



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

**ACTUAL EVAPOTRANSPIRATION
ESTIMATES IN ARID REGIONS USING
MACHINE LEARNING ALGORITHMS WITH
IN-SITU AND REMOTE SENSING DATA**

JOSEFINA AMANDA MOSRE POLLER

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Master of Science in Engineering

Advisor:

FRANCISCO SUÁREZ POCH

Santiago de Chile, (May, 2020)

© 2020, Josefina Mosre Poller



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

**ACTUAL EVAPOTRANSPIRATION
ESTIMATES IN ARID REGIONS USING
MACHINE LEARNING ALGORITHMS WITH
IN-SITU AND REMOTE SENSING DATA**

JOSEFINA AMANDA MOSRE POLLER

Members of the Committee:

FRANCISCO SUÁREZ POCH

CARLOS BONILLA MELÉNDEZ

FRANCISCO MEZA DABANCENS

CÉSAR SÁEZ NAVARRETE

Thesis submitted to the Office of Research and Graduate Studies in
partial fulfillment of the requirements for the Degree of Master of
Science in Engineering

Santiago de Chile, (May, 2020)

ACKNOWLEDGEMENTS

I want to thank my thesis advisor, Dr. Francisco Suárez for being present every time that I requested his support, and for allowing me to take all the time that I needed. Not all students can boast about it. I also want to thank Tomás Oportus because without him it would have been impossible to start this thesis. Thanks for the data, the manual, the explanations, for everything.

I would also like to thank my family, my friends and my boyfriend, Luciano, for supporting me when the anxiety was stronger than me. Thanks for listen to me every time, although I know that you understood half of my explanations. I swear never to say the word “evapotranspiration” for a while.

I want to thank my teachers, my partners and the whole circus world for teaching me that the constant work and learn to ask for help are the keys to achieve everything without never lose my smile.

I thank the Agencia Nacional de Investigación y Desarrollo (ANID) for funding grant ANID/FONDECYT/1170850, as well as the Centro de Desarrollo Urbano Sustentable (CEDEUS – ANID/FONDAP/15110020) and the Centro de Excelencia en Geotermia de los Andes (CEGA – ANID/FONDAP/15090013) for supporting this investigation.

This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The ERA-Interim reanalysis data are provided by ECMWF and processed by LSCE. The FLUXNET eddy covariance data processing and harmonization was carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC and ICOS Ecosystem Thematic Center, and the OzFlux, ChinaFlux and AsiaFlux offices.

CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
CONTENTS	iv
LIST OF TABLES	v
LIST OF FIGURES	v
RESUMEN	vi
ABSTRACT	viii
1. INTRODUCTION	1
1.1. Objectives.....	5
1.2. Hypothesis.....	5
2. Materials and Methods.....	6
2.1. Study sites	6
2.2. ET _a fluxes and meteorological data	9
2.3. Remote sensing and vegetation indices	10
2.4. Determination of main variables and ET _a estimates using machine learning	13
3. Results	16
3.1. Footprints	16
3.2. Remote sensing information.....	19
3.3. Variables predicting ET _a	24
3.4. ET estimate formulas	25
4. Discussion	30
5. Conclusions	35
References	36
Appendixes	48

LIST OF TABLES

	Page
Table 2.1: Name, country, vegetation type and period of measure of each sites.	8
Table 2.2: Variables analyzed in this study and its respective units of measure.....	15
Table 3.1: Comparison between VI's values obtained with the footprints calculated with the Kljun et al. (2015) and the Schuepp et al. (1990) approaches.....	18
Table 3.2: Pearson correlation between measured ET_a and VI's values.....	20
Table 3.3: Number of times that every factor is selected in a site-specific equation for daily, monthly and monthly with VI estimates.	25
Table 3.4: Linear regression formulas developed with the training data set.....	28

LIST OF FIGURES

	Page
Figure 2.1: Study sites.....	7
Figure 2.2: Example of footprint calculation with the Schuepp et al. (1990) and the Kljun et al. (2015) approaches.....	10
Figure 2.3: Flowchart of methods.....	14
Figure 3.1: Footprints areas calculated with the Kljun et al. (2015) approach and the Schuepp et al. (1990) approach for (a) CH-AT1, (b) CH-AT2 and (c) CH-AT3 sites....	17
Figure 3.2: Temporal evolution of monthly precipitation (a), monthly ET_a (b), and monthly VI's (c) in US-Wkg.....	21
Figure 3.3: Temporal evolution of monthly precipitation (a), monthly ET_a (b), and monthly VI's (c) in AU-Ync.	22
Figure 3.4: Temporal evolution of monthly precipitation (a), monthly ET_a (b), and monthly VI's (c) in CH-AT1.....	23
Figure 3.5: Regression formulas for the training data set.....	27
Figure 3.6: Performance of the “global” formulas in the validation sites.....	29

RESUMEN

La evapotranspiración (ET) es un proceso hidrológico relevante en regiones áridas, donde el agua es vital para el desarrollo de comunidades locales y ecosistemas. Históricamente, realizar estimaciones de ET ha sido un gran desafío en estas áreas, debido a que sus paisajes se componen principalmente de vegetación dispersa adaptada a las condiciones de sequía, lo que se contradice con muchas de las suposiciones usadas en los métodos tradicionales de estimación de ET. Sin embargo, existen varios estudios realizados en zonas áridas que han mostrado buenos resultados cuando se implementan fórmulas empíricas de regresión que, a pesar de su simplicidad, son comparables en exactitud con modelos más complejos. Aunque existen muchos tipos de fórmulas de regresión para estimar ET, no existe un consenso respecto a qué variables se deben considerar en el análisis. En esta investigación se usaron algoritmos de aprendizaje automático para encontrar las principales variables que predicen la ET diaria y mensual en regiones áridas mediante el uso de ecuaciones de regresión lineal. Se utilizó como datos de entrada en las estimaciones mensuales solo información meteorológica y luego combinada con índices vegetacionales de percepción remota (VI's). Se recolectaron datos meteorológicos y flujos de ET de 10 sitios en Chile, Australia y Estados Unidos. Las estimaciones diarias y mensuales fueron evaluadas en tres sitios de validación, uno por país, en donde se obtuvo desempeños diferentes. Los resultados obtenidos indican que la energía disponible es la principal variable que predice la ET en los sitios de estudio, incluso cuando las regiones áridas son típicamente descritas como ambientes con agua limitada. El VI que representa mejor la ET es el Índice de Agua de Diferencia Normalizada (NDWI) que, a diferencia de otros VI's, representa la disponibilidad de agua en plantas y el suelo en vez de la actividad de la vegetación. El mejor desempeño obtenido en las ecuaciones de regresión se obtuvo en la estimación mensual con la incorporación de un VI el sitio de validación de E.E.U.U. ($R^2 = 0.82$), mientras que el peor se obtuvo en la estimación mensual del sitio de validación de Australia cuando solo se consideró el uso de información meteorológica. Incorporar información de percepción remota resulta en mejores estimaciones de ET, en contraste a cuando solo se incluye información meteorológica en el análisis.

Palabras claves: Evapotranspiración, percepción remota, aprendizaje automático, regiones áridas.

ABSTRACT

Evapotranspiration (ET) is a relevant hydrological process in arid regions where water is vital for the development of local communities and ecosystems. ET estimations in arid regions have been historically a great challenge, because these landscapes mainly consist of sparse vegetation adapted to drought conditions, which do not comply with many of the assumptions used in traditional ET estimation methods. Nevertheless, in arid areas several studies have shown good results when implementing empirical regression formulas that, despite their simplicity, are comparable in accuracy to more complex models. Although many types of regression formulas to estimate ET exist, there is no consensus on what variables must be included in the analysis. In this research, I used machine learning algorithms to find the main variables that predicts daily and monthly ET in arid regions using linear regression equations. Meteorological data alone and then combined with remote sensing vegetation indices (VI's) were used as input in monthly estimations. In-situ ET fluxes and meteorological data were obtained from ten sites in Chile, Australia and United States. Daily and monthly ET estimations were evaluated in three validation sites, one from each country, obtaining different performance. My results indicate that the available energy is the main meteorological variable that predicts ET fluxes in the assessed sites, even when arid regions are typically described as water-limited environments. The VI that represents better the in-situ ET fluxes is the Normalized Difference Water Index (NDWI), which unlike other VI's, represents water availability in plants and soil instead of vegetation activity. The best performance of the linear regression equations was obtained for monthly estimates with the incorporation of VI's at the U.S. validation site ($R^2 = 0.82$), whereas the worst performance of these equations was obtained for monthly ET estimates at the Australia validation site when only meteorological data are considered. The incorporation of remote-sensing information results in better ET estimations in contrast with estimations obtained when only meteorological data are included in the analysis.

Keywords: Evapotranspiration, remote sensing, machine learning, arid regions.

1. INTRODUCTION

Arid and semi-arid regions cover approximately 41% of the world's land and are inhabited by more than 2.500 million people. These regions are expected to expand following land degradation due to unsustainable land and water use, as well as climate change that exacerbates desertification (Gaur & Squires, 2018). In this context, an accurate quantification of evapotranspiration (ET), a relevant hydrological process, is important for water resources management to ensure availability for human and environmental needs (Bunting et al., 2014; Carter & Liang, 2018; Gaur & Squires, 2018; Nagler et al., 2005). ET is mainly driven by energy exchange and water availability, but there are several meteorological and vegetational characteristics that determine its magnitude, making its estimations more complex (Allen et al., 2011; Allen et al., 1998). Major challenges in ET estimation are those that make the process even more dynamic over time and variable in space. In arid regions, a large proportion of the low and sporadic precipitation returns to the atmosphere, whereas a small proportion infiltrates on soil. Also, the plant available water depends of the nature, the hydraulic properties and the water retention capacity of the soil. Further, the varying vegetation density, tree height and physiological plant adaptations due to water stress influences the transpiration process (Allen et al., 2011; Jovanovic & Israel, 2012).

Several direct and indirect methods have been developed to improve ET estimations. Lysimeters, eddy covariance systems (EC) and scintillometers are classified as direct ET estimations methods, which have been cataloged as the most accurate approaches by several authors (Carter & Liang, 2018; Meijninger et al., 2004; Nagler et al., 2005). However, these methods are often expensive and can only be fully operated by trained personnel. Also, they have limited footprints in the order of hundreds of meters or a few kilometers, which limit their applicability in basin or regional studies (Carter & Liang, 2018). Nevertheless, they are important for the evaluation of indirect methods (Allen et al., 1998).

Indirect ET estimation methods are empirical or semi-empirical formulas computed from meteorological data. The Penman-Monteith is the most used formula to estimate potential evapotranspiration (ET_o) under standard conditions (Allen et al., 1998). From ET_o , actual evapotranspiration (ET_a) can be estimated for different crops through the use of a crop coefficient (K_c), i.e., $ET_a = K_c \cdot ET_o$, where K_c represent four effects that distinguish the crop from reference grass: aerodynamic resistance, albedo, surface resistance and soil evaporation (Allen et al., 1998). Since state agencies usually provides K_c values for several crops, this method is widely used by farmers to estimate crop irrigation in a simple way. The crop coefficient method provides good results for agronomic applications since the inherent assumptions of the method are typically met, i.e. the crops are fully irrigated (Allen et al., 1998; Mata-González et al., 2005). Nevertheless, arid lands are vastly different than irrigated farms and do not satisfy this assumptions, especially because arid lands native vegetation displays high resistance to transpiration and low ground cover, and are normally under drought conditions (Mata-González et al., 2005). Recently, remote sensing methods have been developed to estimate ET_a and have been positioned as the only feasible method for wide areas of mixed landscapes, allowing to improve water balance estimations over basin and regional scales (El Masri et al., 2019; Glenn et al., 2010; Nagler et al., 2001).

The most common remote sensing ET_a approaches are based on the surface energy balance (SEB) equation, where sensible heat (H) is estimated using land surface temperature (LST) derived from thermal infrared (TIR) sensor on satellites (Glenn et al., 2010). Although these methods have been cataloged as operational, there are difficulties on their implementation: small errors in the estimation of the LST translate into large errors in H estimates, and only few sensors offer open source TIR data (Glenn et al., 2010; Yebra et al., 2013). Also, TIR bands are always coarser than visible (VIS) and infrared bands (IR) which limit field-scale implementation (Bisquert et al., 2016). For example, Landsat TIR band have a 60 m spatial resolution, whereas its VIS and IR bands have 30 m. Landsat has a better spatial resolution than MODIS and Sentinel-3 TIR bands (60 m vs 1 km spatial

resolution), however it has a poorer time frequency (Bisquert et al., 2016; Prikaziuk & van der Tol, 2019).

Vegetation index (VI)-based methods to estimate ET were developed to take advantage of remote sensing avoiding the disadvantages associated with the methods based on SEB. VI's were developed for vegetation monitoring, due to spectral reflectance signature reveal information about the state, biogeochemical composition, and structure of a leaf and canopy, but VI's can also give information about water and carbon cycles (Huete, 2012). ET estimation methods based on VI's depend on an estimate of the density of green vegetation over the landscape, as measured by VI's or related products that combine the VIS and IR bands (Glenn et al., 2010). For example, the Normalized Difference Vegetation Index (NDVI), one of the most common VI's, captures the contrast in light reflection from green leaves between the red and near infrared (NIR) bands, because red light is strongly absorbed by chlorophyll and nearly all the NIR is transmitted (Glenn et al., 2011). The Enhance Vegetation Index (EVI) and the Soil Adjusted Vegetation Index (SAVI) (Glenn et al., 2010) were developed to improve vegetation signal, whereas the Normalized Difference Water Index (NDWI) aimed to be sensitive to other properties, e.g., leaf water content (Ji et al., 2011). NDWI is a VI that is capable to represent both, canopy and soil water content. Thus, it is able to represent plant water stress. (Huang, et al., 2016; Sriwongsitanon et al., 2016). It is also less sensitive to atmospheric scattering effects than NDVI, but it does not remove completely the effects of soil background reflectance. An advantage of the NDWI over other VI's is that it represent the natural interaction between rainfall, soil moisture and leave water content (Sriwongsitanon et al., 2015). VI's have several advantages for use in ET algorithms: they are available from multiple sensors, VI's change on time scales of weeks to months, so is feasible to interpolate VI values with observations obtained several days apart, and VI methods are usually simple and resilient in the presence of data gaps (Carter & Liang, 2018).

