) methodology, the FLUXNET 2015 tier 1 data set (https://fluxnet.fluxdata.org/data/download-data/) was downloaded with information from 206 sites. Afterwards, the corresponding plant function type was assigned to each site based on International Geosphere-Biosphere Programme (IGBP) vegetation classification scheme from information already available in the FLUXNET database. Data filtering was then done following the criteria already established for the model. This consisted on using only measured and good quality gap filtered data, data sampled during rainy days as well as the following day was removed to avoid data quality drops, only daytime data was used (sensible heat > 5 W/m^2 and incoming shortwave radiation > 50 W/m^2) to avoid stable boundary layer conditions, and records with negative gross primary productivity, LE or vapor pressure deficit were filtered out. In addition to this, the authors applied a percentile filter for carbon dioxide concentration, net radiation and

ground heat flux keeping only the data between the 5th and 95th percentile. As a requirement for the model to work, the dataset was corrected to achieve energy balance closure with a Bowen ratio method when closure was under 0.8 or over 1.2. After applying the previous filters, the dataset was reduced to 86 sites, 4 more than the number mentioned in the cited research.

In addition to the FLUXNET data, the methodology required remotely sensed data corresponding to the fraction of absorbed photosynthetically active radiation. This data was retrieved on site basis from the MODIS product MCD15A3H using the Global Subset Tool (https://modis.ornl.gov/cgi-bin/MODIS/global/subset.pl). Since this product has a temporal resolution of 4 days, and following the authors methodology, this value was used for the half-hourly inputs. However, since it's not clear how the unmodified value was used in the original research the 4 days values were interpolated to daily values, allowing only for gaps of up to 7 days or 2 measurements. In addition to this, vegetation height was estimated for each flux site record following the analytical equation derived by Pennypacker and Baldocchi (2016) as:

$$h = \frac{z}{0.6 + 0.1 \exp\left(\frac{k \cdot u}{u_*}\right)}$$

Where:
h = Canopy height (m)
z = Height of the instruments (m)
k = von Karman constant (k = 0.4)
u = Horizontal wind speed at height z ($m \cdot s^{-1}$)
u_* = Friction velocity ($m \cdot s^{-1}$)

Since the model developed by Zhao et al. (2019) is based on estimating r_s, this is an additional variable that needs to be calculated from the available data to be able to train the model. To achieve this, a quadratic equation for sensible heat flux can be used to solve for r_s if we assume a turbulent derived approach. This equation, shown below as described by the authors, is derived

from the Penman-Monteith equation, a second-order Taylor approximation for the saturation vapor pressure function, and an equation for LE and H expressed using a big-leaf resistance.

$$a \cdot H^2 + b \cdot H + c = 0$$

Where:

$$a = \frac{1}{2}\beta\left(\frac{r_a}{\rho C_p}\right)^2$$

$$b = \frac{\Delta r_a}{\rho C_p} + \frac{\gamma\left(r_a + r_s\right)}{\rho C_p}$$

$$c = VPD - \frac{\gamma\left(r_a + r_s\right)\left(R_n - G\right)}{\rho C_p}$$

And:

$$\beta = -\frac{5006.12\left(T_a - 1811.79\right)e^{\frac{17.27T_a}{237.3+T_a}}}{\left(237.3 + T_a\right)^4}$$

T_a= Air temperature (°C)

To help the model stabilize faster the methodology requires all the variables used as inputs to be normalized using the z-score, which uses the mean and standard deviation to account for the differences in the range of values, as this helps the optimization algorithm to converge more quickly without affecting the distribution of the data (Lindholm et al., 2022). In addition to this, the logarithm value of r_s is required as it will be the target of the model given its more normal distribution.

The methodology described above was replicated for the two dryland located sites. In the case of the US-PeA dataset, additional processing was needed. The data was filtered by signal strength, plausibility, and diagnostic flags, as well as processed for removal of spikes, following

Mauder and Foken (2011) using EddyPro (h https://www.licor.com/env/support/EddyPro/home.html). In addition to this, the Webb method was be used for density correction (Webb et al., 1980) along with coordinate planar rotation (Wilczak et al., 2001). Gap filling and U star filtering was done using the web REddyProc tool (https://bgc.iwww.mpg.de/5622399/REddyProc) (Wutzler et al., 2018) for short periods of no data or where extreme flux values were filtered. Ground heat flux was calculated using two heat flux plates with overlying soil thermocouples and a soil water content probe located beneath the trees' canopy.

Model definition

The initial parameters for the model used in this work will closely follow those defined by Zhao et al. (2019), as this research aims to evaluate the performance of this previously established hybrid model on two different dryland sites.

The model's architecture includes an input layer, five hidden layers with 64 neurons each, and an output layer, using a rectified linear unit as activation function. The aggregated dataset was shuffled randomly across time and sites and then separated with a distribution of 80 - 20% of the FLUXNET based dataset for training and testing, respectively. The two drylands sites based datasets were left out for additional testing after the base model results were considered satisfactory. The type of machine learning model used was an Artificial Neural Network model, or Neural Network (NN) for short, using the Python package ScikitLearn (Pedregosa et al., 2011). The structure used followed the one optimized by the authors through the Aikake Information Criterion, with six total layers, which includes the input layer, five hidden layers with 64 neurons for each one, and an output layer with a single neuron. However, further calibration and optimization of the model was needed as parameters such as batch size and learning rate were not

specified. To make this model hybrid, a modified quadratic Penman-Monteith equation was used to predict LE from r_s, meaning that the NN model will be set to predict r_s instead of LE. In addition to this, to constrain the model to the surface energy balance and the diffusion-like process, a corrected second order PM equation was used as the loss function. To simplify the equation, the loss function solves for sensible heat flux instead of for LE. The equation is defined as follows:

$$Loss(\widehat{H}, H) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{b_i} - b_i)^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{c_i} - c_i)^2}$$

Where:

$$b = \frac{\Delta r_a}{\rho C_p} + \frac{\gamma(r_a + r_s)}{\rho C_p} \quad , \quad c = VPD - \frac{\gamma(r_a + r_s)(R_n - G)}{\rho C_p}$$

And:

H = Sensible heat flux ($W \cdot m^{-2)}$
n = Number of samples
^ represent the predictions, while the items without ^ represent the true values
Δ = Slope of the saturated vapor pressure curve ($kPa \cdot K^{-1}$)
r_a = Aerodynamic resistance ($m \cdot s^{-1}$)
r_s = Surface resistance ($m \cdot s^{-1}$)
γ = Psychrometric constant ($kPa \cdot K^{-1}$)
ρ = Air density ($kg \cdot m^{-3}$)
C_p = Specific heat of air (C_p = 1012 $J \cdot kg^{-1} \cdot K^{-1}$)
VPD = Vapor pressure deficit (kPa)
R_n = Net radiation ($W \cdot m^{-2}$)
G = Soil heat flux ($W \cdot m^{-2}$)

The predictions from the replica model r_s were then used to generate predictions for LE. The metrics used to evaluate the performance of the replica model were the same as the ones used in the original model: R^2, mean average percentage error and root mean square error.

To assess the model's performance for US-Jo1 and US-PeA sites once the base model performance was considered satisfactory the same prediction accuracy metrics were used. To evaluate the sensitivity of the model to the environmental variables for the dryland located sites,

perturbations were added to each input variable with increasing steps of 10% of the standard deviation to identify potential differences in LE and r_s drivers compared to those of the FLUXNET dataset. An emphasis is made on the word potential, as despite the model having been constrained by physical parameters, correlation might not mean causation.

