

# PROTOTYPE EVALUATION

V1.3  
February 8, 2019

Contract No.  
4000121772/17/I-NB “SENTINELS FOR EVAPOTRANSPIRATION [SEN-ET]”

Submitted by



In cooperation with:





**DOCUMENT RELEASE SHEET**

Authors:	Héctor Nieto (IRTA), Radoslaw Guzinski (DHI), Inge Sandholt (SANDHOLT APS), Georgios Karamitilios (SANDHOLT APS), Mads Olander Rasmussen (DHI), Joaquim Bellvert (IRTA), Christian Jofré (IRTA), Lars Boye Hansen (DHI), Jonas Blüthgen Sølvsteen (DHI), Torsten Bondo (DHI), Benjamin Koetz (ESA)
Approval (internal)	Mads Olander Rasmussen
Approval (ESA)	Benjamin Koetz
Distribution	ESA and partners

**CHANGE RECORD**

Date	Page(s)	Change record
30.11.2018	–	Draft version sent to partners
21.12.2018	–	First version sent to ESA
07.01.2019	3	Updated author list
09.01.2019	–	Fixed values of % points column in tables 1.3-1.8
23.01.2019	26, 34	Additional validation tables separating herbaceous from woody crops
03.02.2019	59 - 70	Added text, tables and figures to Chapter 2: Model benchmarking and selection
08.02.2019	Table 1.1	Added crop information for Klingenberg and Selhausen

This document has been prepared under the DHI Business Management System certified by Bureau Veritas to comply with ISO 9001 (Quality Management)





# Contents

<b>1</b>	<b>High resolution evapotranspiration using Sentinel-2 and 3: Evaluation of METRIC, TSEB and ESVEP models with thermal sharpening methods</b>	<b>9</b>
1.1	Introduction . . . . .	9
1.2	Materials and Methods . . . . .	12
1.2.1	Description of ET models . . . . .	12
1.2.2	Validation sites . . . . .	15
1.2.3	Input data sources . . . . .	17
1.2.4	Thermal data sharpening approaches . . . . .	20
1.3	Results . . . . .	21
1.4	Discussions . . . . .	28
1.4.1	ET model intercomparison . . . . .	28
1.4.2	Sharpening and disaggregation . . . . .	29
1.4.3	Effects of ancillary inputs . . . . .	30
1.5	Conclusions . . . . .	31
<b>2</b>	<b>Model benchmarking and selection</b>	<b>59</b>
2.1	Benchmarking criteria . . . . .	59
2.2	Benchmarking results . . . . .	60
2.2.1	Accuracy . . . . .	60
2.2.2	Applicability . . . . .	62
2.2.3	Data Accessibility . . . . .	63
2.2.4	Functionality . . . . .	67
2.2.5	Performance . . . . .	68
2.2.6	Other considerations . . . . .	69
2.2.7	Overall assessment and model selection . . . . .	69
<b>3</b>	<b>Detailed model implementations</b>	<b>71</b>
3.1	Net radiation . . . . .	71
3.2	METRIC . . . . .	73
3.2.1	Automatic search of cold and hot pixels . . . . .	74
3.3	TSEB . . . . .	74
3.4	ESVEP . . . . .	77
3.5	sharpening models . . . . .	77
<b>4</b>	<b>Data preprocessing details</b>	<b>79</b>
<b>5</b>	<b>Scientific questions to be addressed in Phase 2, and beyond</b>	<b>81</b>

**6 Suggestions for future ESA thermal mission candidate****83**

# Preface

This Deliverable (D.3.1) will evaluate the performance of the three evapotranspiration models and two thermal sharpening methods that were previously selected during the review of existing remote sensing ET models (D.1.1, [1]). Part of this deliverable is intended to be submitted for a scientific publications and thus it is presented in manuscript format, as much as possible, in Chapter 1. The following chapters describe in more detail the processing of Sentinel and ancillary data (Ch. 4) and model prototypes (3). Finally we added a chapter proposing future scientific activities/projects that could be carried out in Phase 2 and/or after SET-ET project (Ch. 5) and recommendations for a possible future high-resolution thermal ESA mission (Ch. 6)



## Chapter 1

# High resolution evapotranspiration using Sentinel-2 and 3: Evaluation of METRIC, TSEB and ESVEP models with thermal sharpening methods

### 1.1 Introduction

The fluxes of water (e.g. evapotranspiration - ET) and energy (e.g. of latent and sensible heat) at the surface of the Earth are critical to quantify for many applications in the fields of climatology, meteorology, hydrology and agronomy. Easy access to reliable estimations of ET is considered a key requirement within natural resource management, and if ET can be estimated accurately enough it holds a vast potential to assist in the current attempts of meeting the UN Sustainable Development Goals (SDG), e.g. SDG2 – zero hunger, or SDG6 – clean water and sanitation (<https://sustainabledevelopment.un.org>, last accessed 10 December 2018). Water and energy fluxes show large spatio-temporal variability since they are highly dependent not only on the meteorological conditions, but also on different characteristics and properties of the land surface, such as soil moisture/water availability, land cover type and amount of vegetation biomass and its health. Remote sensing data can provide spatially-distributed information about relevant land surface states and properties used to model the relevant fluxes and hence this technology addresses a key limitation of conventional point scale observations when estimating fluxes at watershed and regional scales. In particular, thermal remote sensing has been widely used for assessing land surface turbulent fluxes [2]. While there are a variety of existing remote sensing ET methods and data options available [3], none is fully satisfying the user needs for reliable, operational and easy accessible estimates and tools able to derive ET at agricultural-parcel scale. The limitations have so far primarily been centred on the lack of suitable satellite-based input data sources.