In the context of precision agriculture, VI's were used to improve the crop coefficient method in order to represent the actual state of single or multiple crop canopies during development (Glenn et al., 2011). For this reason, VI's have been applied to natural

ecosystems reaching good results as shown by Groeneveld et al. (2007) and Nagler et al. (2013). Nevertheless, Mata-González et al. (2005) ensure that crop coefficient methods are not suitable for determining ET_a of vegetation adapted to arid conditions, because transpiration is overestimated when plants encounter suboptimal conditions of soil water, as a result of not considering stomatal regulation and plant adaptations to drought.

With the intention of making ET data more accessible, global ET data sets were derived from satellite information. Two of the most used operational ET data products are MOD16 and LSA-SAF. MOD16, based on MODIS information, is available since 2010 and has a 1 km resolution at 8-day, monthly, and annual intervals (Hu et al., 2015; Mu et al., 2013). LSA-SAF, derived from Meteosat Second Generation (MSG) satellite, has 30 min and daily ET observations with a spatial resolution of 3 km. However, it only covers Europe, Africa and the eastern part of South America (DHI-GRAS, 2020; Hu et al., 2015). Hu et al. (2015) compared the MOD16 and LSA-SAF products with EC measurements, concluding that LSA-SAF have a better performance than MOD16. However, neither products can capture ET_a in a water-limited region. ET data can also be obtained from the USDA-ARS ET dataset and the data provided from ECMWF or GLDAS models (DHI-GRAS, 2020). However, their spatial resolution are even coarser than that of MOD16 and LSA-SAF datasets (DHI-GRAS, 2020). Nowadays, the European Space Agency (ESA) is developing an open-source software application for ET modeling at high (tens of meters) and medium (1 km) spatial resolutions with the observations of Sentinel-2 and Sentinel-3 for field-scale applications named Sen-ET. First validation procedure in latent heat flux results in a correlation of 0.76 with the best performance obtained in croplands (DHI-GRAS, 2020). More detailed validation results are available in the “Prototype benchmarking and description” document (Nieto et al., 2019).

Several studies have demonstrated the potential to combine site-specific ET data with remotely sensed and meteorological parameters to develop empirical models based on statistical correlations for regional-scale ET estimates (Bunting et al., 2014; Glenn et al., 2010; Granata, 2019). Despite their simplicity, empirical regression formulas can produce ET_a values that are comparable in accuracy to more complex models, without as much

computational requirements for specific expertise (Carter & Liang, 2018). However, the estimation of ET_a with a higher degree of accuracy and over extended time scales has forced researchers to look for techniques such as machine learning (Torres et al., 2011).

In machine learning, a computer first learns to perform a task by studying a training set of examples. The computer then performs the same task with data it has not used before (Louridas & Ebert, 2016). This process allows making predictions of complex systems, such as the hydrological cycle and its components. Granata (2019) named some examples of machine learning applications in hydrology and mentions some researches related to ET. However, he states that these investigations are limited and the knowledge on the topic is still partial and fragmented. Besides, studies that use empirical regression formulas and basic machine learning concepts usually focus in the form of the formulas that predict ET instead of the factors that drives ET (Carter & Liang, 2018; Yebra et al., 2013).

The aim of this research is to determine the main factors that predict ET_a in arid cold regions through implementation of empirical regression formulas using machine learning algorithms with meteorological and remote sensing input data and, also, to compare the performance of these formulas when remote sensing data are included.

1.1. Objectives

The main objective of this study is to determine the main variables that predict ET_a in arid cold regions with machine learning algorithms. The specific objectives are: (i) to represent ET_a in arid areas through empirical formulas based on regressions; (ii) to determine the main variables that conform each of the regression formulas in different arid cold sites; and (iii) to determine the impact of incorporating VI's to these empirical formulas.

1.2. Hypothesis

A machine learning algorithm can identify the main variables that predict ET_a in arid regions: availability of energy, expressed as the difference between net radiation and soil

heat flux, and the availability of water, expressed as soil water content or as the Normalized Difference Water Index (NDWI).

2. MATERIALS AND METHODS

2.1. Study sites

In this study, I used 10 sites located around the world that, according to the Köppen climate classification system correspond to arid cold climate (BSk and BWk) (Kottek et al., 2006). Three of them are located in the Chilean Altiplano, two in Australia and five in the United States. Figure 2.1 presents the location of the study sites and Table 2.1 shows their main characteristics. Chilean sites are classified as desert cold climates, while other sites are cold steppe. Also, the Chilean sites are located above 4,000 m.a.s.l., whereas the sites in Australia and United States are located between 125 and 1,530 m.a.s.l. The study sites represent different ecosystems of arid cold environments, which includes grasslands, savannah and shrubland. A general description of each site is presented in Appendix A.

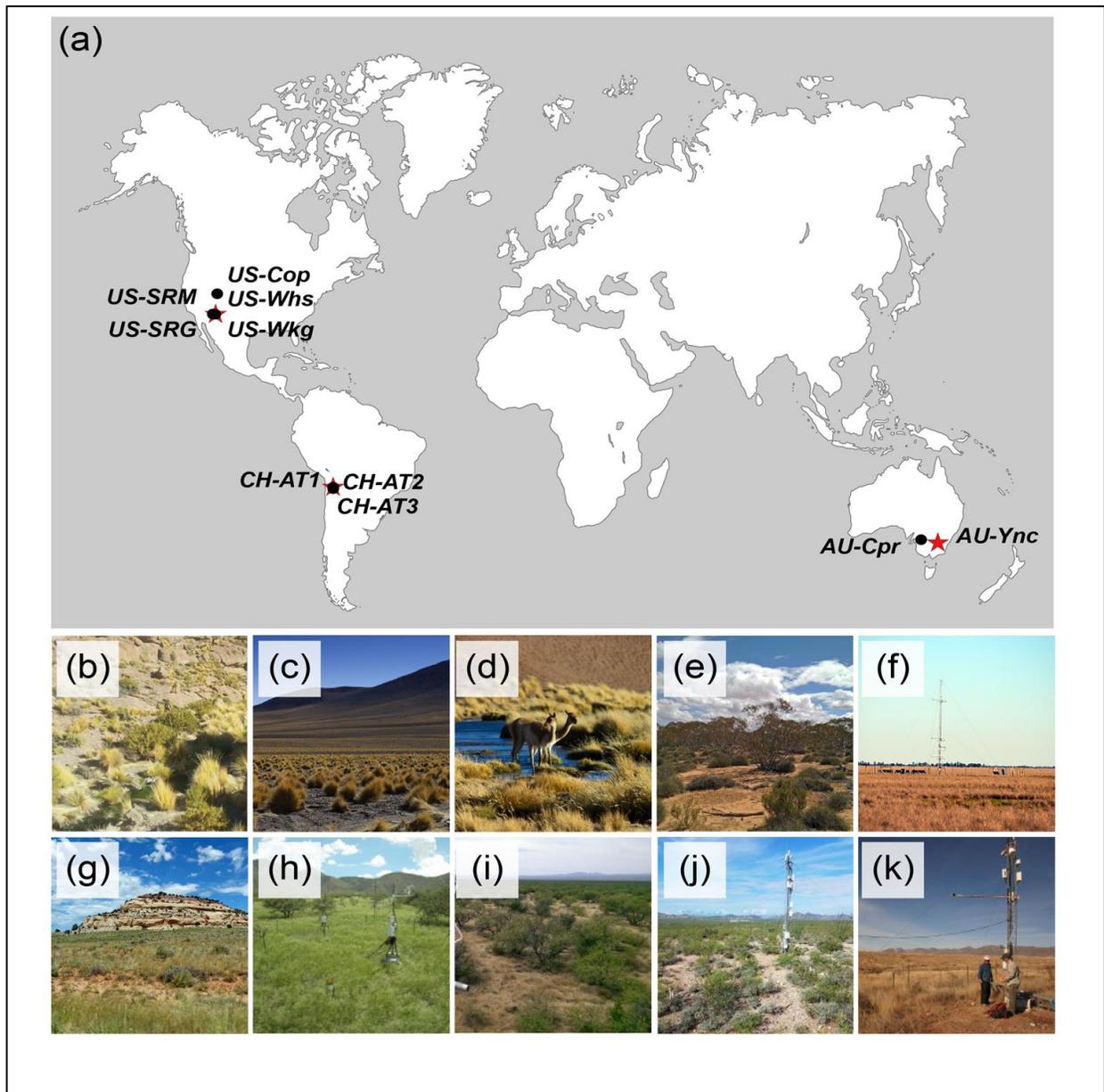


Figure 2.1: Study sites. Panel a shows the location of study sites, where red stars correspond to the sites used to validate ETa estimates. The bottom of the figure shows pictures of the environment of the study sites; (b) CH-AT1, (c) CH-AT2, (d) CH-AT3, (e) AU-Cpr (TERN, 2017a), (f) AU-Ync (TERN, 2017b), (g) US-Cop (Google Earth, n.d.), (h) US-SRG (Scott, 2012), (i) US-SRM (Scott, 2014), (j) US-Wks (Scott, 2017) and (k) US-Wkg (Scott, 2015).

Table 2.1: Name, country, vegetation type and period of measure of each sites (Beringer & Walker, 2015; D. Bowling, 2015; Meyer et al., 2015; ORNL DAAC, 2017; SCM El Abra, 2016; Scott, 2015a, 2015b, 2015c, 2015d).

Site	Country	Location		Vegetation type	Time period	
		Latitude (°)	Longitude (°)		Start date	End date
CH-AT1	Chile	-22.0268	-68.0478	Grassland with presence of hydrophytes and some shrubs	18/01/2018	29/05/2019
CH-AT2	Chile	-22.0136	-68.0456	Grassland	22/02/2018	25/04/2019
CH-AT3	Chile	-22.5247	-68.0179	Grassland with presence of hydrophytes	19/04/2018	28/05/2019
AU-Cpr	Australia	-34.0021	140.5891	Savannah	01/01/2019	31/12/2014
AU-Ync	Australia	-34.9893	146.2907	Grassland	01/01/2012	31/12/2014
US-Cop	United States	38.09	-109.39	Grassland	01/01/2001	31/12/2007
US-SRG	United States	31.789379	-110.827675	Grassland	01/01/2008	31/12/2014
US-SRM	United States	31.8214	-110.8661	Woody savannah	01/01/2004	31/12/2014
US-Whs	United States	31.7438	-110.0522	Open shrubland	01/01/2007	31/12/2014
US-Wkg	United States	31.7365	-109.9419	Grassland	01/01/2004	31/12/2014

2.2. ET_a fluxes and meteorological data

ET_a data in the Chilean sites were obtained from three Eddy Covariance systems (IRGASON, Campbell Sci., UT, USA), each one having a meteorological station that allowed measuring net radiation (R_n) (CNR4, Kipp & Zonen, The Netherlands), soil heat flux (G) (HFP01SC, Hukseflux, The Netherlands), precipitation (PPT) (TE525, Campbell Sci., Logan, UT, USA), atmospheric pressure (P) (PTB110, Vaisala, Helsinki, Finland), air temperature (T) and relative humidity (RH) (CS215, Campbell Sci., Logan, UT, USA), soil temperature (Ts) (TCAV, Campbell Sci., Logan, UT, USA) and soil's volumetric water content (VWC) (CS655, Campbell Sci., Logan, UT, USA). Vapor pressure deficit (VPD) was estimated using the previous meteorological data, and the wind speed (WS) in these sites was calculated using the measurements of the Eddy covariance sonic anemometer. On the other hand, the data from Australia and the United States were obtained from the FLUXNET 2015 dataset (Beringer & Walker, 2015; Bowling, 2015; Meyer et al., 2015; Scott, 2015a, 2015b, 2015c, 2015d).

To incorporate the remote sensing data, as described below, it is important to estimate the footprint of the ET_a measurements. Here, I approximated the footprint to a circle which radius correspond to the fetch, following the Schuepp approach (1990).

$$x_{max} = \frac{u}{u_*} \frac{(z - d)}{2k}$$

where x_{max} correspond to the fetch (m), u is the average wind speed (m/s), u^* is the average friction velocity (m/s), z is the measuring height (m), d is the displacement height (m) and k is the von Kármán constant (Leclerc et al., 2014). The footprint was calculated for each month with data from each site. This approximation was chosen instead of a more complex approach such as the Kljun et al. (2015) model, for one principal reason: the required information for more complex footprint models is not available in the FLUXNET dataset, e.g. the crosswind distance standard deviation (σ_y) (Kljun et al., 2015). The performance of the Schuepp et al. (1990) approximation was assessed by comparing the area and VI's values obtained with this model and with the 80% flux footprint calculated with the Kljun

et al. (2015) approach in the Chilean sites, where all the required information was available (Figure 2.2).