Table 1. List of variables used as inputs for machine learning model.

Variable	*Units*	*Data source*	*Frequency*
Fraction of photosynthetically active radiation	ratio	MODIS (MCD15A3H)	4-day
Soil water content	$m^3 \cdot m^{-3}$	FLUXNET / Study sites	Half-hourly
Vapor pressure deficit	kPa	FLUXNET / Study sites	Half-hourly
Plant function type	category	FLUXNET / Study sites	Half-hourly
Air temperature	°C	FLUXNET / Study sites	Half-hourly
CO_2 concentration	$umol \cdot mol^{-1}$	FLUXNET / Study sites	Half-hourly
Wind speed	$m \cdot s^{-1}$	FLUXNET / Study sites	Half-hourly
Atmospheric pressure	kPA	FLUXNET / Study sites	Half-hourly
Relative humidity	%	FLUXNET / Study sites	Half-hourly
Net radiation	$W \cdot m^{-2}$	FLUXNET / Study sites	Half-hourly
Soil heat flux	$W \cdot m^{-2}$	FLUXNET / Study sites	Half-hourly
Photosynthethically active radiation	$\mu mol \cdot s^{-1} \cdot m^{-2}$	FLUXNET / Study sites	Half-hourly
Canopy height	m	Estimated following Pennypacker and Baldocchi (2016)	Half-hourly
Surface resistance	$s \cdot m^{-1}$	Inverted from PM equation	Half-hourly
Latent heat flux	$W \cdot m^{-2}$	FLUXNET / Study sites	Half-hourly

Results

Model development and performance

The statistical metrics for r_s showed that similar results were achieved between the established model and the replica created for this project, with slightly improved performance in the R^2 and MAPE for the replicated one (R^2=0.63 and MAPE=58.26% against and average R^2=0.76 and MAPE=41.55%) but considerably worse performance for the RMSE (RMSE = 90.63 against an average RMSE = 357.7) (Figure 2Figure 5). These differences are mainly attributed to variations in data processing and model calibration, as it is suspected that information related to some of these aspects was missing in the description of the established model.

In the case of the statistical metrics for LE, the performance of the replica model was preserved better, with only major differences between the values of the MAPE (MAPE=21.85% against an average MAPE=87.81%). A summary of these metrics is available in Table 2. In particular, the difference in in sample sizes between the models (N=32998 against N=542612) is of concern as the number of FLUXNET sites used for training is very similar, with 82 sites for the established model and 86 for the replica.

The performance of the replica model in the target sites decreases considerably for both r_s and LE. The overall performance of the replica model for predicting r_s values was slightly better for US-Jo1, while predicted LE values were considerably better for US-PeA. However, the performance of the model for US-PeA is considered highly uncertain due to the relatively small amount of samples available.

Table 2. Summary of statistical metrics of hybrid model performance (training, validation, test) for predicting r_s and LE against the established model and the study sites.

	r_s			**LE**			
Dataset	*R^2*	*MAPE*	*RMSE*	*R^2*	*MAPE*	*RMSE*	*N*
Training	0.76	41.40	356.77	0.75	81.62	60.91	347270
Validation	0.76	41.72	359.68	0.75	82.46	60.01	86832
Test	0.75	41.53	356.86	0.74	99.36	61.10	108510
Established model	0.63	58.26	90.63	0.76	21.85	55.80	32998
US-Jo1	0.55	58.64	872.09	0.36	503.99	161.83	26844
US-PeA	0.46	56.29	533.32	0.67	524.04	371.25	2033

High levels of dispersion are observed when comparing r_s model predictions against estimated values (Figure 5a and c). An apparent gap seems to appear consistently around the r_s prediction value of 170, this is believed to be related to the filters required by the methodology. The plots showing these relationships below show only the range of values consistent with the one shown in the established model methodology, however in reality the range of values is wider due to the presence of extreme outliers which are likely to be the source of the high values for the r_s RMSE and the LE MAPE. However, despite the dispersion of the r_s prediction values, the distribution and trend of the values of LE derived mathematically from them (Figure 5d and f) remain highly consistent with those shown on Zhao et al. (2019) albeit with a considerably higher MAPE. It is unclear if additional filters related to the possible existence of two positive solutions for the quadratic equation to calculate LE where used, or if the difference is mainly related to the higher number of samples used in the replica model.

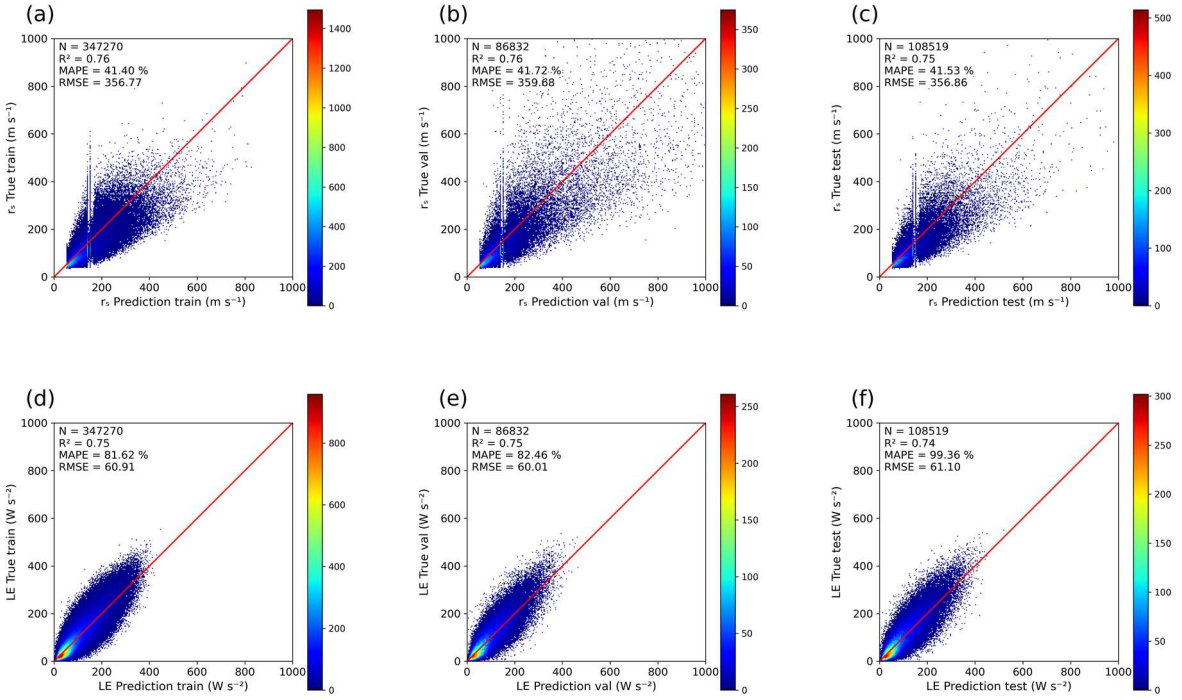

Figure 5. Comparison between the true values from the FLUXNET dataset and the predicted values derived from replica model following Zhao et al. (2019) methodology of r_s (a-c) and LE (d-f) for the train, validation and testing datasets. The red line represents a 1:1 slope while the gradient indicates the scatter point's density.