With the recent launch of Sentinel 2 and 3, the data foundation for producing operational ET maps has been set [4]. Coupled with extensive existing research, and not least the maturity of open source software, the reliability and accessibility of ET estimates can now be increased substantially. The combined use of Sentinel 3 and 2 constellation of satellites provides a unique opportunity for providing operational estimates of ET. Sentinel 3 SLSTR

instrument acquires daily thermal infrared (TIR) information of the surface at ca. 1 km scale. However, the reliable estimation of ET in agricultural and heterogeneous landscapes requires that the model's spatial resolution matches the dominant landscape feature scale, usually tens or hundreds of meters. Sentinel 2, with a spatial resolution ranging from 10 to 60 m and 5 day revisit time with Sentinel 2A & B combined, can resolve part of these scaling issues, although it lacks a TIR instrument at high spatial resolution such as in the Landsat missions. Therefore sharpening [5, 6, 7] and/or disaggregation methods [8] are required to bridge the spatial gap between the currently available Sentinel constellation's thermal and optical observational capabilities in order to optimally exploit the synergies of both types of sensors for field-scale ET estimations.

Several data fusion methods have been proposed to merge low resolution thermal infrared imagery with high resolution optical imagery in order to obtain estimates of surface temperature ( $T_{rad}$ ) and/or ET at high spatial resolution. In this study we focus on different, but possibly complementary, approaches: empirical and semi-empirical methods that exploit relationships between optical bands and thermal or ET data (hereinafter called image sharpening methods); and physically-based ET downscaling methods (hereinafter called ET disaggregation). Thermal image sharpening uses information from the thermal and optical images themselves to calibrate empirical or semi-empirical models. Those models relate coarse resolution  $T_{rad}$  (or ET) with coarse resolution (or fine resolution aggregated to coarse resolution) optical bands, and then apply the calibrated model to the fine scale optical image, producing either a sharpened  $T_{rad}$ , or directly an ET product. This image sharpening approach relies on the direct or indirect relationship that different regions of the optical spectrum have with the radiometric temperature and/or the ET process. For instance the temperature of denser canopies, with higher contrast between visible and near-infrared bands, is lower than the temperature of bare soils [9, 10]. On the other hand, surfaces with higher water content (i.e. larger absorption in the short-wave infrared) have a larger evaporative capability and hence lower temperature [11]. Likewise, higher chlorophyll concentrations (i.e. larger absorption in the red and red-edge regions) might lead to higher light and water use efficiency and hence lower temperatures.

One of the first attempts to sharpen  $T_{rad}$  was TsHARP [12], in which tested different regression models between  $T_{rad}$  and NDVI. Since then, TsHARP has been utilised as reference method for developing and testing other sharpening methods [6, 13, 14]. The Data Mining Sharpening (DMS) approach [6] used local and global regression trees between reflective bands and  $T_{rad}$  of homogeneous samples at coarse scale (based on coefficient of variation threshold). Residual analysis was performed to ensure energy conservation (based on emitted radiances) between original resolution and sharpened images. To avoid overfitting of regression trees such as in DMS the use of random forests was proposed instead [15]. Following with the machine learning algorithms, Yang et al. [16] used an Artificial Neural Network with Genetic Algorithm and Self-Organizing Feature Mapping trained with different land surface parameters for each land cover class (vegetation, bare soil, urban and water). The use of a contextual algorithm can also be applied in sharpening, such as is the case of DISPATCH-LST (DISaggregation based on Physical And Theoretical scale CHange). Merlin et al. [5] which used optical information on fractional vegetation cover and fractional photosynthetically active vegetation cover in contextual scatterplots of fractional green vegetation cover versus  $T_{rad}$  and albedo versus  $T_{rad}$  to define minimum and maximum soil and canopy endmember temperatures. Finally, two or more different

methods can be used together and combined through weighted averaging, such as in Chen et al. [17], who combined TsHARP and a Thin Plate Spline interpolation by weighting their corresponding residuals. Besides of the fact that all methods described above can be used as well to sharpen ET, other studies have already suggested methods to directly downscale coarse scale ET using optical data [18, 19, 20, 21]. In any case, optical images provide limited information related to some surface energy balance processes, such as turbulent transport, soil moisture, and meteorological forcing. Therefore ancillary variables could be included in  $T_{rad}$  or ET sharpening such as land cover maps (to account for different aerodynamic roughness), local meteorology, or surface geometry [22]. Nevertheless, a step that was found to significantly enhance the accuracy of the fluxes derived with sharpened  $T_{rad}$ , was to use a “disaggregation” approach [23, 8] to ensure spatial consistency between fluxes at fine and coarse spatial scales [4]. In that approach, the fluxes are first estimated at the coarse scale at which the thermal observations were acquired. Afterwards, all the high-resolution flux pixels falling within one low-resolution flux pixel are adjusted (by varying the low-resolution air temperature) until there is a consistency between the two scales. This is done under the assumption that the coarse scale estimates are more accurate since they are derived with the  $T_{rad}$  with original spatial resolution. This disaggregation approach usually improves the results compared to when applying the ET models with either coarse or fine resolution alone [23, 24, 4].