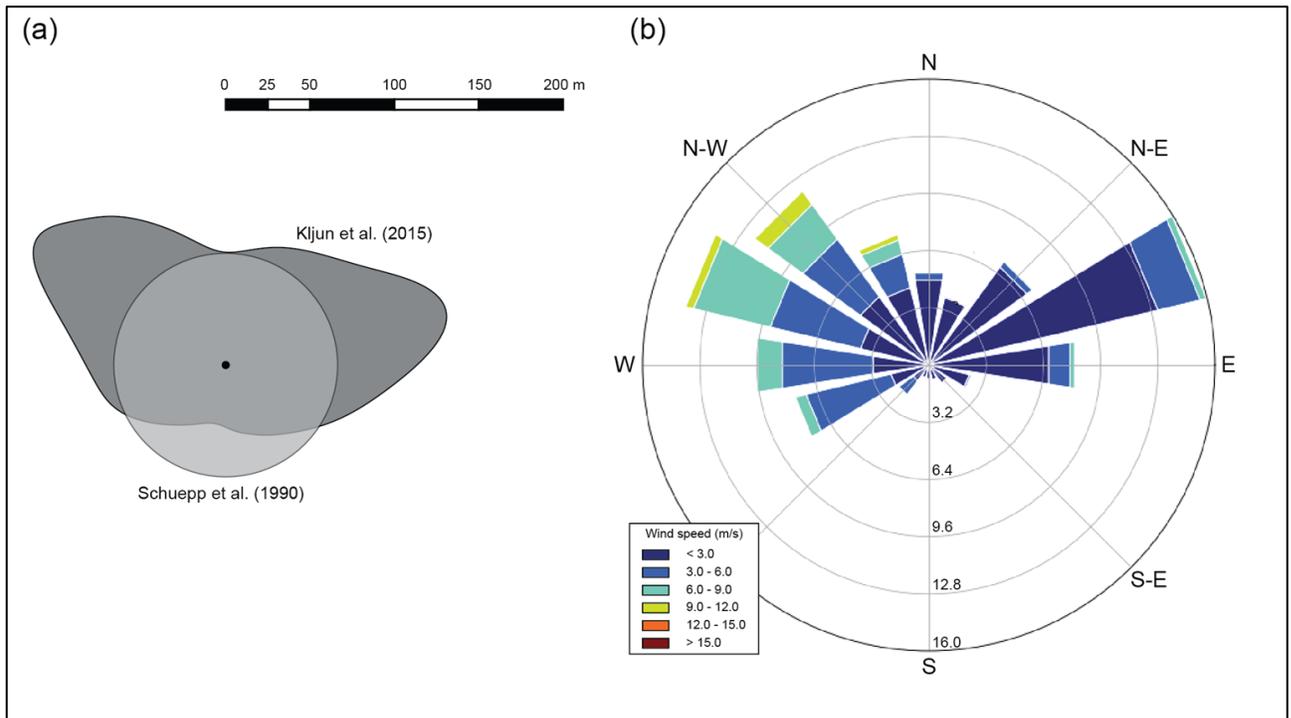


Figure 2.2: (a) Example of footprint calculation with the Schuepp et al. (1990) and the Kljun et al. (2015) approaches. The monthly footprint were calculated at CH-AT3 during November 2019. The Kljun et al. (2015) approach results in a series of irregular but consistent footprints. The Schuepp et al. (1990) approach estimate the monthly footprint as a circle that agrees with the Kljun et al. (2015) footprint in the N-S direction. (b) Wind rose for November 2019 at the CH-AT3, which determines the trend of the Kljun et al. (2015) footprint.

2.3. Remote sensing and vegetation indices

Reflectance images were obtained from the Landsat 7 satellite mission and then analyzed through Google Earth Engine (Gorelick et al., 2017) to estimate different VI's to be incorporated into the ET_a estimates. Every selected image corresponded to the less cloudy image of each month.

For every selected image, the following VI's were calculated: Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), Enhance Vegetation Index (EVI), Normalized Difference Water Index (NDWI) and Normalized Difference Greenness Index (NDGI). Then, the average of each VI was obtained in the footprint area.

The NDVI is the most utilized VI because it is strongly correlated with several biophysical characteristics and physiological processes of plants, including ET (Glenn et al., 2011). NDVI range between -1 to 1, where negative values correspond to water pixels, positives values but near 0 correspond to bare soil and values near 1 are related to dense canopy. NDVI is calculated as (Glenn et al., 2010, 2011):

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$$

where ρ_{NIR} correspond to the reflectance of the NIR band (0.77 - 0.90 μm) and ρ_{RED} (0.63 - 0.69 μm) is the reflectance of the visible red band (USGS, 2019).

SAVI is a VI derived from the NDVI that includes a correction factor L , which minimizes the variations produced by the soil presence in heterogeneous surfaces. This index is calculated as (Glenn et al., 2011):

$$SAVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED} + L} (1 + L)$$

The optimal value of L decreases as vegetation cover increases, i.e., $L=1$ when the density is low, $L = 0.5$ for intermediate vegetation cover and $L=0.25$ for high density. For this investigation $L=0.5$ was used because this value has shown good performance reducing the noise produced by the presence of bare soil in a great range of vegetation cover densities (Odi-Lara et al., 2016) .

The EVI was developed to improve the sensitivity of the signal in high-biomass regions and to reduce the atmosphere influence. EVI responds better than NDVI to structural changes in plants and extend the range over which the NDVI respond to increases in foliage density (Glenn et al., 2010; Yebra et al., 2013). EVI is calculated as:

$$EVI = G_f \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + C1\rho_{RED} + C2\rho_{BLUE} + L_c}$$

where $C1$ and $C2$ are correction coefficients used to account for aerosol resistance, which uses the blue band to correct the influence of the aerosol in the red band. ρ_{BLUE} (0.45 - 0.52 μm) is the reflectance of the blue band (USGS, 2019), G_f is the gain factor (set as 2.5; Glenn et al., 2011) and L_c is the canopy background adjustment (set to 1; Glenn et al., 2011). $C1$ and $C2$ were set as 6 and -7.5, respectively (Glenn et al., 2011).

The NDWI, unlike the others VI's, focuses on identifying trends in the humidity of the studied surface, combining the water content of bare soil and vegetation. The NDWI is defined as (Jovanovic et al., 2014) :

$$NDWI = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}}$$

where ρ_{SWIR} is the reflectance of the short wave infrared (SWIR) band (1.55 - 1.75 μm) (USGS, 2019). Ji et al. (2011) suggested to name this index as Normalized Difference Infrared Index (NDII) because NDWI was first used in MODIS, which SWIR band is between 1.23–1.25 μm . However, in this study I prefer to use NDWI.

Finally, the NDGI is a VI developed to minimize variations between background reflectance of different surfaces and to maximize the contrast between vegetation and other background components, in order to prevent the effects of snow in phenology estimation (Yang et al., 2019). NDGI is calculated as follows (Yang et al., 2019):

$$NDGI = \frac{\varepsilon\rho_{GREEN} + (1 - \varepsilon)\rho_{NIR} - \rho_{RED}}{\varepsilon\rho_{GREEN} + (1 - \varepsilon)\rho_{NIR} + \rho_{RED}}$$

where ρ_{GREEN} is the reflectance of the green band (0.52 – 1.75 μm), and ε is a coefficient that depends on the satellite ($\varepsilon = 0.63$ for Landsat 7).

2.4. Determination of main variables and ET_a estimates using machine learning

The general procedure to generate ET estimates is shown in Figure 2.3. Remote sensing, meteorological and flux data were used as an input in the Exhaustive Feature Selection (EFS) algorithm (Yildirim et al. 2013) to determine the main variables that predicts ET. The EFS algorithm selects the subset of the original features that achieves better an objective, usually finding the high value of a performance metric given by an arbitrary regressor or classifier (Raschka, 2019). The EFS algorithm is the most computationally expensive feature selection method, because its needs to evaluated all possible M-feature combinations of the original N features, with M the number of features that are wanted to be selected (Wang et al., 2016). However, the EFS is the optimal feature selection method when the size of dataset and the number of required features allow this method to be computationally feasible (Schadl et al., 2018) In this research, a subset of 4 features were selected and a maximum of 18 features were evaluated (Table 2.2). The coefficient of determination (R^2) of the linear regression was chosen as the performance metric.

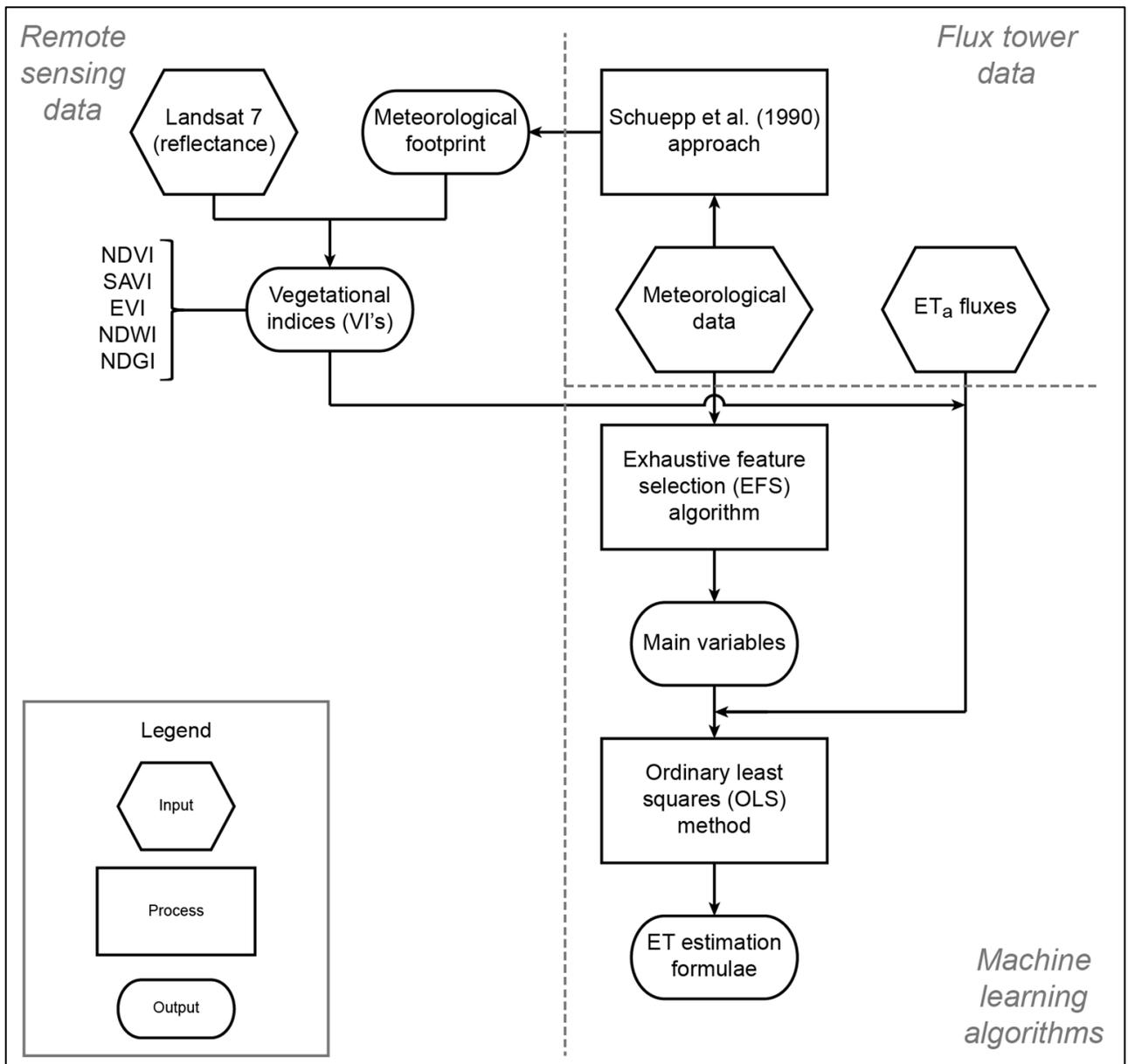


Figure 2.3: Flowchart of methods. Remote sensing, and flux and meteorological data area used to obtain the main variables and ET_a estimate formulas by the application of machine learning algorithms.

Table 2.2: Variables analyzed in this study and its respective units of measure.

Variable	Symbol	Units
Available energy (net radiation minus soil heat flux)	$R_n - G$	MJ m ⁻² d ⁻¹
Precipitation	PPT	mm d ⁻¹ and mm month ⁻¹
Mean temperature	T	°C
Minimum temperature	T_{min}	°C
Maximum temperature	T_{max}	°C
Soil temperature	T_s	°C
Minimum soil temperature	T_{Smin}	°C
Maximum soil temperature	T_{Smax}	°C
Relative humidity	RH	-
Volumetric water content	VWC	-
Vapor pressure deficit	VPD	kPa
Wind speed	WS	m s ⁻¹
Potential evapotranspiration	ET_o	mm d ⁻¹ and mm month ⁻¹
Normalized difference vegetation index	NDVI	-
Soil-adjusted vegetation index	SAVI	-
Enhanced vegetation index	EVI	-
Normalized difference water index	NDWI	-
Normalized difference greenness index	NDGI	-

A linear equation to estimate ET was constructed using the main variables found with the EFS algorithm. The regression coefficients were found with the Ordinary Least Squares method, which minimized the squared distance between the measured data and the estimated line (Stoyan & Baran, 2016). To find the main variables and the regression coefficients, the input data were normalized, i.e., each one of the inputs variables ranged between 0 and 1. This normalization ensured that the EFS chooses the main variables for their contribution to the ET_a variability and not because of its magnitude.

Data were separated into two groups: the training data (CH-AT2, CH-AT, AU-Cpr, US-Cop, US-SRG, US-SRM and US-Wkg) and the validation data (CH-AT1, AU-Ync and US-Wkg). The training data were used to generate a “global” estimation that could fit for all sites. The performance of this equation was evaluated with the validation data. Also, site-specific formulas were constructed with the data of each site. Global and site-specific equations were found for daily and monthly time scales, both of them expressed as mm/day, with only meteorological data. Then, VI’s were incorporated in monthly estimations to evaluate the relevance of incorporating remote sensing data into estimations that consider places with different cover types, but the same climate.

3. RESULTS

3.1. Footprints

Figure 3.1 shows the footprint areas calculated with the Kljun et al. (2015) and the Schuepp et al. (1990) approaches. The Kljun et al. (2015) footprints are larger than the Schuepp et al. (1990) footprints at CH-AT1 and CH-AT3, whereas the opposite is true at CH-AT2. Bigger differences between areas occur in summer months in CH-AT2. For all sites a high correlation was found ($R^2 > 0.84$ and $RMSE < 0.17$) between VI’s values calculated with both footprint approach, despite the difference in the footprints areas (Table 3.1). Hence, even when the Schuepp et al. (1990) approach may not represent precisely the footprint, it allows estimating VI’s that agree with those calculated with a more sophisticated footprint method, such as the Kljun et al. (2015) approach.