Model evaluation at study sites

The predictions for US-Jo1 and US-PeA (Figure 6) show a different trend to that of the FLUXNET dataset. In both sites, r_s gaps are present, similar to those observed in the FLUXNET dataset, suggesting that this might be due to a step in the calculations. For US-Jo1 (Figure 6a), r_s is consistently underestimated, while for US-PeA (Figure 6b) the sparseness and dispersion of data makes it difficult to assess patterns other than a slightly less pronounced underestimation.

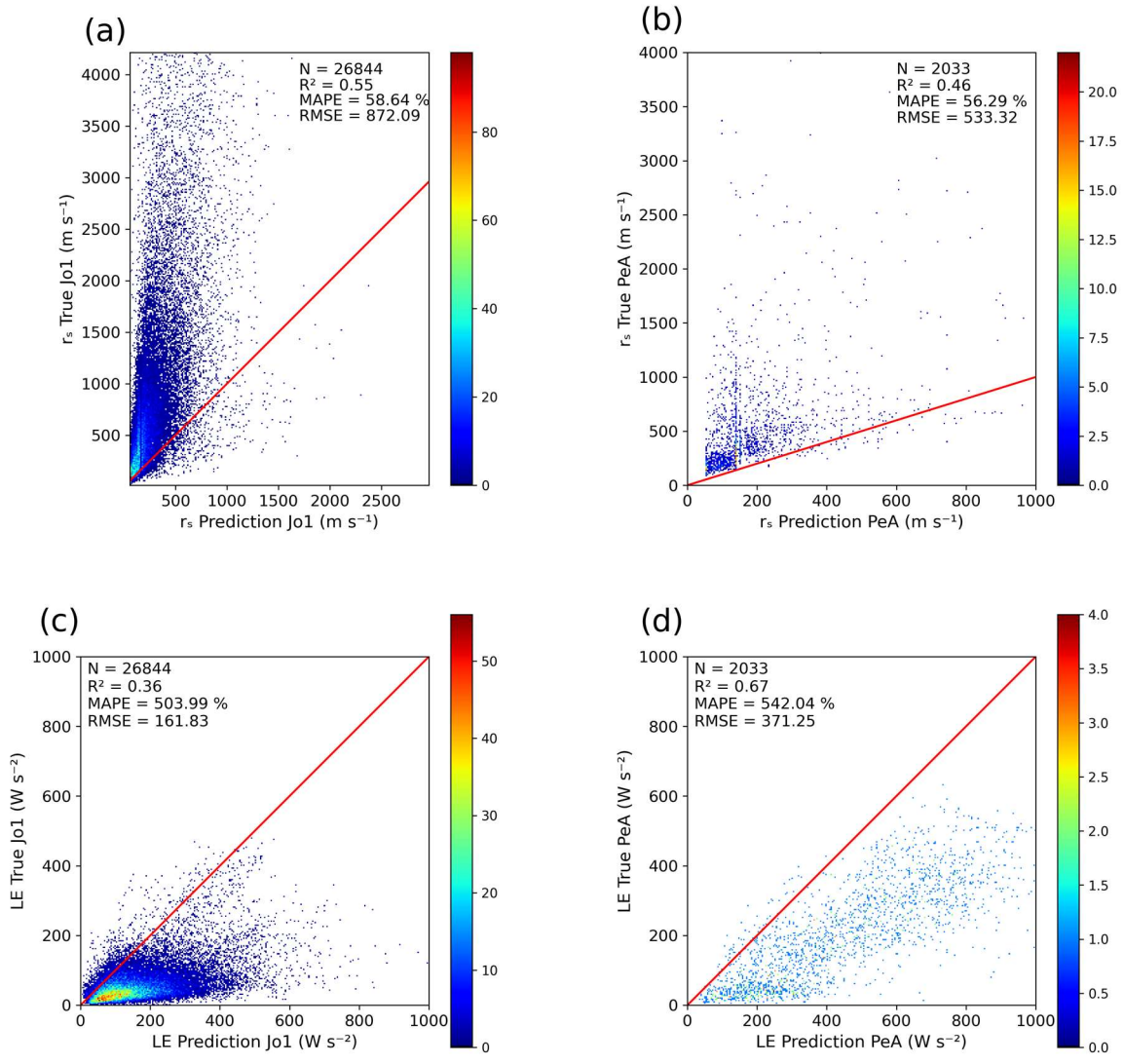

Figure 6. Comparison between measured (true) LE values and r_s values with predicted values for both variables in Jo1 (a, c) and PeA (b, d). The red line represents a 1:1 slope while the gradient indicates the scatter point's density.

The distribution of the LE values across both sites shows that the range of values predicted is higher than that of the measured values. In particular, for US-Jo1 (Figure 6c), it is observed that the measured values range between 0 and 200 Wm^{-2} while the model-predicted values are concentrated between 0 and 600 Wm^{-2}. While a similar trend of overestimation is present for US-PeA (Figure 6d), it is once again less pronounced with a seemingly linear trend that could be associated to a semi constant displacement rather than a completely different range of values.

The results of the sensitivity analysis (Figure 7) does not show clearly which variables are apparently more sensible to perturbations, especially when compared to the clear trends seen in the original research presented by Zhao et al. (2019). However, even small percentages of perturbation seem to lead to a considerable drop in R^2 values for all variables, however, afterwards the R^2 remains almost constant. The sole variable which seems to have a consistent behavior across sites is PAR, which for all cases is the less affected by perturbations.

It is noted that the R^2 values for the r_s sensitivity test in both sites become negative for some variables. While by definition R^2 represents the square of the coefficient of correlation and is intuitively assumed to only range between 0 and 1, the formula for R^2 is defined as:

$$R^2 = 1 - \frac{Residual\ sum\ of\ squares}{Total\ sum\ of\ squares}$$

Where the residual sum of squares represents the variation in data that is not explained by the model and the total sum of squares represents the difference between the expected and the actual values. Thus, negative R^2 values can happen when the residual sum is major than the total sum of squares. This inequality is possible when the model is evaluated in a different dataset from that in which it was fitted for and its variation is larger than that from the original dataset. All negative R^2 values indicate a fit worse than that of the average line (Chicco et al., 2021).