The latent heat flux  $\lambda E$  (or ET) can be estimated as the residual of the surface energy budget, using estimates of the net radiation ( $R_n$ ), soil heat flux ( $G$ ) and sensible heat flux ( $H$ ), assuming other energy terms (heat advection  $A$ , heat storage in the canopy layer  $S$ , and energy for photosynthesis  $\lambda_c F$ ) are negligible. The thermal-based ET models were originally formulated for computing  $H$ , which is governed by the bulk resistance equation for heat transfer [25], and is driven by the gradient between an ensemble surface temperature, called the “aerodynamic surface temperature” ( $T_0$ ), and the surface layer air temperature. Besides of the estimation of that surface-to-air temperature gradient, the estimation of  $H$  requires the modelling of an aerodynamic resistance term, which can be viewed as a simplification of the complex turbulent transport of heat, momentum and water vapour, by using a similarity with Ohm’s law for electric transport. These resistances therefore represent how efficiently a scalar (heat, momentum or water vapour) is transported from one point to another following a gradient (i.e. vertical differences of temperature and/or vapour pressure). Several formulations and/or parametrizations have been proposed to describe these turbulent transport processes but generally they include variables related to surface aerodynamic roughness, wind speed as well as wind attenuation through the canopy, and atmospheric stability [26]

The challenge in resistance energy balance models is that  $T_0$  cannot be directly estimated by remote sensing [27, 28]. Hence, remote sensing ET models differ from each other on how the existing difference between the radiometric temperature ( $T_{rad}$ ) observed by satellite sensors and  $T_0$  is considered. Single-source or bulk transfer schemes for modelling  $H$  treat soil and canopy as a single flux source and often employ an additional resistance term ( $R_{AH}$ , usually dependent on the Stanton number  $kB^{-1}$ ) because heat transport is less efficient than momentum transport from land surface (see e.g. Garratt and Hicks [29] or Verhoef et al. [30]). Appropriately calibrated, one-source energy balance (OSEB) models have shown satisfactory estimates of surface energy fluxes in heterogeneous landscapes [31, 32, 33, 34]. However, due to the difficulty in robustly and parsimoniously parametrizing

$R_{ex}$  for OSEB schemes at different landscapes, climates, and observational configurations [35], the two-source energy balance (TSEB) modelling approach was developed [36]. TSEB models partition the surface energy fluxes and the radiometric temperature between nominal soil and canopy sources, and include a more physical representation of processes related to  $T_{rad}$  and  $T_0$  without requiring any additional input information beyond that needed by single-source models using more sophisticated  $kB^{-1}$  parametrizing. However, because direct measurements of canopy ( $T_C$ ) and soil ( $T_S$ ) temperatures rarely are available, in most applications these component temperatures are derived from a measurement of the bulk surface radiometric temperature  $T_{rad}$ . Partitioning of  $T_{rad}$  between  $T_C$  and  $T_S$  requires some assumptions related to the evaporative efficiency of soil or canopy [36, 37, 38].

Finally, like all remote sensing retrievals, satellite radiometric temperature is prone to uncertainty due to sensor noise, surface emissivity and atmospheric effects. To overcome this issue in ET estimation, several methods have been proposed based on either contextual models [39, 40, 41], by constraining the ET range between hot (no ET) and cold (potential ET) pixels [31, 32], or using time-differenced morning temperature rise [42, 43]. Regarding the contextual methods, all of them require homogeneous forcing and coupling between land surface/atmosphere which is a disadvantage when applied at large scales. In addition, those models assume that the coldest pixel in the image means potential transpiration, and the hottest pixel means zero transpiration which is not always the case (e.g. in humid and sub-humid areas).

In this study we will evaluate three different ET models with Sentinel imagery: METRIC [32] is a one source energy balance model that is less sensitive to heat transfer coefficient parametrizing than other OSEB model such as SEBS [33]; TSEB-PT [36] as a widely used two source energy balance model; and ESVEP [44] as a hybrid contextual-two source energy balance model.

## 1.2 Materials and Methods

### 1.2.1 Description of ET models

The energy balance can be expressed as (1.1)

$$R_n \approx G + H + \lambda E \quad (1.1)$$

where net radiation  $R_n$  is a key element in the energy budget of the land surface as it determines the available energy that the land utilises for water evapotranspiration (latent heat flux,  $\lambda E$ ) and for heating up the overlying air layer (sensible heat flux,  $H$ ). Since ET is the combined process of soil evaporation and canopy transpiration,  $R_n$  can be also be partitioned into soil ( $R_{n,S}$ ) and canopy net radiation ( $R_{n,C}$ ), with both sensible and latent heat flux also partitioned between soil (i.e. evaporation process) and canopy (transpiration).

Using remote sensing data to derive  $R_n$  has proven to be a sound alternative to ground-based measurements. Different approaches have been proposed to estimate surface albedo, ranging from empirical relationships between ground measured albedo and the different reflective bands in satellite [45] to more physically based methods relying on modeling the surface anisotropic effects [46, 47]. Indeed, one of the major challenges when estimating albedo with satellite remote sensing data is that such sensors typically measure the outgoing radiance at a given direction while the estimation of albedo needs to account for the

outgoing radiance in all the directions of the hemisphere [48, 49]. Methods based on the modelling of those bidirectional effects have proven to be effective to overcome this challenge. Based on the different spectral behaviour of soil and vegetation for the visible (PAR) and near infrared regions (NIR) of the spectrum, Kustas and Norman [50] proposed a method for retrieving soil and canopy shortwave net radiation. Such approach is based on the radiative transfer model (RTM) described in Campbell and Norman [51] to obtain estimates of soil and canopy albedo as well as canopy transmittance in the PAR and NIR. This approach requires as inputs Leaf Area Index and leaf inclination distribution [52], the different bihemispherical reflectances and transmittances of soil and a single leaf, and the proportion of diffuse radiation. However, this approach assumes homogeneous canopies and it requires some corrections when dealing with clumped canopies [53]. On the other hand, longwave net radiation is primarily driven by the thermal radiation emitted by the surface, which depends on surface emissivity and skin temperature following the Stefan-Boltzman law. Besides, Kirchoff's law can be applied to derive the atmospheric longwave radiation that is absorbed by the surface. As in modelling shortwave net radiation, surface anisotropy can also be considered in estimating the net longwave radiation, considering that leaves and soil have different temperatures and hence emit different amounts of thermal radiation [50].