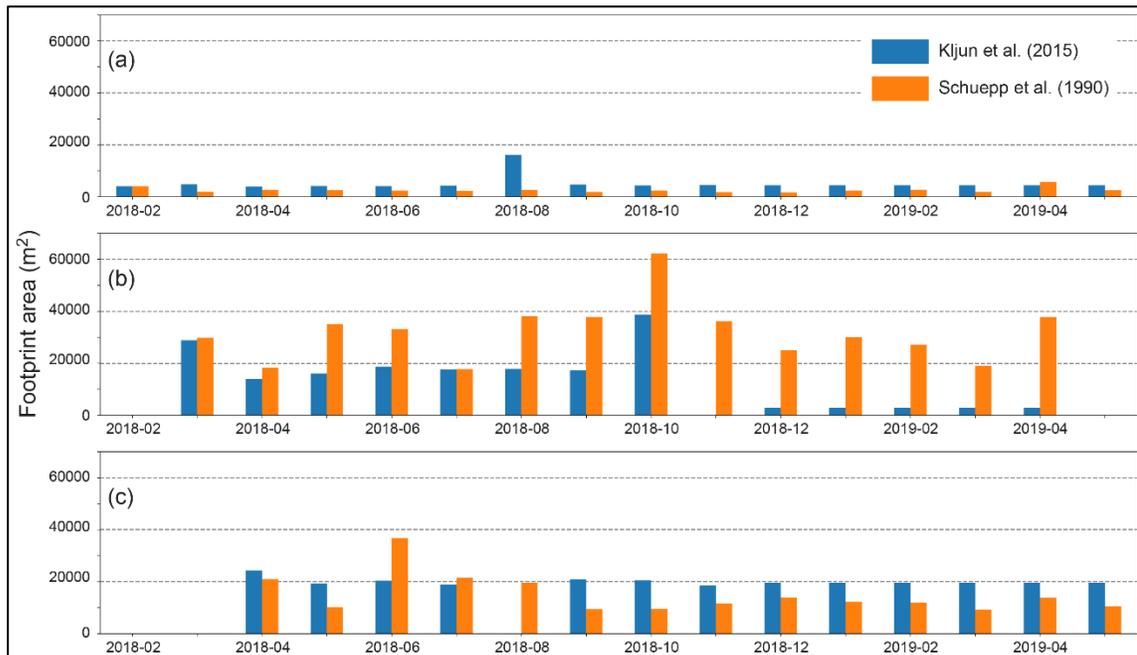


Figure 3.1: Footprints areas calculated with the Kljun et al. (2015) approach and the Schuepp et al. (1990) approach for (a) CH-AT1, (b) CH-AT2 and (c) CH-AT3 sites.

Table 3.1: Comparison between VI's values obtained with the footprints calculated with the Kljun et al. (2015) and the Schuepp et al. (1990) approaches. The mean values of each VI calculated with both approaches in the study period are presented. Also, the R^2 and RMSE of correlations calculated with the Kljun et al. (2015) and the Schuepp et al. (1990) approaches are shown.

		NDVI	SAVI	EVI	NDWI	NDGI
CH-AT1	Kljun mean	0.09	0.14	0.20	-0.03	-0.04
	Schuepp mean	0.08	0.12	0.19	-0.03	-0.04
	R^2	0.91	0.92	0.94	0.88	0.84
	RMSE	0.01	0.02	0.03	0.03	0.01
CH-AT2	Kljun mean	0.04	0.07	0.08	0.00	-0.06
	Schuepp mean	0.05	0.07	0.09	0.01	-0.06
	R^2	0.96	0.96	0.95	0.99	0.93
	RMSE	0.00	0.01	0.01	0.01	0.00
CH-AT3	Kljun mean	0.15	0.22	0.27	0.16	0.00
	Schuepp mean	0.22	0.33	0.41	0.25	0.05
	R^2	0.99	0.99	0.99	0.87	0.99
	RMSE	0.09	0.14	0.17	0.10	0.06

3.2. Remote sensing information

Unlike other studies, in which a high correlation between VI's and ET_a is reported, (Bunting et al., 2014; Groeneveld et al., 2007; Nagler et al., 2013; Yebra et al., 2013), in this research I found that monthly ET_a and VI's are poorly correlated when trying to find a “global” estimation that could be used in all sites. In the case where the “global” monthly ET_a is correlated with VI values, Pearson correlation (corr) ranged between 0.08 and 0.25. However, when the same procedure is performed for each site, the highest corr was of 0.54, which corresponds to NDGI in CH-AT3 (Table 3.2). In many of the cases, VI's have a low variability in relation to the ET_a , with exception of EVI, whose values in some cases exceed the range of -1 and 1 (Figure 3.2). However, when the VI's temporal evolution is studied, it is observed that ET_a peaks are typically accompanied by peaks in the VI's (Figure 3.2, Figure 3.3 and Figure 3.4). However, most important ET_a peaks, which are usually the result of more water availability due to rainfall events, are only accompanied by peaks in EVI and NDWI.

Table 3.2: Pearson correlation between measured ET_a and VI's values.

	NDVI	NDWI	SAVI	EVI	NDGI
Global	0.08	0.21	0.08	0.17	0.25
AU-Cpr	0.38	0.01	0.38	0.21	0.41
AU-Ync	0.37	0.51	0.37	0.19	0.39
CH-AT1	0.41	0.28	0.41	0.44	0.38
CH-AT2	0.01	0.15	0.01	0.05	0.0010
CH-AT3	0.52	0.07	0.5258	0.53	0.54
US-Cop	0.10	0.21	0.10	0.00	0.14
US-SRG	0.02	0.11	0.02	0.07	0.25
US-SRM	0.0111	0.12	0.11	0.10	0.17
US-Whs	0.08	0.24	0.08	0.17	0.12
US-Wkg	0.01	0.32	0.01	0.22	0.25

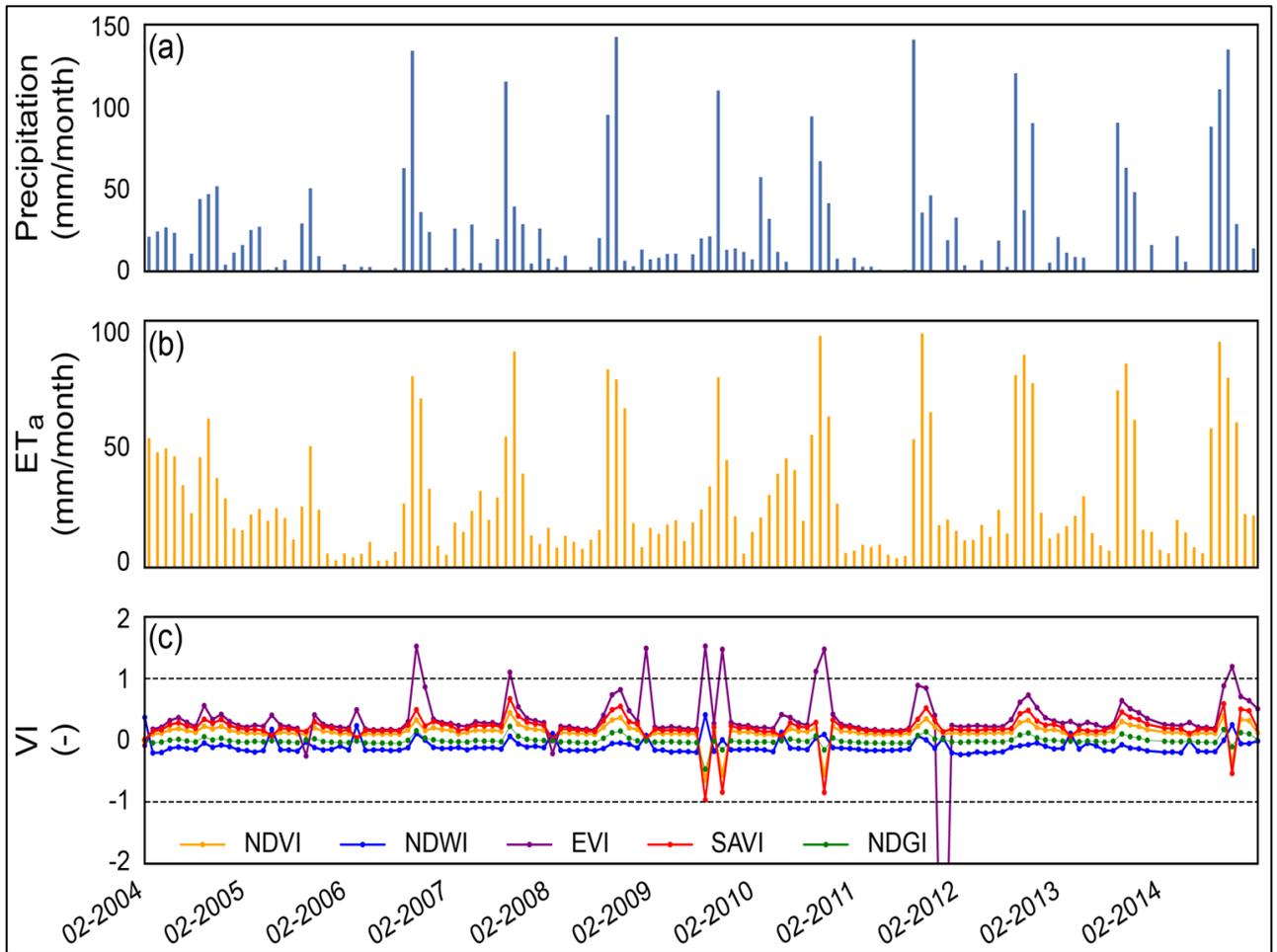


Figure 3.2: Temporal evolution of monthly precipitation (a), monthly ET_a (b), and monthly VI's (c) in US-Wkg

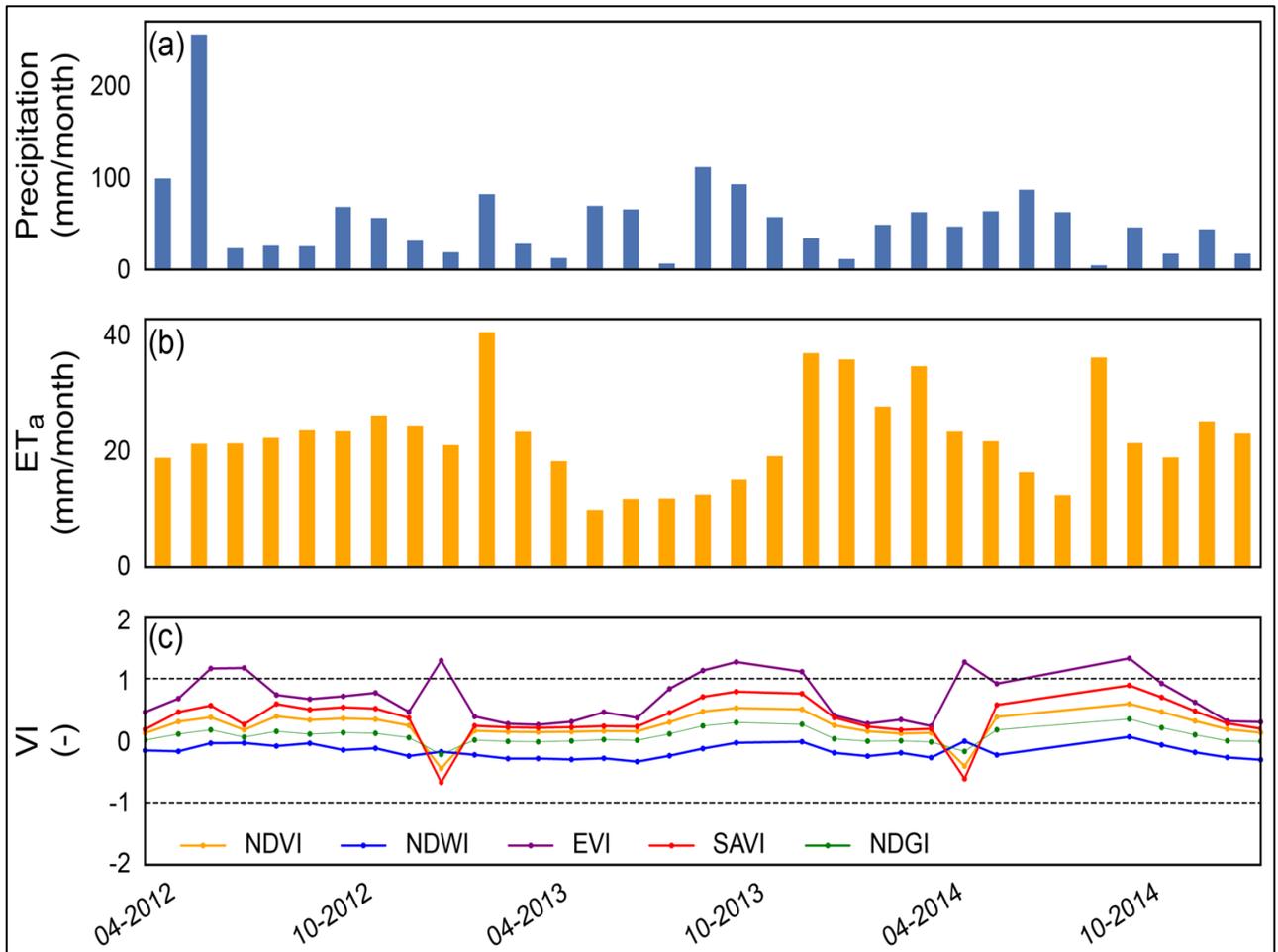


Figure 3.3: Temporal evolution of monthly precipitation (a), monthly ET_a (b), and monthly VI's (c) in AU-Ync.