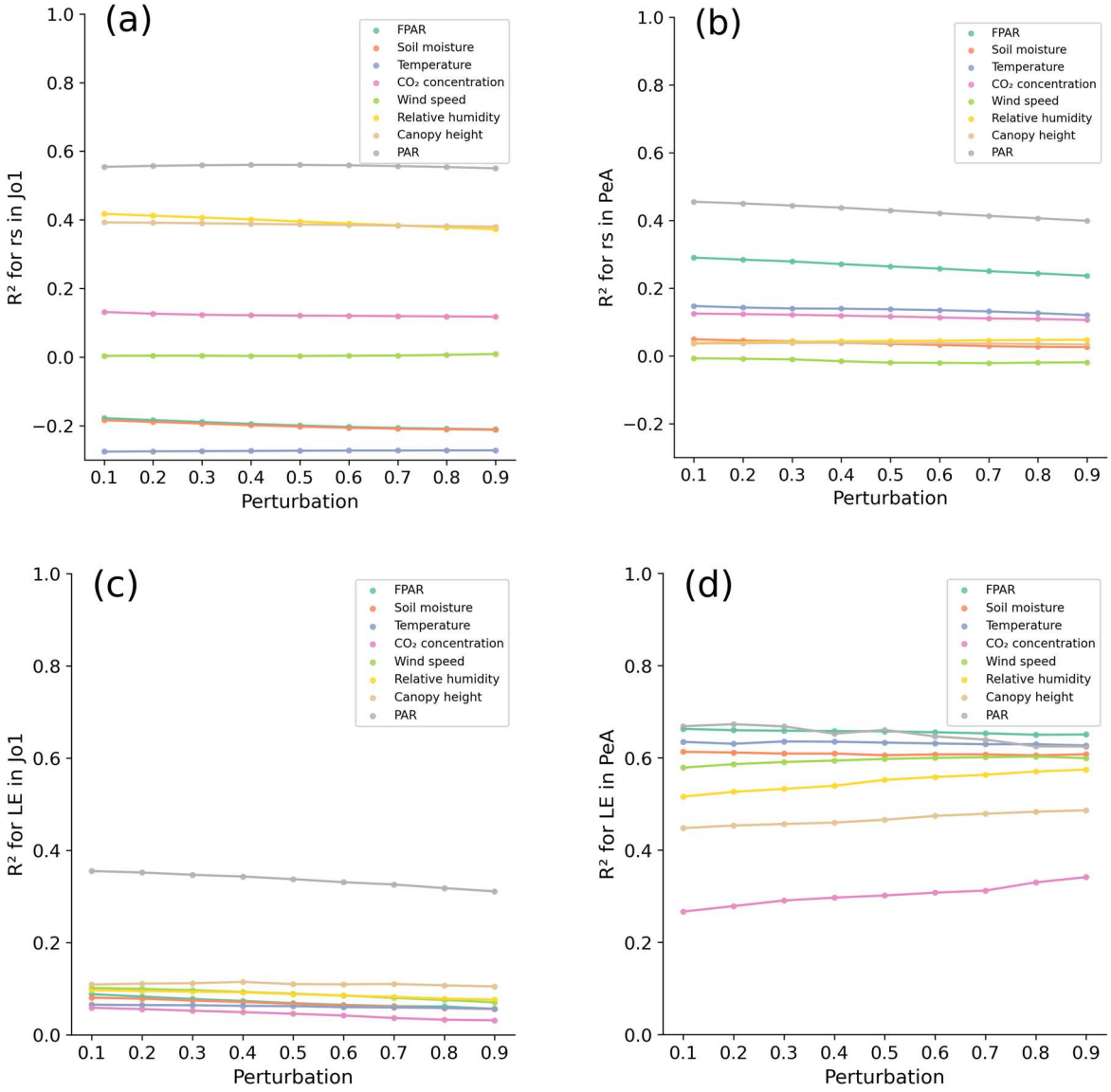

Figure 7. Results of sensitivity analysis for Jo1 (a-c) and PeA (b-d) for r_s (a-b) and LE (c-d).

To further evaluate the differences between sites, the relationship between the input variables and the predicted LE values for both sites was assessed (Figure 8). A clear difference was observed between sites for the CO_2 concentration values and the soil moisture (Figure 8a and b), both of which can be attributed to the difference in vegetation type and distribution as well as to specific site conditions. Otherwise, a general overlap between the remaining predicting variables is observed. The overall similitude between these remaining variables distribution is attributed to

the relative geographical closeness between the sites and the fact that they are all meteorological variables, although the limited data availability for PeA might be obscuring other possible relationships.

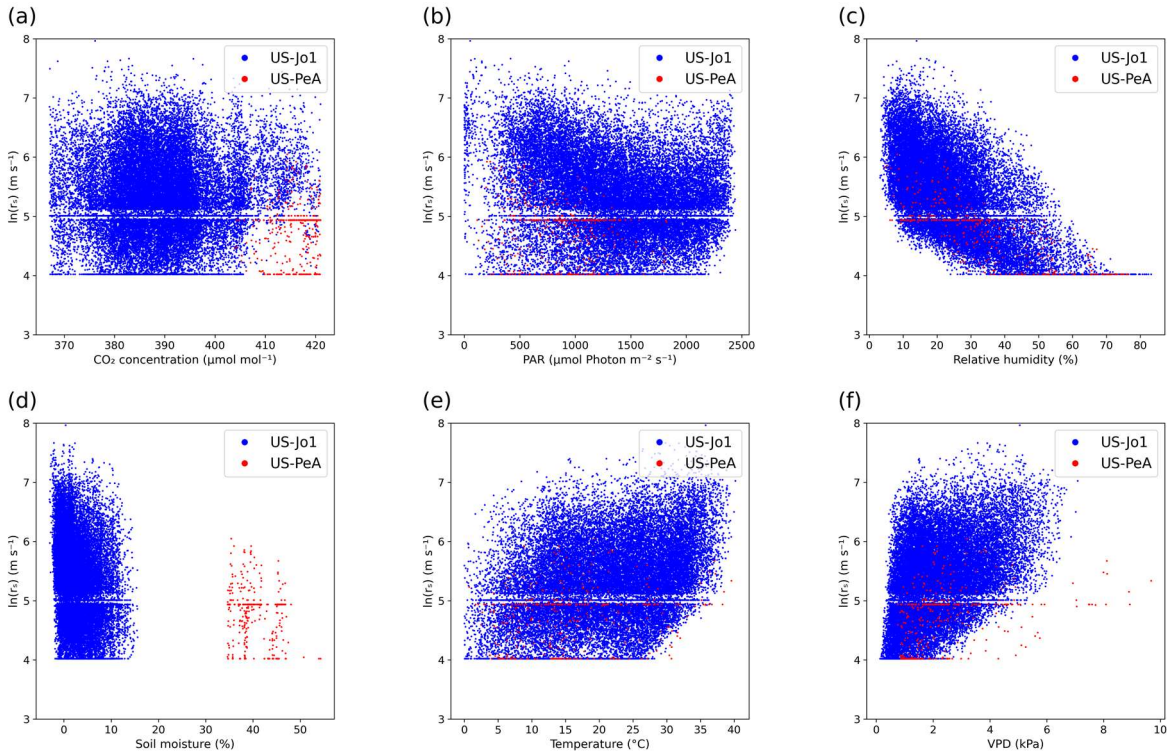

Figure 8. Key predicting variables against predicted ln(r_s) values for both sites including (a) CO_2 concentration, (b) PAR, (c) relative humidity, (d) soil moisture, (e) temperature, and (f) vapor pressure deficit.

The canopy height values for both sites used as an input in the model were evaluated, as they are one of the key variables in the original model but also represent an additional source of uncertainty as a calculated value. Results (Figure 9) show implausible values for Jo1, ranging from 0 to 7 meters, which are not representative of the predominating creosote and mesquite vegetation found at the site. For PeA, while the values are in range close to that of the pecan trees (between 10 and 12 meters), the range goes above plausible values and there is an apparent 'shortening' of the trees related to the trees leafing out.