Soil heat flux  $G$  is usually assumed to be a ratio of the soil net radiation. Choudhury et al. [54], Bastiaanssen et al. [31] suggested that  $G$  is ca. 35% the radiation absorbed by the soil around midday hours. However, in cases when remote sensing data is acquired earlier in the morning or later in the afternoon, the assumption of a constant ratio between  $G$  and  $R_{n,S}$  is less reliable [55], and curve that is function of time is usually applied instead [37].

### Mapping EvapoTranspiration at high Resolution with Internalized Calibration, METRIC

Sensible heat flux in METRIC is derived in a contextual manner by finding hot and cold pixels (Eq. 1.2).

$$H = \rho C_p \frac{\delta T}{R_{AH}} \quad (1.2a)$$

$$\delta T = c + m T_{rad} \quad (1.2b)$$

where  $\delta T$  is the estimated gradient between aerodynamic and air temperature, estimated as a linear equation function of  $T_{rad}$  with  $c$  and  $m$  parameters are linearly solved from expressing Eq. 1.2b from two cold and hot endpoints:

$$m = \frac{\delta T_{hot} - \delta T_{cold}}{T_{hot} - T_{cold}} \quad (1.3a)$$

$$c = \delta T_{hot} - m T_{hot} \quad (1.3b)$$

METRIC scales  $\lambda E$  between these two hot ( $T_{hot}$ ) and cold ( $T_{cold}$ ) endmembers based on a linear relationship between actual ET and reference ET using the standardised ASCE Penman-Monteith equation for an ideal alfalfa field [56]. Therefore, METRIC, as opposed to SEBAL [31], does not assume zero sensible heat flux at the cold pixel, which can have

a positive impact at well watered areas under large vapour pressure deficit conditions. According to Allen et al. [32], cold pixels yield a 5% larger ET than the reference ET ( $\lambda E_{cold} = 1.05\lambda E_{ref}$ ), but earlier in the season and off-season, cold pixel ET is instead a function of fractional cover or NDVI:  $\lambda E_{cold}/\lambda E_{ref} = f(NDVI)$ . On the other hand, METRIC overcomes the issue of estimating  $kB^{-1}$  by computing  $R_{AH}$  using the profile at two different heights above  $z_{0H}$ . Finally the authors stated the need for either computing an “excess resistance” in aerodynamically rough and dry surfaces when using the  $\delta T$  calibration performed over agricultural areas, or calibrating different  $\delta T$  slopes at different land covers/environmental conditions [32].

For Eq. 1.2 to hold true,  $\delta T$  and  $H$  require constant wind speed at the application domain, so the model uses wind speed at blending height to overcome this issue. It also requires constant irradiance and air temperature, i.e.  $\delta T$  changes are only either due to soil moisture or aerodynamic roughness. Furthermore, the model requires heterogeneity in hydrologic and vegetation conditions and therefore we applied METRIC over two different vegetation domains, short vegetation (crops, grass and shrubs) and tall vegetation (broadleaved and conifer forests as well as wooded savannas). Finally, METRIC is sensitive to the definition of hot and cold pixels. Several different methodologies to find those endmember values were proposed, which can be especially challenging in heterogeneous areas where pixels become mixed at coarse spatial resolution. In our case we adopted the Exhaustive Search Algorithm solution proposed by Bhattarai et al. [57]. More details on METRIC and its implementation in this study can be found in Section 3.2 of Chapter 3.

### Priestley-Taylor Two-Source Energy Balance model, TSEB-PT

Two-source energy balance models treat the land surface as two layers, soil and canopy, contributing to the energy and water fluxes (Eq. 1.4)

$$R_{n,C} = H_C + \lambda E_C R_{n,S} = H_S + \lambda E_S + G \quad (1.4)$$

where soil (canopy) sensible heat flux is computed from the gradient between the soil (canopy) temperature ( $T_S$  and  $T_C$  respectively) and the air temperature at the sink-source height (equivalent to  $T_0$ ). Since  $T_C$  and  $T_S$  are unknown *a priori*, they are estimated in an iterative process in which it is first assumed that green canopy (expressed as the fraction of LAI that is green,  $f_g$ ) transpires a potential rate based on Priestley–Taylor formulation ( $\lambda E_C = \alpha_{PT} f_g \frac{\Delta}{\Delta + \gamma} R_{n,veg}$ ,  $\alpha_{PT} = 1.26$ ) [36]. Then the canopy transpiration is sequentially reduced (i.e.  $\alpha_{PT} < 1.26$ ) until realistic fluxes are obtained ( $\lambda E_C \geq 0$  and  $\lambda E_S \geq 0$ )

TSEB-PT probably is the model that requires most accurate retrievals of physical inputs ( $LAI$  and  $T_{rad}$ ), and studies already reported larger uncertainty in senescent vegetation and very dense (high  $LAI$ ) or tall vegetation [43, 58]. It is more complex than METRIC and therefore has a large number of parameters and modelling options. Finally, the Priestley–Taylor formulation was shown to produce larger uncertainty in highly advection conditions, cases in which initializing  $\lambda E_C$  with a Penman-Monteith formulation showed better results [37]. More details on TSEB-PT and its implementation in this study can be found in Section 3.3 of Chapter 3.