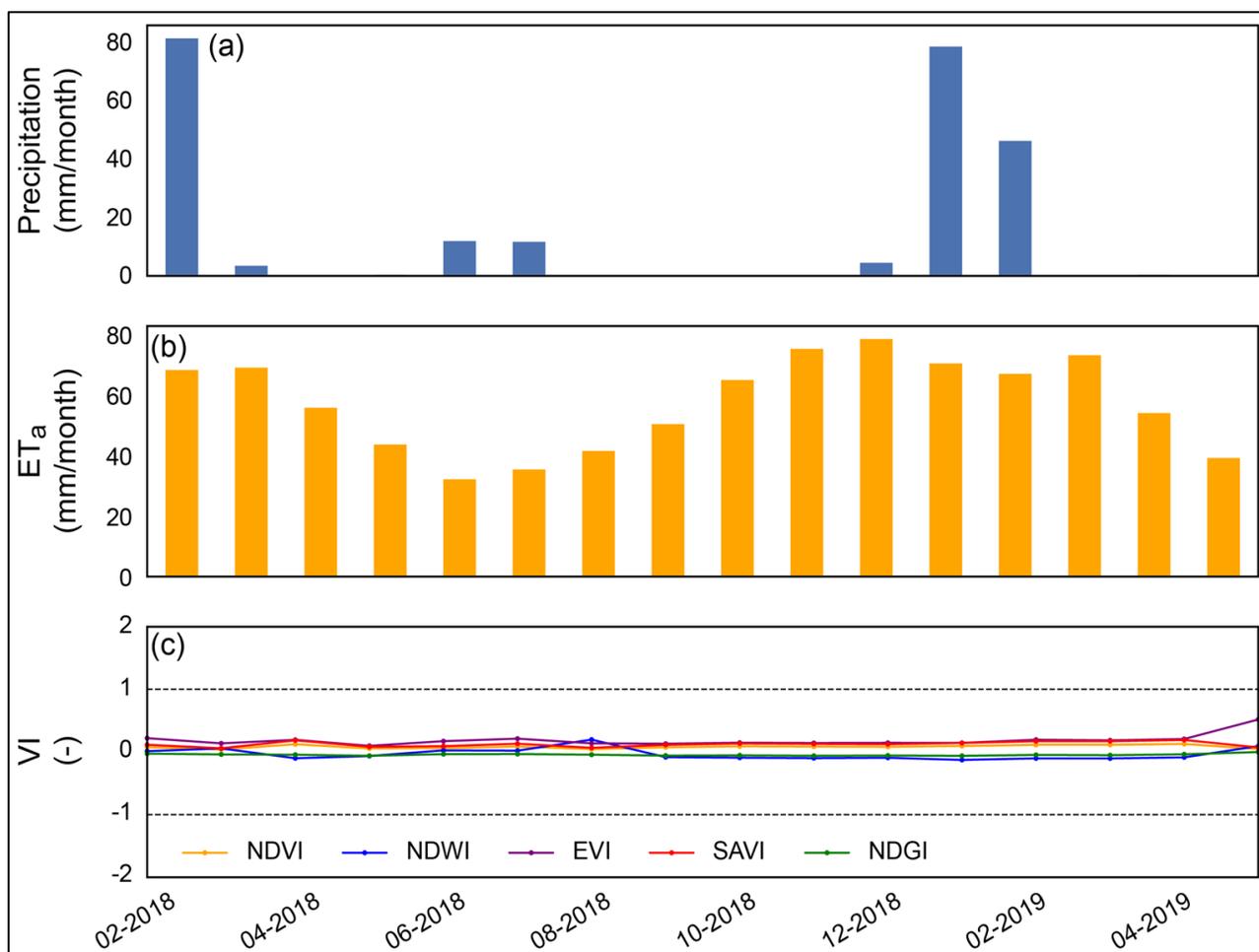


Figure 3.4: Temporal evolution of monthly precipitation (a), monthly ET_a (b), and monthly VI's (c) in CH-AT1.

Figure 3.2 shows the temporal evolution of ET_a, precipitation and VI's on the US-Wkg validation site. Here, ET_a responds to water availability determined by the amount of precipitation. Also, VI's respond in the same way as ET_a, except for SAVI and NDVI, whose values decrease drastically in the presence of precipitation events. This behavior is not common for all the sites, for example, Figure 3.3 shows the temporal evolution of ET_a, precipitation and VI's in the AU-Ync validation site. Here, the relationship between precipitation and ET_a weak. However, it seems that ET_a responds to water availability,

represented as the NDWI. Also, EVI explains some of the ET_a peaks. In the CH-AT1 validation site, no relationship between VI's and ET_a was found (Figure 3.4).

3.3. Variables predicting ET_a

For both the daily and monthly “global” equations, the main variables that influence ET_a are $R_n - G$, ET_o , T_{min} and T_{smax} . In both cases $R_n - G$ is the variable that has the highest regression coefficients, and hence, is the variable that can represent better temporal evolution of ET_a . In the case where remote sensing information is used, the main variables are R_n-G , PPT, NDGI and NDWI. The variable with the greatest contribution to this equation is the NDGI and with the lowest was the R_n-G .

Table 3.3 shows the occurrence of the main variables found for all the sites for the site-specific equations. For the daily estimates, the most important variables are $R_n - G$, VWC and ET_o , which show that daily ET_a depends on both, energy and water availability. In the case of the monthly estimations, the main variables are $R_n - G$, T and Ts . The monthly estimates that includes VI's have VPD and NDWI as the principal variables. Unlike daily site-specific estimates, monthly site-specific estimates do not have a regular behavior; in both cases only two variables are repeated in more than five sites. At daily timescale, the principal variables give information about energy availability, whereas at the monthly timescale, the main variables are related to water availability.

Table 3.3: Number of times that every factor is selected in a site-specific equation for daily, monthly and monthly with VI estimates.

Factor	Daily	Monthly	Monthly with VI
R _n -G	8	7	4
VPD	3	3	5
VWC	6	2	2
RH	1	2	2
T	5	5	4
T _{min}	2	2	1
T _{max}	1	1	0
T _s	3	6	3
T _{Smin}	1	1	0
T _{Smax}	3	1	1
PPT	0	3	4
ET _o	6	4	3
WS	1	3	1
NDVI	-	-	0
NDWI	-	-	8
SAVI	-	-	0
EVI	-	-	0
NDGI	-	-	2

3.4. ET estimate formulas

The daily “global” equation developed with the seven sites of the training data reached an R² of 0.56 and an RMSE of 0.64 mm/day. When this equation is applied to each one of the ten sites, the R² ranges between 0.00 and 0.69, corresponding to the sites CH-AT2 and US-Wkg, respectively. At the same time, at a monthly timescale and only

considering meteorological information, the “global” equation developed with the training sites reaches an R^2 of 0.70 and an RMSE of 0.47 mm/day. For each of the ten sites, the R^2 of this equation varies between 0.00 and 0.82. The monthly “global” equation that does include VI’s reaches an R^2 of 0.67 and an RMSE of 0.49 mm/day (Figure 3.5). When this equation is applied to each one of the ten sites, the R^2 ranges from 0.16 to 0.82, corresponding to the sites AU-Ync and US-Whs, respectively. The linear regression formulas developed with the training data set are shown in **Table 3.4**. In all cases, monthly estimations were more accurate than daily estimations, especially because monthly averages are able to mask outliers. In general, the equation that only considers meteorological information performs better than the equation that includes VI’s. However, when the site-specific equation is applied to each one of the sites, the equation that includes the VI’s results in better outcomes.

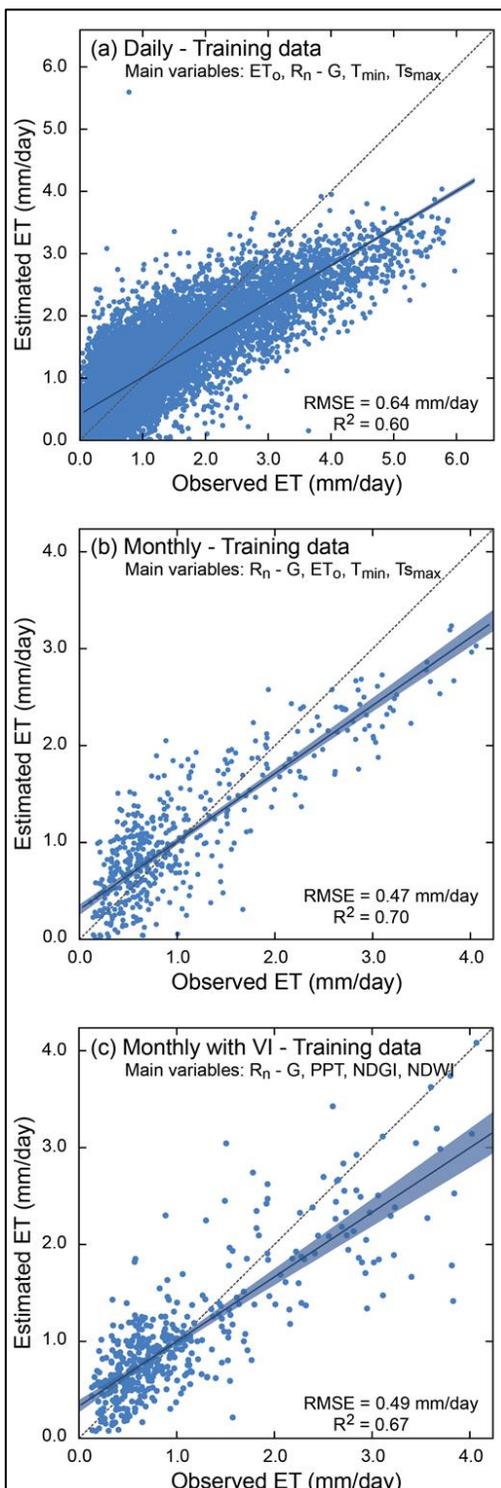


Figure 3.5: Regression formulas for the training data set. (a) Daily estimates; (b) monthly estimates only taking into account meteorological data; (c) monthly estimates considering meteorological data and VI's. Each panel includes the main variable selected for the construction of each formula, the RMSE, and the R^2 .

Table 3.4: Linear regression formulas developed with the training data set. Each variable was previously normalized, consequently they are dimensionless. Original units of measure from each variable are shown in **Table 2.2**.

Type of “global” estimation	Linear regression formulas
Daily	$10.25 R_n - G - 2.86 ET_o + 5.17 T_{min} - 4.17 Ts_{max} - 3.31$
Monthly	$4.26 R_n - G - 3.46 ET_o + 3.21 T_{min} - 2.26 Ts_{max} - 0.07$
Monthly with VI	$1.17 R_n - G + 2.36 PPT + 2.41 NDGI + 1.95 NDWI - 2.14$

In the validation sites, only acceptable results were obtained for the US-Wkg cases, most likely because a large amount of the training data came from a site located near US-Wkg. Figure 3.6 shows the R^2 and RMSE values for the three validation sites. Daily estimations were usually less accurate than monthly estimates, except for the AU-Ync site. Also, estimations that include a VI performed better than those that only considered meteorological information. The case with the best performance corresponds to the monthly estimate that includes a VI in US-Wkg (R^2 of 0.82 and RMSE of 0.42 mm/day).

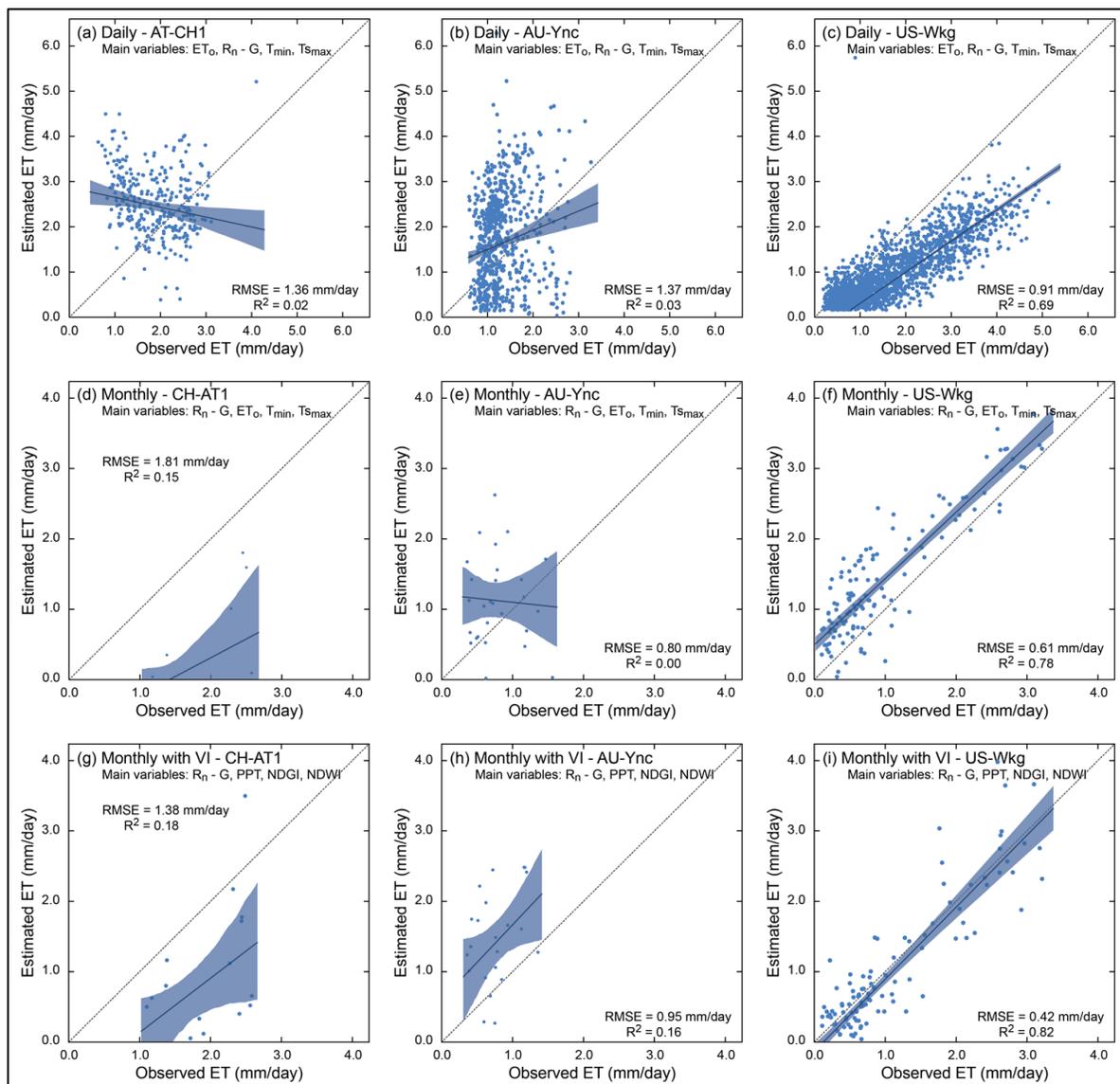


Figure 3.6: Performance of the “global” formulas in the validation sites. CH-AT1, AU-Ync and US-Wkg are shown from left to right. Daily, monthly and VI monthly formulas are shown from the top to the bottom. Each panel includes the main variable selected for the construction of each formula, the RMSE, and R^2 .