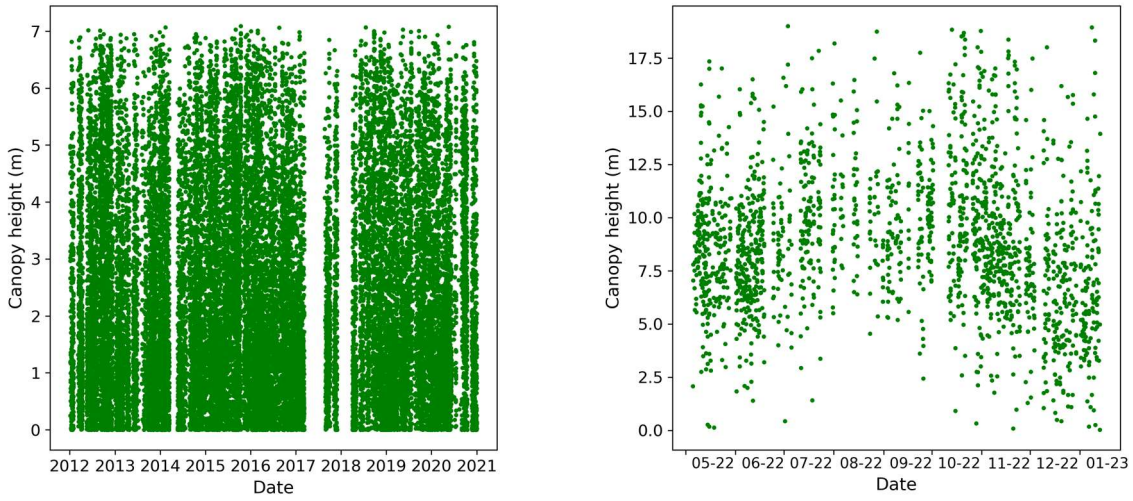

Figure 9. Time series of calculated canopy height values for both sites.

Given the marked seasonality of key events such as the North American monsoon season for both sites, and the irrigation season for PeA, the time series of the predicted and true values for both sites was compared (Figure 10). Overall, the predicted values were consistently overestimated across time for both sites, however, the overall pattern for both the predicted and the true values remained similar suggesting the prediction error is mostly related to the magnitude of the values.

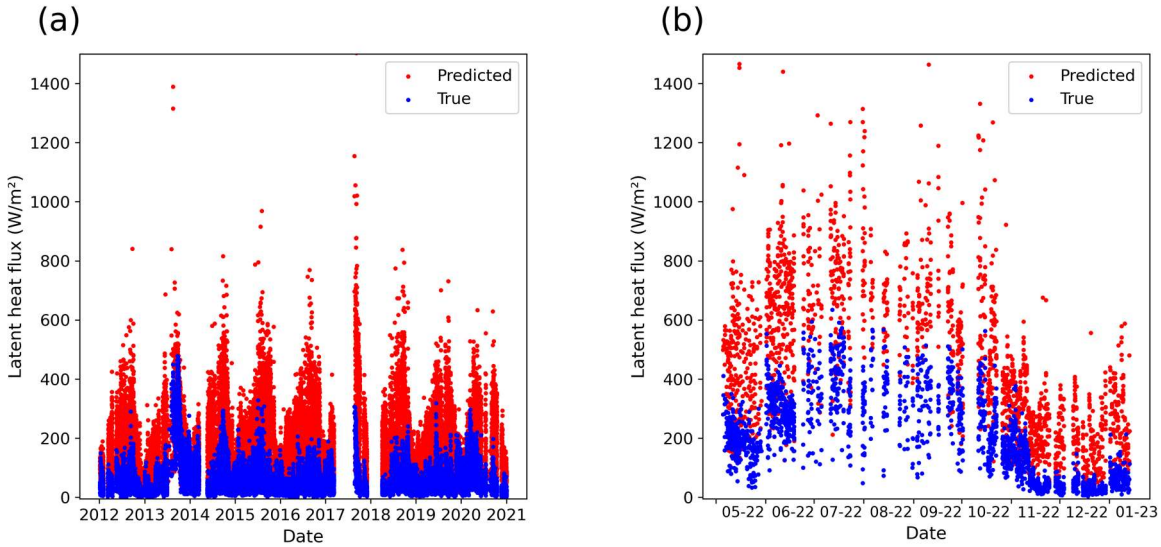

Figure 10. Overlapping time series of predicted and true LE values for (a) Jo1 and (b) PeA.

Discussion

The performance of the replica model was considered satisfactory, with RMSE being the only metric with a considerable discrepancy (Table 2). However, RMSE is a metric susceptible to the effect of data outliers that are likely present in the replica and not in the established model due to the large difference in sample sizes. This work was not able to find the source of the difference in sample sizes, which is even more disconcerting considering the number of sites used from the FLUXNET dataset after applying the filters required was almost the same (i.e. 82 against 86). It is also noted that in order to fulfill the assumptions required by the model, a large number of sites and PFT from the original FLUXNET dataset are excluded. This is an initial limitation for the applicability of the model.

Overall, the performance of the replica model for the dryland located sites was deficient, more so for Jo1 than for PeA. The distribution of the LE values across both sites (Figure 6c and d) suggests that the range of values predicted by the model is higher than the true values and LE is being overestimated. This predominant lack of agreement between true and predicted values when using globally trained models is consistent with findings by other authors (Barnes et al., 2021; Biederman et al., 2017; Dannenberg et al., 2023) where, for example, models based on MODIS satellite observations have been shown to underestimate mean annual ET by around 50%. It is worth mentioning as well that due to corrections needed to force energy balance closure and fulfill the model assumptions, there are cases where the true value does not match the measured value at the tower site.

The difference between sites level of underestimation is suspected to be related to the increased water availability of water associated to PeA being an irrigated agricultural field as well as the effect of estimating canopy height for Jo1, a variable with high importance in the original

established model. The increased water availability likely make this site conditions similar to a more mesic site, which the original model is better trained for. In fact, the PeA site PFT is classified as a deciduous broadleaf forest (DBF), which was trained for and evaluated in the original established model. However, there is still a considerable difference within statistical metrics between the DBF original model average R^2 of 0.75, MAPE of 20.84 and RMSE of 64.04 and the PeA R^2 of 0.67, MAPE of 524.04 and RMSE of 371.25. The effect of the calculations for canopy height is another possible high impact factor in the case of Jo1. This site is located in a shrubland with non-homogeneous canopy height distribution in the tower footprint. A very different case to that of PeA, a mostly homogeneous canopy. The method used for calculating the canopy height does not consider vegetation sparseness and thus likely produce values that are not representative of the actual footprint, indicating a higher vegetation coverage than it actually exists. This is believed to be the main cause of the extreme LE overestimation for Jo1.

In the case of the sensitivity analysis, the apparent lack of consistency in the order in which variables R^2 value decreases within the same site is unexpected. For example, in US-Jo1 sensitivity analysis results, the R^2 values for perturbation of CO_2 concentration goes from being only mildly affected in the case of r_s to being the most affected variable in the case of LE (Figure 7a and b). The sensitivity analysis results between sites show opposite behaviors, with accuracy for Jo1 being extremely sensible to small perturbations for all variables except PAR while accuracy for PeA appears to be mostly insensible to perturbations except for those in relative humidity, canopy height, and carbon dioxide concentration which additionally appear to increase accuracy as the perturbation increases. There is an apparent insensitivity to perturbations in PAR values likely explained by it being the only remotely sensed variable and being temporally downscaled from 4-day intervals to a single daily value for all the half-hourly records of the day. This is believed to

lead to its limited effect on the prediction of r_s and could be eliminated to simplify the model. A case can be made for both removing remotely sensed variables and adding more. Keeping only measured or measurement-derived predicting variables is a favorable factor for evaluating the possible relationships between variables if enough credence in the model has been established, while integrating multiple remotely sensed variables from different sources can help the precision of the predictions as shown in Dannenberg et al. (2023).