## End-member-based Soil and Vegetation Energy Partitioning, ESVEP

ESVEP is based on a trapezoid  $T_{rad} - f_{cover}$  framework, in which it considers fluxes acting in a “parallel” soil and canopy system [44]. As in TSEB-PT, ESVEP partitions  $T_{rad}$  as a linear weight of emitted radiance. Other similar models to ESVEP are HTEM [59] and TTEM [60], but ESVEP solves the trapezoid in a pixel-per-pixel basis overcoming the need for homogeneous weather forcing and roughness (Eq. 1.5a).

$$T_{S,max} = \frac{r_a (R_{n,soil} - G)}{\rho_a C_p} + T_A \quad (1.5a)$$

$$T_{C,max} = \frac{r_a R_{n,veg}}{\rho_a C_p} \frac{\gamma (1 + r_{b,dry}/r_a)}{\Delta + \gamma (1 + r_{b,dry}/r_a)} - \frac{vpd}{\Delta + \gamma (1 + r_{b,dry}/r_a)} + T_A \quad (1.5b)$$

$$T_{S,min} = \frac{r_a (R_{n,soil} - G)}{\rho_a C_p} \frac{\gamma}{\Delta + \gamma} - \frac{vpd}{\Delta + \gamma} + T_A \quad (1.5c)$$

$$T_{C,min} = \frac{r_a R_{n,veg}}{\rho_a C_p} \frac{\gamma (1 + r_{b,wet}/r_a)}{\Delta + \gamma (1 + r_{b,wet}/r_a)} - \frac{vpd}{\Delta + \gamma (1 + r_{b,wet}/r_a)} + T_A \quad (1.5d)$$

More details on ESVEP and its implementation in this study can be found in Section 3.4 of Chapter 3.

### 1.2.2 Validation sites

Eleven eddy covariance EC sites were used in this study, covering a wide range of land cover types and climates. Sites are summarised in table 1.1 and data used in validation included four component of radiation  $R_n$  (shortwave/longwave downwelling/upwelling), soil heat flux  $G$ , sensible heat flux  $H$ , and latent heat flux  $\lambda E$ . In addition, friction velocity, Monin-Obukhov length, and wind direction data from the EC system was used to estimate the satellite pixel footprint contribution [61, 62] to the turbulent fluxes at the satellite overpass. Validation sites comprise 5 agricultural sites, both irrigated and rainfed, including row crops (e.g. vineyard) and an olive grove. In addition, two sites over grassland, one humid meadow (Hyltemossa) and a semi-arid steppe (Walnut Gulch), two on conifer and one on broadleaved forests are also included in the validation list. Finally, complex heterogeneous landscapes are represented by two wooded savannas. From all these sites, 3 are on Mediterranean climate, and two more on semi-arid climates, whereas the rest of the sites are located in temperate climates.

Error metrics included mean bias error ( $\sum (Obs. - Pred.)/N$ ), root mean squared error ( $RMSE = \sqrt{\sum (Obs. - Pred.)^2/N}$ ), relative RMSE ( $RMSE/Obs$ ), and Pearson correlation coefficient between observed and predicted. Due to the lack of energy closure in the eddy covariance data, previous metrics were computed after adding the energy balance residual (residual =  $R_{n,EC} - G_{EC} - \lambda E_{EC} - H_{EC}$ ) to the latent heat flux. However, additional metrics were computed considering the lack of closure uncertainty in both measured latent and sensible heat fluxes. Observed latent heat flux  $\lambda E_{Obs}$  would range between  $\lambda E_{EC}$  and  $\lambda E_{EC} + \text{residual}$ . Likewise observed sensible heat flux would range between  $H_{EC}$  and  $H_{EC} + \text{residual}$ . Then we counted the number of predicted values in  $H$  and  $\lambda E$  that fall within those expected ranges. Furthermore, we computed then mean squared distance of the predicted turbulent fluxes to the boundaries of these

Table 1.1: Description of eddy covariance sites used for validation. Sites are listed in alphabetic order. Z shows the EC measurement height in meters, while the contact person for the EC tower is credited in PI

Site (abrevation)	Land cover	Climate	Z (m)	PI
Borden (BO)	Cropland, irrigated (vineyard)	Mediterranean	5	William P. Kustas (ARS-USDA)
Choptank (CH)	Cropland, irrigated (rotation of corn and soybean)	Temperate	5	William P. Kustas (ARS-USDA)
Dahra (DA)	Savanna	Semi-arid	9	Torbern Tagesson (Univ. Copenhagen)
Grillenburg (GR)	Grassland, meadow	Temperate	3	J. Thomas Gruenwald (T.U. Dresden)
Harvard Forest (HF)	Broadleaved forest	Temperate	30	J. William Munger (Harvard Univ.)
Hyltemossa (HY)	Conifer forest (spruce)	Temperate	27	Jutta Hols t (Lund Univ.)
Klingenberg (KL)	Cropland (spring barley and catch crops)	Temperate	3.5	J. Thomas Gruenwald (T.U. Dresden)
Majadas de Tieétar (MT)	Savanna	Mediterranean	15.5	Arnaud Carrara (CEAM)
Selhausen (SE)	Cropland (sugar beets and winter barley)	Temperate	2	Mathias Schmidt (Jülich)
Taous (TA)	Cropland, rainfed (olive)	Mediterranean	9.5	Gilles Boulet (CES-BIO)
Walnut Gulch (WG)	Grassland, steppe	Semi arid	4	Russell Scott (ARS-USDA)

intervals (i.e.  $d = 0$  if the predicted value is within the observed interval, otherwise  $d = \min \left[ (X_{Pred.} - X_{EC})^2, (X_{Pred.} - X_{EC} + \text{residual})^2 \right]$  for either  $X = H$  or  $X = \lambda E$ ).

### 1.2.3 Input data sources

The input data required to run the evapotranspiration models came from three main and two ancillary sources. The main sources were: Sentinel-2 optical observations, Sentinel-3 thermal observations and European Center for Medium-range Weather Forecasts (ECMWF) ERA-5 meteorological reanalysis data. The ancillary sources were: a digital elevation model (DEM) from the Shuttle Radar Topography Mission (SRTM) satellite, and land cover map created as part of the ESA Climate Change Initiative (CCI).