4. DISCUSSION

A comparison of ET estimation formulas between studies is difficult due to the many differences between them: (1) calibration and validation procedures; (2) data selection and processing; (3) temporal scale of estimates; (4) number and characteristic of the variables used; and (5) number and location of field sites considered (Yebra et al., 2013). However, in this section, the results obtained in other studies that have used regression formulas or machine learning algorithms to estimate ET_a are discussed and compared to my results. Carter & Liang (2018) evaluated seven regression algorithms for daily ET_a estimations with meteorological and/or remote sensing data of different cover types, reaching R^2 between 0.43 to 0.52 for all sites, similar to the R^2 obtained in this study for daily estimations considering the training data. The algorithms evaluated by Carter & Liang (2018) correspond to simple linear equations, such as the Yebra et al. (2013) formula, to more complex equations, such as that developed by Wang et al. (2010). Granata (2019) fitted three daily ET_a estimations models that include different meteorological data with four different machine learning algorithms in a subtropical humid site located in Florida. All of them reached R^2 values over 0.92. However, better results were obtained in the model with a greater number of variables.

Studies in natural arid zones landscapes are scarce compared to studies performed in agricultural lands located in mesic environments (Mupenzi et al., 2012). Investigations performed in the western of the U.S. have provided the basis for better estimating ET in arid and semi-arid environments (Bunting et al., 2014; Glenn et al., 2013; Jarchowet al., 2017; Nagler et al., 2013; 2009; 2005). Bunting et al. (2014) evaluated three regression equations that estimates ET_a in a period of 16 days in riparian and upland sites in California. One of the equations is a multiple linear regression that includes MODIS EVI and precipitation data ($R^2=0.70$). Nagler et al. (2013; 2005) developed two different regression equations that require meteorological and MODIS EVI information to estimate ET_a in riparian environments of the Colorado, Rio Grande and San Pedro rivers in Colorado, U.S. Both equations are based on the relationship between leaf area index (LAI) and light absorption by the canopy, and the linear relationship between EVI and LAI. Both

equations have good predictive capability ($R^2 = 0.73$ and 0.74 , respectively). Performance of the best results obtained in this research are comparable to the studies reviewed.

Different arid cold climate sites were used to generate linear regression formulas to estimate ET_a . Nevertheless, better results could be reached if only sites with one vegetation cover are selected, instead of one climate type. According to Yebra et al. (2013) a global ET model across land cover types cannot be fitted based on the relationships between observed ET and VI's, as opposed to fitting other variables such as stomatal conductance. However, very different performance in daily, monthly and monthly with VI's estimations were obtained in sites with the same vegetation cover, such as in the case of AU-Ync and US-Wkg, where both of them correspond to grassland ($R^2 = 0.03$, 0.00 , 0.16 and $R^2 = 0.69$, 0.78 , 0.82 , respectively). The linear formulation of the regression formulas should not be an important source of error. Carter & Liang (2018) demonstrated that different regression formulas, with different theoretical bases and same input data, have similar performance.

According to Allen et al. (1998), the main meteorological variables affecting ET_a are radiation, air temperature, air humidity and wind speed. Several researches have studied the relative importance of this variables in ET processes in arid regions. However, these researches normally focused in the behavior of ET_o instead of ET_a , so they do not consider the effects of water stress. For example, Adnan et al. (2017) and Eslamian, et al. (2011) studied the influence of meteorological variables on ET_o estimations in semi-arid, arid and hyper-arid climates (Pakistan and Iran) using the Penman-Monthieth formulation. Both studies concluded that air temperature and humidity are the most important meteorological variables affecting ET_o . One of the few studies that analyze the sensitivity of ET_a estimations to variations in meteorological and remote sensing data in a semi-arid region is that performed by Mokhtari et al. (2013). They analyzed the sensitivity of METRIC (Mapping Evapotranspiration at High Resolution with Internalized Calibration), an ET_a estimation model based in the surface energy balance algorithm for land (SEBAL). Mokhtari et al. (2013) concluded that METRIC is highly sensitive to surface temperature,

net radiation and air temperature, and it is less sensitive to LAI, SAVI, and WS (excepting WS at low level of vegetation cover).

My results indicate that available energy is the main variable that predicts ET_a , which agrees with previous researches that studied ET_a components in most climate and vegetation cover types as possible. For example, Wang et al. (2007) correlated ET_a measurements with radiation, air and land surface temperature, EVI and NDVI, and soil moisture. They concluded that correlation coefficients between R_n and ET are the highest, followed by T 's and VI 's. Carter & Liang (2018) also note that R_n is the most significant variable, which is consistent with the findings of Badgley et al. (2015) and Wang & Liang (2008).

In this research, of the four most important meteorological variables, only wind speed was not decisive to estimate ET_a in any of the cases studied. This fact agree with the findings of Granata (2019), who proved that is possible to generate accurate and precise estimates of daily ET_a through machine learning algorithms only with mean temperature, net solar radiation and relative humidity data, pointing out that the incorporation of wind speed does not improve the ET estimations compared to the case when it is not accounted for. However, he analyzed ET in a subtropical humid climate, where number of sunshine hours is considered to be the more dominant variable, whereas wind speed is an important variable in arid climates (Shahidian et al., 2012; Suárez et al., 2020). Irmak et al. (2008) compared 11 ET models in a crop field in Nebraska, USA, to study their complexity in hourly, daily and seasonal scales. They concluded that wind speed, and other meteorological variables such as temperature, gained importance in daily and hourly calculations, while in seasonal scales radiation is the dominant variable (Shahidian et al., 2012). As shown in these studies, it was expected that wind speed was an important variable in daily ET estimations, though, the method and the number of variables chosen in this research could mask its effects: EFS select the most important meteorological variable or variables that explain ET , in this case R_n-G and $NDWI$, accompanied with variables which their unique objective is to make the equation work numerically; and WS influence could be well represented in ET_o , so R_n-G , VWC and T 's are variables that bring

more information about ET variability than *WS* itself. In arid regions, *WS* has an important role when advection of dry air enhance evaporation and affect the energy balance by horizontal transport of latent heat (Kool, Ben-Gal, & Agam, 2018; Lobos et al., under review.; Suárez et al., 2020). However, when the landscape under study is the same as or similar to the landscape for the surrounding region and experiences similar water inputs, as most of the sites studied in this research with exception of CH-AT1 and CH-AT3, then no advection can occur (Irmak et al., 2012), and other variables are more important predicting ET.

Atmospheric demand and available energy determine ET when water supply is sufficient, whereas soil moisture becomes an important factor predicting ET after soil water supply is deficient (Salvucci, 1997; Wang & Liang, 2008). Bunting et al. (2014) proved that ET estimations in semi-arid upland sites using multiple linear regression improve with the incorporation of a moisture input. However, variables such as precipitation and soil moisture are not usually used for several reasons: (1) surface precipitation and soil moisture measurements are point measurements, limiting the possibilities for upscaling; (2) a lag effect must be considered with precipitation; and (3) soil moisture remote sensing products are difficult to process and its resolution is of several kilometers (Carter & Liang, 2018). In this context, it is expected that in this research few equations incorporated *PPT* and *VWC* in the four most important variables, but when remote sensing information was added several considered NDWI. Unlike others VI's, NDWI is capable to indicate trends in soil and vegetation wetness (Gao, 1996; Sriwongsitanon et al., 2015), so it is a valid water availability input that do not have the same disadvantages of *PPT* and *VWC*, as mentioned above.

The use of remote sensing information is fundamental to estimate ET_a for regional scale and in heterogeneous landscapes (Glenn et al., 2010). This research proved that the incorporation of VI's helps to extrapolate global equations to each one of the sites. However, it has been proven that VI's are not enough to accurately estimate ET_a (Yebra et al., 2013). Carter & Liang (2018) note that, at minimum, ET_a estimates with VI's require

the inclusion of radiation data, although, is preferable to increase the number of input variables. My results demonstrate that acceptable results are achieved with four variables.

Although the contribution of the VI's to the improvement of the ET estimations at the regional level is indisputable, there are several sources of error that must be addressed. One of the most important is the influence of bare soil in the reflectance response, especially in high-resolution satellites, such as Landsat. Jarchow et al., (2018) compared Landsat 5 and Landsat 8 EVI values to MODIS EVI in a riparian zone of the Colorado River, Mexico, finding low correlations over bare soil and sparsely vegetated areas. Also, they suggest being cautious when high-resolution Landsat EVI data are analyzed over heterogeneous areas with low vegetation densities, such as those commonly encountered in arid and semi-arid environments, because soil presence contributes to increased variability in the response of the NIR and red bands.

The low correlations obtained in this study between VI's and ET_a could be explained by several factors. Firstly, as mentioned before, the presence of bare soil can perturb the calculation of VI's (Jarchow et al., 2018). Secondly, in this research only ET_a outliers were extracted, whereas other studies selected data that accomplished some characteristics. For example, Yebra et al. (2013) selected data of days where only transpiration was expected to be dominant, and Scott et al. (2009) excluded data from precipitation events and outliers of meteorological variables. In the presence of important rainfall events, most of the VI's considered in this study, with the exception of NDWI and EVI, have negative values. These values of the VI's indicate that there should be a lower ET rate when it rains, since it actually increases. VI's values obtained in this research are different than those reported in previous studies, probably because the Landsat 7 satellite was used, which is not recommended in arid areas due to its high spatial resolution, and because data was poorly selected (Meyer et al., 2015; Restrepo-Coupe et al., 2016; Scott et al., 2010). However, they are different from each other, highlighting the importance of the satellite selection.

5. CONCLUSIONS

In this study I generated linear regression formulas to estimate daily and monthly ET_a in arid cold sites. Different performances were obtained for every site. However the following trends were identified: (1) better results were obtained for monthly than for daily estimates; (2) incorporation of remote sensing information allows to extrapolate formulas to other sites to get better results than estimations with only meteorological data; (3) the available energy is the most important meteorological variable in ET_a estimations for the sites evaluated in this research; and (4) in arid regions is important to incorporate estimations of water availability. As precipitation and soil moisture are point measurements that do not allow to extrapolate estimations in wide areas, the NDWI could be incorporated as a proxy for water availability in heterogeneous landscapes. Also, more studies that analyze variables predicting ET_a in arid natural landscapes are needed, because ET in drylands is exposed to different factors than in more humid environments, such as water stress, advection and vegetation with adaptations to drought. Global ET_o researches cannot study the complexity of ET_a in arid regions in deep.

REFERENCES

- Adnan, S., Ullah, K., Khan, A. H., & Gao, S. (2017). Meteorological impacts on evapotranspiration in different climatic zones of Pakistan. *Journal of Arid Land*, 9(6), 938–952. <https://doi.org/10.1007/s40333-017-0107-2>
- Allen, R. G., Pereira, L. S., Howell, T. A., & Jensen, M. E. (2011). Evapotranspiration information reporting: I. Factors governing measurement accuracy. *Agricultural Water Management*, 98(6), 899–920. <https://doi.org/10.1016/j.agwat.2010.12.015>
- 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. *Irrigation and Drainage*, 1–300. <https://doi.org/10.1016/j.eja.2010.12.001>
- Badgley, G., Fisher, J. B., Jiménez, C., Tu, K. P., & Vinukollu, R. (2015). On uncertainty in global terrestrial evapotranspiration estimates from choice of input forcing datasets. *Journal of Hydrometeorology*, 16(4), 1449–1455. <https://doi.org/10.1175/JHM-D-14-0040.1>
- Beringer, J., & Walker, J. (2015). *FLUXNET2015 AU-Ync Jaxa*. <https://doi.org/10.18140/FLX/1440208>
- Bisquert, M., Sánchez, J. M., & Caselles, V. (2016). Evaluation of Disaggregation Methods for Downscaling MODIS Land Surface Temperature to Landsat Spatial Resolution in Barrax Test Site. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(4), 1430–1438. <https://doi.org/10.1109/JSTARS.2016.2519099>
- Bowling, D. (2015). *FLUXNET2015 US-Cop Corral Pocket*. <https://doi.org/10.18140/FLX/1440100>
- Bowling, D. R., Grote, E. E., & Belnap, J. (2011). Rain pulse response of soil CO₂ exchange by biological soil crusts and grasslands of the semiarid Colorado Plateau, United

States. *Journal of Geophysical Research: Biogeosciences*, 116(3).
<https://doi.org/10.1029/2011JG001643>

Bunting, D. P., Kurc, S. A., Glenn, E. P., Nagler, P. L., & Scott, R. L. (2014). Insights for empirically modeling evapotranspiration influenced by riparian and upland vegetation in semiarid regions. *Journal of Arid Environments*, 111, 42–52.
<https://doi.org/10.1016/j.jaridenv.2014.06.007>

Carter, C., & Liang, S. (2018). Comprehensive evaluation of empirical algorithms for estimating land surface evapotranspiration. *Agricultural and Forest Meteorology*, 256–257(April), 334–345. <https://doi.org/10.1016/j.agrformet.2018.03.027>

Centro de Ciencias del Clima y la Resiliencia. (2019). *Explorador climático*. Retrieved from <http://explorador.cr2.cl/>

DHI-GRAS. (2020). *User Manual for SEN-ET SNAP Plugin*.