It is important to mention that the type of sensitivity analysis performed here was only one-at-a-time, which while simple and commonly used, has the disadvantage of not considering the factors interactions making it specially limited when analyzing more complex processes, such as ET. However, this was the method used in the original model by Zhao et al. (2019), and was thus the one implemented.

The poor model performance of the replica model in the two dryland-located sites suggests multiple issues. First, the assumptions made by the original model limit the ecosystems for which the model can be trained by requiring the energy balance to close and by excluding extreme events. In sites where the storage component plays a bigger role, forcing the energy balance closure leads to, either removing large portions of the datasets, or using modified and non-representative data for training the model. In addition to this, large pulse events captured by the dryland datasets might be excluded as outliers, by following the original model criteria. This means that even if more dryland sites are included in the training of the model, the parameterization extracted from them might not be representative, as key dryland-specific processes are not captured and factors such as canopy height are misrepresented.

The hybrid nature of the model replicated here presents a novel approach that could lead to exploratory modeling where if enough credibility in a model is established it could lead to

inferences about the driving factors or processes behind the predicted output. While different and successful modelling efforts have been made for predicting carbon and water fluxes using ML in the North American Southwest (Barnes et al., 2021; Dannenberg et al., 2023), they lack the implicit coupling of the physical processes involved which limits their ability to deepen the understanding of the unique processes behind, as they are not interpretable ML models.

While the results show that the hybrid ML model proposed by Zhao et al. (2019) is not adequate for the study sites as is, it is considered a starting point which with further evaluation and refinement could lead to a more physically realistic model of carbon and water fluxes in drylands that integrates both its spatial and its temporal context. At this point, there is no potential for the model to inform management decisions, as the performance is extremely poor for both sites. However, the hybrid nature of the model and its expected interpretability once properly refined are expected to change this, especially if the prediction is not only time series based but spatially based as well.

Conclusions

Evaluating the already established hybrid ML model by Zhao et al. (2019) in two dryland-located sites has shown an apparent deficit in its ability to predict LE accurately, with a marked trend towards overestimation of ET across sites. However, the factors associated to this overestimation are not completely clear and need to be further studied. No major inferences about the structure or parameterization of the model were derived from the results of the sensitivity analysis, other than an apparent insensitivity to perturbations in PAR values and PeA having no sensitivity to the predictors, which could be related to the lack of data. Evaluation of the temporal variation of true against predicted value shows no evidence of the overestimation being related to a seasonal or temporal factor.

Existing models for estimation of carbon and water fluxes using ML model have successfully shown the inaccuracy of major globally trained models for predicting these fluxes in drylands. However, they do not fully integrate the coupling of the processes, albeit specific variables are integrated in order to explicitly account for them to some degree. A dryland based hybrid ML model could help gain further understanding of the time and spatial variability of carbon and water fluxes, making it an actual tool for management purposes. The evaluated hybrid ML model will be used as a stepping-stone to be further modified and refined for this purpose.

While research has shown the capabilities of hybrid ML models for fluxes prediction at larger scales, further work is needed to improve their performance and generalizability at sites outside their training dataset based on the physical constraints they integrate. The use of hybrid ML models for prediction of water and carbon fluxes could represent a new tool towards improved time series and spatial variability estimations for regions with low data records availability, such as drylands.

Future directions

Further improvements in the model parameters are needed to achieve accurate predictions for LE in drylands. Future work includes modifying the approach to the physical constraining as well as integrating new parameters to improve the performance of the model and better capture the LE patterns and processes of drylands. The parameters to include will be based on those included in existing dryland models. More dryland sites will be integrated for evaluation. Once the LE hybrid model is properly refined, additional modifications are expected to expand the model for carbon flux prediction with the final goal of moving from a prediction model based only on time series to a spatially based prediction model.

References

Ahlström, A., Raupach, M.R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J.G., Friedlingstein, P., Jain, A.K., Kato, E., Poulter, B., Sitch, S., Stocker, B.D., Viovy, N., Wang, Y.P., Wiltshire, A., Zaehle, S., Zeng, N., 2015. The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink. Science 348, 895–899. https://doi.org/10.1126/science.aaa1668

Alfieri, J. g., Kustas, W. p., Anderson, M. c., 2020. A Brief Overview of Approaches for Measuring Evapotranspiration, in: Agroclimatology. John Wiley & Sons, Ltd, pp. 109–127. https://doi.org/10.2134/agronmonogr60.2016.0034

Amani, S., Shafizadeh-Moghadam, H., 2023. A review of machine learning models and influential factors for estimating evapotranspiration using remote sensing and ground-based data. Agricultural Water Management 284, 108324. https://doi.org/10.1016/j.agwat.2023.108324

Austin, A.T., 2011. Has water limited our imagination for aridland biogeochemistry? Trends in Ecology & Evolution 26, 229–235. https://doi.org/10.1016/j.tree.2011.02.003

Barnes, M.L., Farella, M.M., Scott, R.L., Moore, D.J.P., Ponce-Campos, G.E., Biederman, J.A., MacBean, N., Litvak, M.E., Breshears, D.D., 2021. Improved dryland carbon flux predictions with explicit consideration of water-carbon coupling. Commun Earth Environ 2, 1–9. https://doi.org/10.1038/s43247-021-00308-2

Bestelmeyer, B.T., Okin, G.S., Duniway, M.C., Archer, S.R., Sayre, N.F., Williamson, J.C., Herrick, J.E., 2015. Desertification, land use, and the transformation of global drylands. Frontiers in Ecology and the Environment 13, 28–36. https://doi.org/10.1890/140162

Biederman, J.A., Scott, R.L., Bell, T.W., Bowling, D.R., Dore, S., Garatuza-Payan, J., Kolb, T.E., Krishnan, P., Krofcheck, D.J., Litvak, M.E., Maurer, G.E., Meyers, T.P., Oechel, W.C., Papuga, S.A., Ponce-Campos, G.E., Rodriguez, J.C., Smith, W.K., Vargas, R., Watts, C.J., Yepez, E.A., Goulden, M.L., 2017. CO2 exchange and evapotranspiration across dryland ecosystems of southwestern North America. Global Change Biology 23, 4204–4221. https://doi.org/10.1111/gcb.13686

Carter, C., Liang, S., 2019. Evaluation of ten machine learning methods for estimating terrestrial evapotranspiration from remote sensing. International Journal of Applied Earth Observation and Geoinformation 78, 86–92. https://doi.org/10.1016/j.jag.2019.01.020

Chen, H., Huang, J.J., Dash, S.S., Wei, Y., Li, H., 2022. A hybrid deep learning framework with physical process description for simulation of evapotranspiration. Journal of Hydrology 606, 127422. https://doi.org/10.1016/j.jhydrol.2021.127422

Chen, Z., Zhu, Z., Jiang, H., Sun, S., 2020. Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. Journal of Hydrology 591, 125286. https://doi.org/10.1016/j.jhydrol.2020.125286

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 7, e623. https://doi.org/10.7717/peerj-cs.623

Collins, S.L., Belnap, J., Grimm, N.B., Rudgers, J.A., Dahm, C.N., D'Odorico, P., Litvak, M., Natvig, D.O., Peters, D.C., Pockman, W.T., Sinsabaugh, R.L., Wolf, B.O., 2014. A Multiscale, Hierarchical Model of Pulse Dynamics in Arid-Land Ecosystems. Annual Review of Ecology, Evolution, and Systematics 45, 397–419. https://doi.org/10.1146/annurev-ecolsys-120213-091650