#### Satellite data

The main satellite data inputs come from the Sentinel-2 (both A and B) and Sentinel-3 (A only) satellites. In particular, the high-resolution shortwave observations needed to characterise the state of vegetation in the evapotranspiration model as well as to sharpen TIR data were obtained by the MultiSpectral Instrument (MSI) on board of the Sentinel-2A & 2B satellites. MSI acquires reflectance information in 13 spectral bands (with central wavelength ranging from 444 nm to 2202 nm) with a spatial resolution of 10 m, 20 m, or 60 m (depending the spectral band) and global geometric revisit of at least 5 days when both satellites are considered [63]. The MSI sensor has 3 spectral bands in the leaf-pigment sensitive red-edge part of the electromagnetic spectrum and two bands in water-content sensitive shortwave-infrared part of the spectrum, in addition to the more traditional visible and near-infrared bands, which makes it well suited for vegetation mapping and monitoring [64]. For each of the validation sites, all Sentinel-2 images for year 2017 were visually scanned and the ones which were cloud, fog and shadow free in the closes vicinity of the flux towers locations were selected for processing.

L1C top of the atmosphere images were converted to bottom-of-atmosphere (BOA) reflectances (L2A) using the Sen2Cor atmospheric correction processor [65] v2.5.5. BOA reflectance values were then used as input to the Biophysical Processor [66] available in the SNAP software v6.0.1 ([step.esa.int](http://step.esa.int) - last accessed 28.11.2018) in order to obtain effective values of green Leaf Area Index (LAI), Fractional Vegetation Cover (FVC), Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), Canopy Chlorophyll Content (CCC) and Canopy Water Content (CWC). The fraction of vegetation which is green and transpiring ( $f_g$ ) was estimated based on Fisher et al. [67] (Eq. 1.6).

$$f_g = FAPAR/FIPAR \quad (1.6)$$

where FIPAR is the fraction of photosynthetically active radiation intercepted by green and brown vegetation. FAPAR was obtained from the biophysical processor as described above, while FIPAR was derived iteratively from Eq. 1.7 Campbell and Norman [51].

$$FIPAR = 1 - \exp \frac{-0.5PAI}{\cos \theta} \quad (1.7)$$

where  $\theta$  is the solar zenith angle at the time of the S2 overpass, and PAI is the plant area index with initial PAI equal to LAI and in subsequent iterations

$$PAI = LAI/f_g \quad (1.8)$$

until  $f_g$  converges. Two assumptions made in Eq. 1.7 are that all intercepted PAR comes from the solar beams, and that both FAPAR and FIPAR are computed from a canopy with a spherical leaf inclination distribution. Indeed, from the the average leaf angle histogram, from which the training database was built in Weiss and Baret [66], most training cases in the Biophysical processor correspond to a spherical distribution (mode at  $60^\circ$ ). Equation 1.8 was subsequently used within the land surface models to convert LAI, which was assumed to represent green LAI [66], into PAI. Afterwards, PAI, leaf bi-hemispherical reflectance and transmittance, together with constant values for soil reflectance in the visible ( $VIS = [400 - 700]$  nm,  $\rho_{soil,VIS} = 0.15$ ) and near infrared ( $NIR = [700 - 2500]$  nm,  $\rho_{soil,NIR} = 0.25$ ) were used to quantify the shortwave net radiation of the soil and canopy (see section 3.1 for algorithm details). Leaf chlorophyll concentration (i.e.  $C_{a+b} = CCC/LAI$ ) was used to derive the leaf bihemispherical reflectance and transmittance in the visible after a curve fitting of 45000 ProspectD [68] simulations. Likewise, equivalent water thickness (i.e.  $C_w = CWC/LAI$ ) was used to retrieve leaf bihemispherical reflectance and transmittance in the NIR region.

The thermal data needed to drive the evapotranspiration model was obtained from the Sea and Land Surface Temperature Radiometer (SLSTR) on board of the Sentinel-3A satellite [69]. SLSTR contains 3 thermal infrared (TIR) channels (with two dynamic range settings - for fire monitoring and for sea/land surface temperature monitoring) with 1 km spatial resolution and less than two days temporal resolution with one satellite (less than one day with both A and B satellites) at the equator. For each selected S2 scene, all the S3 scenes falling on the day of S2 overpass or within four days after, were selected for processing. In the current study two SLSTR brightness temperature (BT) bands (S8 - centred on  $10.85 \mu\text{m}$  and S9 - centred on  $12 \mu\text{m}$ ) were used within a split-window algorithm described by [70] that, together with inputs of Total Column Water Vapour and surface emissivity, derived the 1 km *Trad* dataset.

Finally, the parameters in the ET models that could not be directly retrieved from optical observations (e.g. vegetation height or leaf inclination angle) were set based on a land cover map and a look-up table (see Table 1.2). The CCI landcover map from 2015 [71], which was produced with a global coverage and 300 m spatial resolution, was used as the initial input layer before being reclassified into the smaller number of classes as shown in Table 1.2. Out of the parameters set according to the look-up table, the vegetation height ( $h_C$ ) has the largest influence on the modelled fluxes as it effects aerodynamic roughness [72, 73]. Therefore in herbaceous classes where it can change throughout the growing season (grasslands and croplands) it was scaled with PAI using a power law, with maximum value  $h_{C,max}$  indicated in Table 1.2 reached at a prescribed maximum PAI  $PAI_{max}$  (5 in croplands and 4 in grasslands) and a minimum value set to 10 % of the maximum value.