Dirección General de Aguas [DGA]. (2004). Diagnóstico y clasificación de los cuerpos de agua según objetivos de calidad. Cuenca Salar de Atacama. In *Cade-Idepe Consultores en Ingeniería. Dirección General de Aguas. Ministerio de Obras Públicas. Gobierno de Chile*.

El Masri, B., Rahman, A. F., & Dragoni, D. (2019). Evaluating a new algorithm for satellite-based evapotranspiration for North American ecosystems: Model development and validation. *Agricultural and Forest Meteorology*, 268(June 2017), 234–248.
<https://doi.org/10.1016/j.agrformet.2019.01.025>

Eslamian, S., Khordadi, M. J., & Abedi-Koupai, J. (2011). Effects of variations in climatic parameters on evapotranspiration in the arid and semi-arid regions. *Global and Planetary Change*, 78(3–4), 188–194. <https://doi.org/10.1016/j.gloplacha.2011.07.001>

Gao, B.-C. (1996). NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 72(2)(April 1995), 257–266.

Gaur, M. K., & Squires, V. R. (2018). Geographic Extent and Characteristics of the World's Arid Zones and Their Peoples. In M. Gaur & V. Squires (Eds.), *Climate Variability Impacts on Land Use and Livelihoods in Drylands* (pp. 3–20). Cham, Switzerland: Springer.

Glenn, E. P., Mexicano, L., Garcia-Hernandez, J., Nagler, P. L., Gomez-Sapiens, M. M., Tang, D., Lomeli, M.A., Ramírez-Hernández, J. & Zamora-Arroyo, F. (2013). Evapotranspiration and water balance of an anthropogenic coastal desert wetland: Responses to fire, inflows and salinities. *Ecological Engineering*, 59, 176–184. <https://doi.org/10.1016/j.ecoleng.2012.06.043>

Glenn, E. P., Nagler, P. L., & Huete, A. R. (2010). Vegetation Index Methods for Estimating Evapotranspiration by Remote Sensing. *Surv Geophys*, 31, 531–555. <https://doi.org/10.1007/s10712-010-9102-2>

Glenn, E. P., Neale, C. M. U., Hunsaker, D. J., & Nagler, P. L. (2011). Vegetation index-based crop coefficients to estimate evapotranspiration by remote sensing in agricultural and natural ecosystems. *Hydrological Processes*, 25(26), 4050–4062. <https://doi.org/10.1002/hyp.8392>

Google Earth. (n.d.). *Photography of Monticello, Utah, USA*. Retrieved from https://lh5.googleusercontent.com/p/AF1QipOyCbXpgbxq-mCTVhM_JbJBKDJ5WPuaKfQnmEPo=h1440

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). *Google Earth Engine: Planetary-scale geospatial analysis for everyone*.

- Granata, F. (2019). Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agricultural Water Management*, 217(August 2018), 303–315. <https://doi.org/10.1016/j.agwat.2019.03.015>
- Groeneveld, D. P., Baugh, W. M., Sanderson, J. S., & Cooper, D. J. (2007). Annual groundwater evapotranspiration mapped from single satellite scenes. *Journal of Hydrology*, 344(1–2), 146–156. <https://doi.org/10.1016/j.jhydrol.2007.07.002>
- Hu, G., Jia, L., & Menenti, M. (2015). Comparison of MOD16 and LSA-SAF MSG evapotranspiration products over Europe for 2011. *Remote Sensing of Environment*, 156, 510–526. <https://doi.org/10.1016/j.rse.2014.10.017>
- Huang, C., Li, Y., Gu, J., Lu, L., & Li, X. (2016). Estimating regional evapotranspiration under water limited conditions based on SEBS and modi data in arid regions. *IGARSS*, 238–241.
- Huete, A. R. (2012). Vegetation Indices, Remote Sensing and Forest Monitoring. *Geography Compass*, 6(9), 513–532. <https://doi.org/10.1111/j.1749-8198.2012.00507.x>
- Irmak, A., Allen, R. G., Kjaersgaard, J., Huntington, J., Kamble, B., Trezza, R., & Ratcliffe, I. (2012). Operational Remote Sensing of ET and Challenges. In A. Irmak (Ed.), *Evapotranspiration - Remote Sensing and Modeling* (pp. 467–492). <https://doi.org/10.5772/25174>
- Irmak, S., Istanbuluoglu, E., & Irmak, A. (2008). An Evaluation of Evapotranspiration Model Complexity Against Performance in Comparison with Bowen Ratio Energy Balance Measurements. *Science and Public Policy*, 51(4), 1–8.
- Jarchow, C. J., Didan, K., Barreto-Muñoz, A., Nagler, P. L., & Glenn, E. P. (2018). Application and comparison of the MODIS-derived enhanced vegetation index to VIIRS, landsat 5 TM and landsat 8 OLI platforms: A case study in the arid colorado river delta, Mexico. *Sensors (Switzerland)*, 18(5). <https://doi.org/10.3390/s18051546>

- Jarchow, C. J., Nagler, P. L., Glenn, E. P., Ramírez-Hernández, J., & Rodríguez-Burgueño, J. E. (2017). Evapotranspiration by remote sensing: An analysis of the Colorado River Delta before and after the Minute 319 pulse flow to Mexico. *Ecological Engineering*, *106*, 725–732. <https://doi.org/10.1016/j.ecoleng.2016.10.056>
- Ji, L., Zhang, L., Wylie, B. K., & Rover, J. (2011). On the terminology of the spectral vegetation index (NIR - SWIR)/(NIR+SWIR). *International Journal of Remote Sensing*, *32*(21), 6901–6909. <https://doi.org/10.1080/01431161.2010.510811>
- Jovanovic, N, Garcia, C., Bagan, R., Teich, I., & Rodriguez, C. (2014). Validation of remotely-sensed evapotranspiration and NDWI using ground measurements at Riverlands, South Africa. *Water SA*, *40*(2), 211. <https://doi.org/10.4314/wsa.v40i2.3>
- Jovanovic, Nebo, & Israel, S. (2012). Critical Review of Methods for the Estimation of Actual Evapotranspiration in Hydrological Models. In A. Irmak (Ed.), *Evapotranspiration - Remote Sensing and Modeling* (pp. 329–350). Retrieved from <http://www.intechopen.com/books/evapotranspiration-remote-sensing-and-modeling/critical-review-of-methods-for-the-estimation-of-actual-evapotranspiration-in-hydrological-models>
- Kljun, N., Calanca, P., Rotach, M. W., & Schmid, H. P. (2015). A simple two-dimensional parameterisation for Flux Footprint Prediction (FFP). *Geoscientific Model Development*, *8*(11), 3695–3713. <https://doi.org/10.5194/gmd-8-3695-2015>
- Kool, D., Ben-Gal, A., & Agam, N. (2018). Within-field advection enhances evaporation and transpiration in a vineyard in an arid environment. *Agricultural and Forest Meteorology*, *255*, 104–113. <https://doi.org/10.1016/J.AGRFORMET.2017.10.018>
- Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, *15*(3), 259–263. <https://doi.org/10.1127/0941-2948/2006/0130>

- Leclerc, M. Y., Foken, T., Savage, M. J., & Göckede, M. (2014). Footprints in micrometeorology and ecology. In *Footprints in Micrometeorology and Ecology*. <https://doi.org/10.1007/978-3-642-54545-0>
- Lobos Roco, F., Hartogensis, O., Vilà-Guerau de Arellano, J., de la Fuente, A., Muñoz, R., Rutllant, J., & Suárez, F. Local evaporation controlled by regional atmospheric circulation in the Altiplano of the Atacama. *Under review in Journal of Geophysical Research: Atmospheres*.
- Louridas, P., & Ebert, C. (2016). Machine Learning. *IEEE SOFTWARE*, 33(5), 110–115. <https://doi.org/10.1109/MS.2016.114>
- Mata-González, R., McLendon, T., & Martin, D. W. (2005). The inappropriate use of crop transpiration coefficients (Kc) to estimate evapotranspiration in arid ecosystems: A review. *Arid Land Research and Management*, 19(3), 285–295. <https://doi.org/10.1080/15324980590951469>
- Meijninger, W. M. L., Beyrich, F., Ludi, A., Kohsiek, W., & Bruin, H. a R. (2004). Scintillometer Fluxes of Sensible and Latent Heat over a Heterogeneous area – A Contribution to Litfass-2003. *Boundary-Layer Meteorology*, 121(August 2015), 89–110. <https://doi.org/10.1007/s10546-005-9022-8>
- Meyer, W., Cale, P., Koerber, G., Ewenz, C., & Sun, Q. (2015). *FLUXNET2015 AU-Cpr Calperum*. <https://doi.org/10.18140/FLX/1440195>
- Meyer, W. S., Kondrlovà, E., & Koerber, G. R. (2015). Evaporation of perennial semi-arid woodland in southeastern Australia is adapted for irregular but common dry periods. *Hydrological Processes*, 29(17), 3714–3726. <https://doi.org/10.1002/hyp.10467>
- Mokhtari, M. H., Ahmad, B., Hoveidi, H., & Busu, I. (2013). Sensitivity analysis of METRIC-based evapotranspiration algorithm. *International Journal of Environmental Research*, 7(2), 407–422.

Mu, Q., Zhao, M., & Running, S. W. (2013). MODIS Global Terrestrial Evapotranspiration (ET) Product (MOD16A2/A3). *Algorithm Theoretical Basis Document, Collection*, 66.

Mupenzi, J. de la P., Li, L., Ge, J., Ngamije, J., Achal, V., Habiyaremye, G., & de Dieu Habumugisha, J. (2012). Water losses in arid and semi-arid zone: Evaporation, evapotranspiration and seepage. *Journal of Mountain Science*, 9(2), 256–261. <https://doi.org/10.1007/s11629-012-2186-z>

Nagler, P., Glenn, E., Nguyen, U., Scott, R., & Doody, T. (2013). Estimating Riparian and Agricultural Actual Evapotranspiration by Reference Evapotranspiration and MODIS Enhanced Vegetation Index. *Remote Sensing*, 5(8), 3849–3871. <https://doi.org/10.3390/rs5083849>

Nagler, P. L., Glenn, E. P., & Huete, A. R. (2001). Assessment of spectral vegetation indices for riparian vegetation in the Colorado River delta, Mexico. *Journal of Arid Environments*, 49(1), 91–110. <https://doi.org/10.1006/jare.2001.0844>

Nagler, P. L., Morino, K., Murray, R. S., Osterberg, J., & Glenn, E. P. (2009). An empirical algorithm for estimating agricultural and riparian evapotranspiration using MODIS enhanced vegetation index and ground measurements of ET. I. Description of method. *Remote Sensing*, 1(4), 1273–1297. <https://doi.org/10.3390/rs1041273>

Nagler, P., Scott, R., Westenburg, C., Cleverly, J. R., Glenn, E. P., & Huete, A. R. (2005). Evapotranspiration on western U.S. rivers estimated using the Enhanced Vegetation Index from MODIS and data from eddy covariance and Bowen ratio flux towers. *Remote Sensing of Environment*, 97, 337–351. <https://doi.org/10.1016/j.rse.2005.05.011>

Nieto, H., Guzinski, R., Sandholt, I., Karamitilios, G., Rasmussen, M. O., Bellvert, J., Jofré, C., Hansen, L.B., Sølvsteen, J.B., Bondo, T. & Koetz, B. (2019). Prototype Benchmarking and Description - ESA Sentinels for Evapotranspiration (Sen-et). (4000121772), 1–31.

Odi-Lara, M., Campos, I., Neale, C. M. U., Ortega-Farías, S., Poblete-Echeverría, C., Balbontín, C., & Calera, A. (2016). Estimating evapotranspiration of an apple orchard using a remote sensing-based soil water balance. *Remote Sensing*, 8(3). <https://doi.org/10.3390/rs8030253>

Prikaziuk, E., & van der Tol, C. (2019). Global sensitivity analysis of the SCOPE model in Sentinel-3 Bands: Thermal domain focus. *Remote Sensing*, 11(20). <https://doi.org/10.3390/rs11202424>

Raschka, S. (2019). Exhaustive Feature Selector. mlxtend documentation. Retrieved from http://rasbt.github.io/mlxtend/user_guide/feature_selection/ExhaustiveFeatureSelector/

Restrepo-Coupe, N., Huete, A., Davies, K., Cleverly, J., Beringer, J., Eamus, D., van Gorsel, E., Hutley, L. B. & Meyer, W. S. (2016). MODIS vegetation products as proxies of photosynthetic potential along a gradient of meteorologically and biologically driven ecosystem productivity. *Biogeosciences*, 13(19), 5587–5608. <https://doi.org/10.5194/bg-13-5587-2016>

Salvucci, G. D. (1997). Soil and moisture independent estimation of stage-two evaporation from potential evaporation and albedo or surface temperature. *Water Resources Research*, 33(1), 111–122. <https://doi.org/10.1029/96WR02858>

Schadl, K., Vassar, R., Cahill-Rowley, K., Yeom, K. W., Stevenson, D. K., & Rose, J. (2018). Prediction of cognitive and motor development in preterm children using exhaustive feature selection and cross-validation of near-term white matter microstructure. *NeuroImage: Clinical*, 17(November 2017), 667–679. <https://doi.org/10.1016/j.nicl.2017.11.023>

Schuepp PH, Leclerc MY, MacPherson JI, D. R. (1990). Footprint prediction of scalar fluxes from analytical solutions of the diffusion equation. *Boundary-Layer Meteorology*, 50, 355–373.

SCM El Abra. (2016). *Estudio de tasa de crecimiento de humedales*. Centro de ecología aplicada.