Dannenberg, M.P., Barnes, M.L., Smith, W.K., Johnston, M.R., Meerdink, S.K., Wang, X., Scott, R.L., Biederman, J.A., 2023. Upscaling dryland carbon and water fluxes with artificial neural networks of optical, thermal, and microwave satellite remote sensing. Biogeosciences 20, 383–404. https://doi.org/10.5194/bg-20-383-2023

Dhungel, R., Anderson, R.G., French, A.N., Skaggs, T.H., Saber, M., Sanchez, C.A., Scudiero, E., 2023. Remote sensing-based energy balance for lettuce in an arid environment: influence of management scenarios on irrigation and evapotranspiration modeling. Irrig Sci 41, 197–214. https://doi.org/10.1007/s00271-023-00848-9

Dou, X., Yang, Y., 2018. Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. Computers and Electronics in Agriculture 148, 95–106. https://doi.org/10.1016/j.compag.2018.03.010

Elfarkh, J., Simonneaux, V., Jarlan, L., Ezzahar, J., Boulet, G., Chakir, A., Er-Raki, S., 2022. Evapotranspiration estimates in a traditional irrigated area in semi-arid Mediterranean. Comparison of four remote sensing-based models. Agricultural Water Management 270, 107728. https://doi.org/10.1016/j.agwat.2022.107728

Fisher, J.B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., McCabe, M.F., Hook, S., Baldocchi, D., Townsend, P.A., Kilic, A., Tu, K., Miralles, D.D., Perret, J., Lagouarde, J.-P., Waliser, D., Purdy, A.J., French, A., Schimel, D., Famiglietti, J.S., Stephens, G., Wood, E.F., 2017. The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. Water Resources Research 53, 2618–2626. https://doi.org/10.1002/2016WR020175

Granata, F., 2019. Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. Agricultural Water Management 217, 303–315. https://doi.org/10.1016/j.agwat.2019.03.015

Greenland, D., Anderson, J., 1997. A Climatic Analysis Of Long-Term Ecological Research Sites: Chapter 9 Jornada Basin.

Gutschick, V.P., Sheng, Z., 2013. Control of atmospheric fluxes from a pecan orchard by physiology, meteorology, and canopy structure: Modeling and measurement. Agricultural Water Management 129, 200–211. https://doi.org/10.1016/j.agwat.2013.08.004

Hu, X., Shi, L., Lin, G., Lin, L., 2021. Comparison of physical-based, data-driven and hybrid modeling approaches for evapotranspiration estimation. Journal of Hydrology 601, 126592. https://doi.org/10.1016/j.jhydrol.2021.126592

Irmak, S., 2008. Evapotranspiration, in: Jørgensen, S.E., Fath, B.D. (Eds.), Encyclopedia of Ecology. Academic Press, Oxford, pp. 1432–1438. https://doi.org/10.1016/B978-008045405-4.00270-6

Kannenberg, S.A., Guo, J.S., Novick, K.A., Anderegg, W.R.L., Feng, X., Kennedy, D., Konings, A.G., Martínez-Vilalta, J., Matheny, A.M., 2022. Opportunities, challenges and pitfalls in characterizing plant water-use strategies. Functional Ecology 36, 24–37. https://doi.org/10.1111/1365-2435.13945

Knapp, A.K., Beier, C., Briske, D.D., Classen, A.T., Luo, Y., Reichstein, M., Smith, M.D., Smith, S.D., Bell, J.E., Fay, P.A., Heisler, J.L., Leavitt, S.W., Sherry, R., Smith, B., Weng, E., 2008. Consequences of More Extreme Precipitation Regimes for Terrestrial Ecosystems. BioScience 58, 811–821. https://doi.org/10.1641/B580908

Lindholm, A., Wahlström, N., Lindsten, F., Schön, T.B., 2022. Machine Learning: A First Course for Engineers and Scientists, 1st ed. Cambridge University Press. https://doi.org/10.1017/9781108919371

Mauder, M., Foken, T., 2011. Documentation and Instruction Manual of the Eddy Covariance Software Package TK2. Arbeitsergebnisse, Universität Bayreuth, Abteilung Mikrometeorologie, ISSN 1614-8916 46. https://doi.org/10.5194/bg-5-451-2008

Mauritz Tozer, M., Esquivel Herrera, K., n.d. Site Info for US-PeA [WWW Document]. US-PeA: 5R Pecan Farm Above Canopy. URL https://ameriflux.lbl.gov/sites/siteinfo/US-PeA (accessed 5.18.23).

Meng, X.H., Evans, J.P., McCabe, M.F., 2014. The Impact of Observed Vegetation Changes on Land–Atmosphere Feedbacks During Drought. Journal of Hydrometeorology 15, 759–776. https://doi.org/10.1175/JHM-D-13-0130.1

Miyamoto, S., Henggeler, J., Storey, J.B., 1995. Water Management in Irrigated Pecan Orchards in the Southwestern United States. HortTechnology 5, 214–218. https://doi.org/10.21273/HORTTECH.5.3.214

Myneni, R., Knyazikhin, Y., Park, T. 2015. MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. Accessed 2023-12-04 from https://doi.org/10.5067/MODIS/MCD15A3H.006

Ortiz, A.C., Jin, L., Ogrinc, N., Kaye, J., Krajnc, B., Ma, L., 2022. Dryland irrigation increases accumulation rates of pedogenic carbonate and releases soil abiotic CO2. Sci Rep 12, 464. https://doi.org/10.1038/s41598-021-04226-3

Palmate, S.S., Kumar, S., Poulose, T., Ganjegunte, G.K., Chaganti, V.N., Sheng, Z., 2022. Comparing the effect of different irrigation water scenarios on arid region pecan orchard using a system dynamics approach. Agricultural Water Management 265, 107547. https://doi.org/10.1016/j.agwat.2022.107547

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Ribeca, A., van Ingen, C., Zhang, L., Amiro, B., Ammann, C., Arain, M.A., Ardö, J., Arkebauer, T., Arndt, S.K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L.B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T.A., Blanken, P.D., Bohrer, G., Boike, J., Bolstad, P.V., Bonal, D., Bonnefond, J.-M., Bowling, D.R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S.P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T.R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B.D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P.S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K.J., De Cinti, B., de Grandcourt, A., De Ligne, A., De Oliveira, R.C., Delpierre, N., Desai, A.R., Di Bella, C.M., di Tommasi, P., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H.A.M., Eugster, W., Ewenz, C.M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., Gielen, B., Gioli, B., Gitelson, A., Goded, I., Goeckede, M., Goldstein, A.H., Gough, C.M., Goulden, M.L., Graf, A., Griebel, A., Gruening, C., Grünwald, T., Hammerle, A., Han, S., Han, X., Hansen, B.U., Hanson, C., Hatakka, J., He, Y., Hehn, M., Heinesch, B., Hinko-Najera, N., Hörtnagl, L., Hutley, L., Ibrom, A., Ikawa, H., Jackowicz-Korczynski, M., Janouš, D., Jans,