## Meteorological data source

The meteorological data used in this study consists of air temperature at 2 m, dew point temperature at 2 m, wind speed at 10 m, surface pressure, TCWV and aerosol optical thickness (AOT) at 550 nm, surface roughness and surface geopotential. Those inputs are obtained from the ECMWF ERA5 reanalysis ensemble means dataset [74]. The only two exceptions are the AOT and surface roughness which come from the Copernicus Atmosphere Monitoring Service (CAMS) forecast dataset [75], since they are not included in

Table 1.2: Land cover based Look-Up-Table for ancillary parameters used in ET models. CCI-LC is the land cover code for the ESA’s CCI land cover legend; IGBP, code for the equivalent IGBP standard land cover legend;  $h_{C,max}$  is the maximum canopy height occurring when PAI reaches  $PAI_{max}$ ;  $f_C$  is fraction of the ground occupied by a clumped canopy ( $f_C = 1$  for a homogeneous canopy);  $w_C$  is canopy shape parameter, representing the canopy width to canopy height ratio;  $l_w$  is the average leaf size;  $\chi$  Campbell [52] leaf angle distribution parameter. See Chapter 3 for more details on the contribution of these parameters to the different models

CCI-LC (-)	IGBP (-)	$h_{C,max}$ (m)	$PAI_{max}$ (-)	$f_C$ (-)	$w_C/h_C$ (-)	$l_w$ (m)	$\chi$
0	255	0	0	0	0	0	0
10	12	1.2	5	1	1	0.02	0.5
11	12	1	5	1	1	0.02	0.5
12	12	2	5	0.5	2	0.1	1
20	12	1.2	5	1	1	0.02	0.5
30	14	1.2	5	0.5	1	0.05	0.5
40	14	1.2	5	0.5	1	0.1	0.5
50	2	10	5	1	1	0.15	1
60	4	10	5	1	1	0.15	1
61	4	10	5	1	1	0.15	1
62	4	10	5	0.4	1	0.15	1
70	1	20	5	1	2	0.05	1
71	1	20	5	1	2	0.05	1
72	1	20	5	0.4	2	0.05	1
80	3	20	5	1	2	0.05	1
81	3	20	5	1	2	0.05	1
82	3	20	5	0.4	2	0.05	1
90	5	15	5	1	1.5	0.1	1
100	8	8	5	0.75	1.5	0.15	0.8
110	9	8	5	0.25	1	0.02	0.5
120	6	1.5	4	1	1	0.05	1
121	6	1.5	4	1	1	0.05	1
122	6	1.5	4	1	1	0.05	1
130	10	0.5	4	1	1	0.02	0.5
140	16	0.05	1	1	1	0.001	1
150	9	2	2	0.15	1	0.05	1
151	9	10	5	0.15	1	0.1	1
152	7	1.5	4	0.15	1	0.05	1
153	10	0.5	4	0.15	1	0.02	0.5
160	11	10	5	1	1	0.1	1
170	11	10	5	1	1	0.1	1
180	11	1	5	1	1	0.02	0.5
190	13	20	0	0	0	0	0
200	16	0	0	0	0	0	0
201	16	0	0	0	0	0	0
202	16	0	0	0	0	0	0
210	0	0	0	0	0	0	0
220	15	0	0	0	0	0	0

ERA5. Inputs at the time of the satellite overpass are computed by linear interpolation between the previous and posterior reanalysis timestep. Due to the low spatial resolution of the air temperature and wind speed fields (tens of kilometers) they are assumed to represent the surface conditions derived from conditions above the blending height (100 m above

the surface) rather than the actual surface conditions. Therefore, air temperature at 100 m is calculated using the 2 m estimate, ECMWF surface geopotential, SRTM DEM and lapse rate for moist air. Similarly, wind speed at 100 m is calculated using the ECMWF 10 m estimate together with surface roughness forecast [32]. Those 100 meter estimates are then used as inputs into the land surface flux models. AOT together with TCWV, surface pressure, SRTM DEM elevation and solar zenith angle at the time of Sentinel-3 satellite overpass were used to estimate the instantaneous shortwave irradiance on a horizontal surface at the satellite overpass [76, 77].

### 1.2.4 Thermal data sharpening approaches

Two thermal data sharpening approaches were tested in this study: one based on ensemble of decision trees and one based on ensemble of neural networks. The two approaches share the same basic scheme (Fig. 1.1), data preparation and bias-reduction methods. Those methods are based on Gao et al. [6] and have been previously applied by Guzinski and Nieto [4] to sharpen thermal data to be used as input to evapotranspiration models. Each S3 scene is matched with an S2 scene acquired at most four days previously and the regression model used for sharpening is derived specifically for each scene pair.

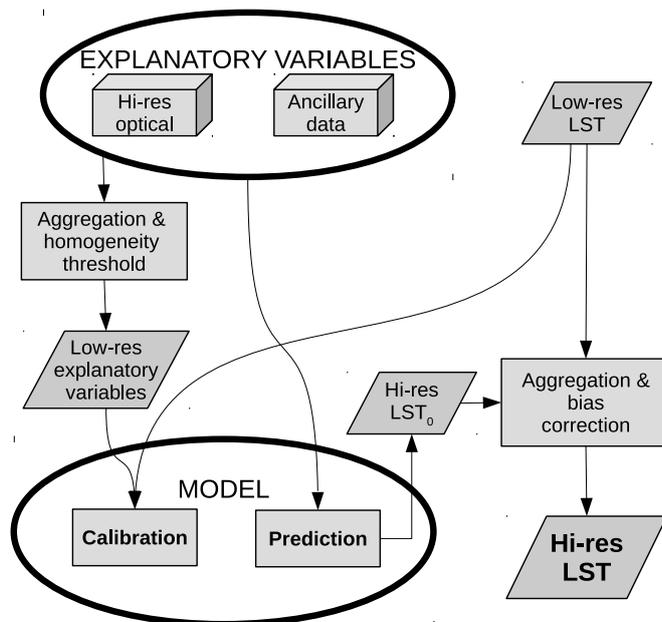


Figure 1.1: General thermal sharpening workflow. Explanatory variables include both multispectral band as well any other ancillary explanatory variable, such as elevation, land cover type or exposure. Model could be any regression model, such as multivariate linear regression or machine learning techniques