Scott, R. (2012). *US-SRG site*. Retrieved from <https://ameriflux.lbl.gov/sites/siteinfo/US-SRG>

Scott, R. (2014). *US-SRM site*. Retrieved from <https://ameriflux.lbl.gov/sites/siteinfo/US-SRM#image-gallery>

Scott, R. (2015a). *FLUXNET2015 US-SRG Santa Rita Grassland*. <https://doi.org/10.18140/FLX/1440114>

Scott, R. (2015b). *FLUXNET2015 US-SRM Santa Rita Mesquite*. <https://doi.org/10.18140/FLX/1440090>

Scott, R. (2015c). *FLUXNET2015 US-Whs Walnut Gulch Lucky Hills Shrub*. <https://doi.org/10.18140/FLX/1440097>

Scott, R. (2015d). *FLUXNET2015 US-Wkg Walnut Gulch Kendall Grasslands*. <https://doi.org/10.18140/FLX/1440096>

Scott, R. (2015e). *US-Wkg site*. Retrieved from <https://ameriflux.lbl.gov/sites/siteinfo/US-Wkg#image-gallery>

Scott, R. (2017). *US-Whs site*. Retrieved from <https://ameriflux.lbl.gov/sites/siteinfo/US-Whs#image-gallery>

Scott, R., Biederman, J. A., Hamerlynck, E. P., & Barron-Gafford, G. A. (2015). The carbon balance pivot point of southwestern U.S. semiarid ecosystems: Insights from the 21st century drought. *Journal of Geophysical Research: Biogeosciences*, *120*(12), 2612–2624. <https://doi.org/10.1002/2015JG003181>

Scott, R., Hamerlynck, E. P., Jenerette, G. D., Moran, M. S., & Barron-Gafford, G. A. (2010). Carbon dioxide exchange in a semidesert grassland through drought-induced

vegetation change. *Journal of Geophysical Research: Biogeosciences*, 115(3), 1–13. <https://doi.org/10.1029/2010JG001348>

Scott, R., Jenerette, G. D., Potts, D. L., & Huxman, T. E. (2009). Effects of seasonal drought on net carbon dioxide exchange from a woody-plant-encroached semiarid grassland. *Journal of Geophysical Research: Biogeosciences*, 114(4), 1–14. <https://doi.org/10.1029/2008JG000900>

Scott, R. L., Jenerette, G. D., Potts, D. L., & Huxman, T. E. (2009). Effects of seasonal drought on net carbon dioxide exchange from a woody-plant-encroached semiarid grassland. *Journal of Geophysical Research: Biogeosciences*, 114(4), 1–13. <https://doi.org/10.1029/2008JG000900>

Shahidian, S., Serralheiro, R., Serrano, J., Teixeira, J., Haie, N., & Santos, F. (2012). Hargreaves and Other Reduced-Set Methods for Calculating Evapotranspiration. In A. Irmak (Ed.), *Evapotranspiration - Remote Sensing and Modeling* (pp. 59–80).

Sriwongsitanon, N., Gao, H., Savenije, H. H. G., Maekan, E., Saengsawang, S., & Thianpopirug, S. (2015). The Normalized Difference Infrared Index (NDII) as a proxy for soil moisture storage in hydrological modelling. *Hydrology and Earth System Sciences Discussions*, 12(8), 8419–8457. <https://doi.org/10.5194/hessd-12-8419-2015>

Sriwongsitanon, N., Nuchanart, Gao, H., Savenije, H. H. G., Maekan, E., Saengsawang, S., & Thianpopirug, S. (2016). Comparing the Normalized Difference Infrared Index (NDII) with root zone storage in a lumped conceptual model. *Hydrology and Earth System Sciences*, 20(8), 3361–3377. <https://doi.org/10.5194/hess-20-3361-2016>

Stoyan, G., & Baran, A. (2016). *Elementary Numerical Mathematics for Programmers and Engineers*. <https://doi.org/10.1007/978-3-319-44660-8>

Suárez, F., Lobos, F., de la Fuente, A., Vilà-Guerau de Arellano, J., Prieto, A., Meruane, C., & Hartogensis, O. (2020). E-DATA: A Comprehensive Field Campaign to Investigate

Evaporation Enhanced by Advection in the Hyper-Arid Altiplano. *Water*, 12(3), 745. <https://doi.org/10.3390/w12030745>

TERN. (2017a). *AU-Cpr site*. Retrieved from http://www.ozflux.org.au/monitoringsites/calperum/calperum_pictures.html

TERN. (2017b). *AU-Ync site*. Retrieved from http://www.ozflux.org.au/monitoringsites/yanco/yanco_pictures.html

Torres, A. F., Walker, W. R., & McKee, M. (2011). Forecasting daily potential evapotranspiration using machine learning and limited climatic data. *Agricultural Water Management*, 98(4), 553–562. <https://doi.org/10.1016/j.agwat.2010.10.012>

USGS. (2019). Landsat 7 (L7) Data Users Handbook. In *Department of the Interior, U.S. Geological Survey*. Retrieved from https://landsat.usgs.gov/sites/default/files/documents/LSDS-1927_L7_Data_Users_Handbook.pdf

Wang, K., Dickinson, R., Wild, M., & Liang, S. (2010). Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2001: 1. Model development. *Journal of Geophysical Research*, 115. <https://doi.org/10.1029/2009JD013671>

Wang, K., & Liang, S. (2008). An improved method for estimating global evapotranspiration based on satellite determination of surface net radiation, vegetation index, temperature, and soil moisture. *Journal of Hydrometeorology*, 9(4), 712–727. <https://doi.org/10.1175/2007JHM911.1>

Wang, K., Wang, P., Li, Z., Cribb, M., & Sparrow, M. (2007). A simple method to estimate actual evapotranspiration from a combination of net radiation, vegetation index, and temperature. *Journal of Geophysical Research Atmospheres*, 112(15), 1–14. <https://doi.org/10.1029/2006JD008351>

Wang, L., Wang, Y., & Chang, Q. (2016). Feature selection methods for big data bioinformatics: A survey from the search perspective. *Methods*, *111*, 21–31. <https://doi.org/10.1016/j.ymeth.2016.08.014>

Yang, W., Kobayashi, H., Wang, C., Shen, M., Chen, J., Matsushita, B., Tang, Y., Kim, Y., Bret-Harte, M.S., Zona, D., Oechel, W. & Kondoh, A. (2019). A semi-analytical snow-free vegetation index for improving estimation of plant phenology in tundra and grassland ecosystems. *Remote Sensing of Environment*, *228*(March), 31–44. <https://doi.org/10.1016/j.rse.2019.03.028>

Yebera, M., Van Dijk, A., Leuning, R., Huete, A., & Guerschman, J. P. (2013). Evaluation of optical remote sensing to estimate actual evapotranspiration and canopy conductance. *Remote Sensing of Environment*, *129*, 250–261. <https://doi.org/10.1016/j.rse.2012.11.004>

Yee, M. S., Pauwels, V. R. N., Daly, E., Beringer, J., Rüdiger, C., McCabe, M. F., & Walker, J. P. (2015). A comparison of optical and microwave scintillometers with eddy covariance derived surface heat fluxes. *Agricultural and Forest Meteorology*, *213*, 226–239. <https://doi.org/10.1016/J.AGRFORMET.2015.07.004>

Yildirim, A. A., Özdoğan, C., & Watson, D. (2013). Parallel data reduction techniques for big datasets. In *Big Data Management, Technologies, and Applications* (pp. 72–93). <https://doi.org/10.4018/978-1-4666-4699-5.ch004>

APPENDIXES

1. APPENDIX A: SITES DESCRIPTION

CH-AT1 (Figure A. 1): this site corresponds to a riparian wetland located in the Chilean Andean Plateau (22.02°S, 68.05°W, elevation: 4182 m.a.s.l.). The annual precipitation is concentrated in summer months due to the effect of the South American monsoon and is of ~78 mm (2007-2016 time period), whereas annual mean temperature is of ~5.8° C (1969-1987 time period) (Centro de Ciencias del Clima y la Resiliencia, 2019). The area is dominated by the presence of the reed *Oxychloe andina* and a grass *Deyeuxia sp.* The *Parastrephia sp.* shrub and some hydrophytes, such as *Lilaeopsis macloviana* and *Myriophyllum quitense*, are also present.



Figure A. 1: the CH-AT1 site

CH-AT2 (Figure A. 2): this site is located 1500 m north of AT-CH1 (22.00°S, 68.05°W, elevation: 4330 m.a.s.l.). Unlike the riparian wetland, only grass and some shrubs are present at this site, where the dominant species is a grass of *Festuca* genera. Because CH-AT1 and CH-AT2 are near to each other, the climate characterization of CH-AT2 is the same as in the riparian wetland (CH-AT1).



Figure A. 2: the CH-AT2 site

CH-AT3 (Figure A. 3): this site is the Putana wetland, which is located in the Altiplano of the Antofagasta Region, Chile (22.55°S, 68.02°W, elevation: 4255 m.a.s.l). The annual precipitation is of ~106 mm (2008-2017 time period), also concentrated in the summer months, and the mean annual temperature is of ~ 1.7° C (2013-2016 time period) (Centro de Ciencias del Clima y la Resiliencia, 2019; Dirección General de Aguas [DGA], 2004). The presence of water in the wetland is due to contributions of the Putana River and groundwater upwelling (SCM El Abra, 2016). The vegetation in the study site is classified as perennial grassland dominated by *Oxychloe andina* and some grass of the *Festuca* and *Deyeuxia* genera. There are also some hydrophytes, such as *Ranunculus uniflorus* and *Azolla filiculoides*.



Figure A. 3: the CH-AT3 site

AU-Cpr (Figure A. 4): This study site is located 25 km north of Renmark in South Australia at Calperum Station (34.00°S, 140.59°E, elevation: ~166 m.a.s.l.). The mean annual precipitation is approximately 250 mm. More rainfall is generally expected in the cooler winter and spring periods, but occasional summer rainfall events occur. Mean annual temperature is 18° C ranging between -3 and 45°C. The vegetation is dominated by several species of *Eucalyptus*, but also it is possible to find mid-storey species belonging to *Eremophila*, *Hakea*, *Olearia*, *Senna* and *Melaleuca* genera (W. S. Meyer et al., 2015).



Figure A. 4: the Au-Cpr site (TERN, 2017a)

AU-Ync (Figure A. 5): The site is located in the Yanco Study Area (35.00°S, 146.29°E, elevation: ~125 m), which is situated within the western plains of the Murrumbidgee River catchment, in New South Wales, Australia. Precipitation is distributed evenly across all months reaching 419 mm per year. Daily mean temperatures vary significantly from 34° C in January to 14.2° C in July. The site consists of a homogeneous flat grassland that is used for the grazing of livestock. The grassland is dominated by perennial tussock grasses, such as kangaroo and wallaby grasses (Yee et al., 2015).



Figure A. 5: the AU-Ync site (TERN, 2017b)

US-Cop (Figure A. 6): this site, named Corral Pocket, is a semiarid grassland located in southeast Utah, USA (38.09°N, 109.39°W, elevation: 1520 m.a.s.l.). Mean annual precipitation and temperature are 216 mm and 12° C, respectively. About 33% of the precipitation occurs during summer. The vegetation is dominated by the perennial *Hilaria jamesii* and *Stipa hymenoides* bunch-grasses and the *Coleogyne ramosissima* shrub, with other grasses and annuals making up a small percentage of total plant cover (Bowling et. al., 2011).



Figure A. 6: Monticello, 3km apart of US-Cop site (Google Earth, n.d.).

US-SRG (Figure A. 7): this site correspond to Santa Rita Grassland, wich is located in the Santa Rita Experimental Range, 45 km south of Tucson, Arizona, USA (31.79°N, 110.83°W, elevation: 1290 m.a.s.l.). Mean annual precipitation is 377 mm. Due to the effect of the North American monsoon, about 50% of rainfall occurs during summer. Mean air temperature is 19°C, with ranges that produce winter freezes in November and daytime maxima that exceed 35° in June (Scott, et al., 2009; 2015). This site is dominated by the South African warm season bunchgrass, Lehmann Lovegrass (*Eragrostis lehmanniana*) and it has a 11% cover of mesquite (*Prosopis velutina*) (Scott et al., 2015).



Figure A. 7: the US-SRG site (Scott, 2012)

US-SRM (Figure A. 8): this site correspond to the Santa Rita mesquite savanna site, which is also located on the Santa Rita Experimental Range, USA, 5km apart of Santa Rita Mesquite site (31.82°N, 110.87°W, elevation: 1116 m.a.s.l.). Site vegetation consist of the leguminous tree *Prosopis velutina* (35% of the vegetation cover) growing in a matrix of native and nonnative perennial grasses, subshrubs and scattered succulents (R. Scott et al., 2009).



Figure A. 8: the US-SRM site (Scott, 2014)

US-Whs (Figure A. 9): this site correspond to the Lucky Hills Shrubland, in the U.S. Department of Agriculture Agricultural Research Service (USDA-ARS) Walnut Gulch Experimental Watershed. It is located 80 km east of Santa Rita sites (31.74°N, 110.05°W,

elevation: 1370 m.a.s.l.). Annual precipitation is lower than that in the Santa Rita sites, reaching 285 mm. The mean air temperature is also quite lower, reaching 17,6°C. This site has a large diversity of shrubs that are typically found throughout the Sonoran and Chihuahuan Deserts, such as *Parthenium incanum*, *Acacia constricta*, *Larrea tridentata*, and *Flourensia cernua* (R. Scott et al., 2015).



Figure A. 9: the US-Whs site (R. Scott, 2017)

US-Wkg (Figure A. 10): this site correspond to the Walnut Gulch Kendall Grasslands, which is located 10 km apart of US-Whs, also in the USDA-ARS Walnut Gulch Experimental Watershed (31-74°N, 109.94°W, elevation: 1530 m.a.s.l.). In the period of 2005 to 2014 a mean annual temperature of 17.3°C and an annual precipitation of 294 mm have been reported. The dominant species are *Eragrostis lehmanniana*, *Bouteloua eripoda*, and *Aristida spp*, all of them belonging to the *Poaceae* family. It is also possible to see woody species as *Ephedra viridis* and *Artemisa filifolia* (Scott et al., 2015).



Figure A. 10: the US-Wkg site (Scott, 2015).