W., Jassal, R., Jiang, S., Kato, T., Khomik, M., Klatt, J., Knohl, A., Knox, S., Kobayashi, H., Koerber, G., Kolle, O., Kosugi, Y., Kotani, A., Kowalski, A., Kruijt, B., Kurbatova, J., Kutsch, W.L., Kwon, H., Launiainen, S., Laurila, T., Law, B., Leuning, R., Li, Yingnian, Liddell, M., Limousin, J.-M., Lion, M., Liska, A.J., Lohila, A., López-Ballesteros, A., López-Blanco, E., Loubet, B., Loustau, D., Lucas-Moffat, A., Lüers, J., Ma, S., Macfarlane, C., Magliulo, V., Maier, R., Mammarella, I., Manca, G., Marcolla, B., Margolis, H.A., Marras, S., Massman, W., Mastepanov, M., Matamala, R., Matthes, J.H., Mazzenga, F., McCaughey, H., McHugh, I., McMillan, A.M.S., Merbold, L., Meyer, W., Meyers, T., Miller, S.D., Minerbi, S., Moderow, U., Monson, R.K., Montagnani, L., Moore, C.E., Moors, E., Moreaux, V., Moureaux, C., Munger, J.W., Nakai, T., Neirynck, J., Nesic, Z., Nicolini, G., Noormets, A., Northwood, M., Nosetto, M., Nouvellon, Y., Novick, K., Oechel, W., Olesen, J.E., Ourcival, J.-M., Papuga, S.A., Parmentier, F.-J., Paul-Limoges, E., Pavelka, M., Peichl, M., Pendall, E., Phillips, R.P., Pilegaard, K., Pirk, N., Posse, G., Powell, T., Prasse, H., Prober, S.M., Rambal, S., Rannik, Ü., Raz-Yaseef, N., Reed, D., de Dios, V.R., Restrepo-Coupe, N., Reverter, B.R., Roland, M., Sabbatini, S., Sachs, T., Saleska, S.R., Sánchez-Cañete, E.P., Sanchez-Mejia, Z.M., Schmid, H.P., Schmidt, M., Schneider, K., Schrader, F., Schroder, I., Scott, R.L., Sedlák, P., Serrano-Ortíz, P., Shao, C., Shi, P., Shironya, I., Siebicke, L., Šigut, L., Silberstein, R., Sirca, C., Spano, D., Steinbrecher, R., Stevens, R.M., Sturtevant, C., Suyker, A., Tagesson, T., Takanashi, S., Tang, Y., Tapper, N., Thom, J., Tiedemann, F., Tomassucci, M., Tuovinen, J.-P., Urbanski, S., Valentini, R., van der Molen, M., van Gorsel, E., van Huissteden, K., Varlagin, A., Verfaillie, J., Vesala, T., Vincke, C., Vitale, D., Vygodskaya, N., Walker, J.P., Walter-Shea, E., Wang, H., Weber, R., Westermann, S., Wille, C., Wofsy, S.,

Wohlfahrt, G., Wolf, S., Woodgate, W., Li, Yuelin, Zampedri, R., Zhang, J., Zhou, G., Zona, D., Agarwal, D., Biraud, S., Torn, M., Papale, D., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Sci Data 7, 225. https://doi.org/10.1038/s41597-020-0534-3

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, É., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Pennypacker, S., Baldocchi, D., 2016. Seeing the Fields and Forests: Application of Surface-Layer Theory and Flux-Tower Data to Calculating Vegetation Canopy Height. Boundary-Layer Meteorol 158, 165–182. https://doi.org/10.1007/s10546-015-0090-0

Petrie, M.D., Peters, D.P.C., Burruss, N.D., Ji, W., Savoy, H.M., 2019. Differing climate and landscape effects on regional dryland vegetation responses during wet periods allude to future patterns. Global Change Biology 25, 3305–3318. https://doi.org/10.1111/gcb.14724

Plaza, C., Zaccone, C., Sawicka, K., Méndez, A.M., Tarquis, A., Gascó, G., Heuvelink, G.B.M., Schuur, E.A.G., Maestre, F.T., 2018. Soil resources and element stocks in drylands to face global issues. Sci Rep 8, 13788. https://doi.org/10.1038/s41598-018-32229-0

Ramirez-Valle, O., 2022. Impact Of Agricultural Practices On The Consumptive Water Use Of Irrigated Annual And Perennial Crops Of Northern Mexico And The U. S. Southwest. University of Texas at El Paso.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1

Reynolds, J.F., Smith, D.M.S., Lambin, E.F., Turner, B.L., Mortimore, M., Batterbury, S.P.J., Downing, T.E., Dowlatabadi, H., Fernández, R.J., Herrick, J.E., Huber-Sannwald, E., Jiang, H., Leemans, R., Lynam, T., Maestre, F.T., Ayarza, M., Walker, B., 2007. Global Desertification: Building a Science for Dryland Development. Science 316, 847–851. https://doi.org/10.1126/science.1131634

Smith, W.K., Dannenberg, M.P., Yan, D., Herrmann, S., Barnes, M.L., Barron-Gafford, G.A., Biederman, J.A., Ferrenberg, S., Fox, A.M., Hudson, A., Knowles, J.F., MacBean, N., Moore, D.J.P., Nagler, P.L., Reed, S.C., Rutherford, W.A., Scott, R.L., Wang, X., Yang, J., 2019. Remote sensing of dryland ecosystem structure and function: Progress, challenges, and opportunities. Remote Sensing of Environment 233, 111401. https://doi.org/10.1016/j.rse.2019.111401

Taylor, S.D., Browning, D.M., Baca, R.A., Gao, F., 2021. Constraints and Opportunities for Detecting Land Surface Phenology in Drylands. Journal of Remote Sensing 2021. https://doi.org/10.34133/2021/9859103

Tweedie, C.E., 2023. AmeriFlux BASE US-Jo1 Jornada Experimental Range Bajada Site. https://doi.org/10.17190/AMF/1767833

Webb, E.K., Pearman, G.I., Leuning, R., 1980. Correction of flux measurements for density effects due to heat and water vapour transfer. Quarterly Journal of the Royal Meteorological Society 106, 85–100. https://doi.org/10.1002/qj.49710644707

Wilczak, J.M., Oncley, S.P., Stage, S.A., 2001. Sonic Anemometer Tilt Correction Algorithms. Boundary-Layer Meteorology 99, 127–150. https://doi.org/10.1023/A:1018966204465

Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., Menzer, O., Reichstein, M., 2018. Basic and extensible post-processing of eddy covariance flux data

with REddyProc. Biogeosciences 15, 5015–5030. https://doi.org/10.5194/bg-15-5015-2018

Zhang, K., Zhu, G., Ma, J., Yang, Y., Shang, S., Gu, C., 2019. Parameter Analysis and Estimates for the MODIS Evapotranspiration Algorithm and Multiscale Verification. Water Resources Research 55, 2211–2231. https://doi.org/10.1029/2018WR023485

Zhao, W.L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., Qiu, G.Y., 2019. Physics-Constrained Machine Learning of Evapotranspiration. Geophysical Research Letters 46, 14496–14507. https://doi.org/10.1029/2019GL085291

Vita

Katya Esquivel Herrera was born in Juarez, Chihuahua, Mexico. She entered Universidad Autonoma de Ciudad Juarez and in December 2020 received the degree of Bachelor of Science in Geosciences Engineering. Then she entered University of Texas at El Paso in August 2021 and received a Masters of Science in Environmental Sciences in December 2023.

Contact Information (optional) : kesquivelh@miners.utep.edu