Briefly, the atmospherically corrected Sentinel-2 optical data with a spatial resolution of 20 m is resampled to match the pixel sampling of the SLSTR sensor (around 1 km spatial resolution). Concurrently, the SRTM DEM is used to derive slope and aspect maps which, together with S3 overpass time, are used to estimate the solar radiation incident of a flat tilted surface. The DEM and the radiation maps are also resampled

to the SLSTR resolution. A multivariate regression model is then trained with the three resampled datasets used as predictors and the  $T_{rad}$  used as the dependent variable. The selection of training samples is performed automatically by estimating the coefficient of variation (CV) of all the high-resolution pixels falling within one low-resolution pixel and selecting 80% of pixels with lowest CV. The regression model is based on bagging ensemble [78] of either decision trees or neural networks. The decision trees are additionally modified such that all samples within a regression tree leaf node are fitted with a multivariate linear model, as proposed by [6].

The regression models are trained on the whole S2 tile (100 km by 100 km) as well as on subsets of 30 by 30 S3 pixels in a moving window fashion. Once they are trained they are also applied on the whole scene and on each window. The bias between the predicted high-resolution  $T_{rad}$  pixels aggregated to the low-resolution and the original low-resolution  $T_{rad}$  is calculated and the outputs of the whole-scene and moving-window regressions are combined based on a weight inversely proportional to the bias [6]. Finally, the predicted  $T_{rad}$  by the regression model is corrected by comparing the emitted longwave radiance of the sharpened fine  $T_{rad}$  versus the original coarse  $T_{rad}$ . A bias-corrected  $T_{rad}$  is therefore re-calculated by adding an offset all fine scale pixels falling within coarse scale pixel in order to remove any residual bias. This is done to ensure the conservation of energy between the two thermal images with different spatial resolutions [6].

### 1.3 Results

The overall performance of the tested models using sharpened temperatures from Decision Trees regressor (hereinafter  $T_{rad,DT}$ ) is shown in Table 1.3. Likewise, overall errors from Artificial Neural Network sharpened temperatures ( $T_{rad,NN}$ ) are listed in Table 1.4. We have filtered all cases where estimated  $R_n \leq 50 \text{ W m}^{-2}$ , assuming that noisy outputs will be produced under low available energy, as well as those yielding unrealistic fluxes during daytime ( $\leq -500 \text{ W m}^{-2}$  and  $\geq 1000 \text{ W m}^{-2}$ ). After filtering the data, more than 200 cases were available overall for the following analyses. However, it is worth noting that METRIC yielded significantly fewer valid retrievals. This issue might be due to the fact that METRIC requires a contextual search of hot and cold pixels, and under cloudy conditions (i.e. when few pixels are available within a scene) as well as in sites where there is a lower range of soil moisture conditions and vegetation density, that search would not result in a reliable endmember selection from the automatic search algorithm. All models returned a similar performance regarding the estimation of  $R_n$ , with mean bias between -5 and  $20 \text{ W m}^{-2}$  and RMSE ranging between 46 and  $55 \text{ W m}^{-2}$ . This similar behaviour is explained by the fact that all models share the same approach and same inputs in modelling net shortwave radiation, which is the component with larger magnitude of  $R_n$ . Likewise, G showed similar behaviour as well, but in this case  $G_{METRIC}$  is computed differently as it is a function of surface  $R_n$  [32, 31] as opposed to TSEB and ESVEP that, as two-source models, G is computed from  $R_{n,S}$  [36, 44].

The main differences in model performance are therefore in the estimation of turbulent fluxes (i.e. sensible and latent heat fluxes), and TSEB usually produced more accurate estimates in terms of RMSE ( $\approx 75 \text{ W m}^{-2}$ , 43% relative error, in H, and  $\approx 90 \text{ W m}^{-2}$ , 49% relative error, in  $\lambda E$ ) and correlation between observed and predicted ( $> 0.7$ ), compared to METRIC and ESVEP, where RMSE are in all cases higher than  $100 \text{ W m}^{-2}$  and with

Table 1.3: Error metrics for METRIC, TSEB-PT, disTSEB (TSEB-PT with flux disaggregation) and ESVEP modelled fluxes using Decision Tree sharpened temperatures. N, number of valid cases;  $\overline{\text{Obs.}}$ ; mean of observed values ( $\text{W m}^{-2}$ ); bias, mean difference between predicted and observed ( $\text{W m}^{-2}$ ); MAE, Mean Absolute Error ( $\text{W m}^{-2}$ ), RMSE, Root Mean Square Error ( $\text{W m}^{-2}$ ); rRMSE, Relative RMSE (-); r, Pearson correlation coefficient (-); % points in plausible range from closure error; RDE, Range Difference Error ( $\text{W m}^{-2}$ )

variable	model	N	$\overline{\text{Obs.}}$	bias	MAE	RMSE	rRMSE	r	% points	RDE
H	METRIC	242	171	15	95	146	0.854	0.286	39.7	119
	TSEB-PT	309	176	-44	60	75	0.426	0.734	20.1	73
	disTSEB	300	177	-36	58	73	0.413	0.728	21.8	69
	ESVEP	259	166	67	137	207	1.246	0.413	21.5	171
$\lambda E$	METRIC	229	199	9	103	130	0.654	0.334	37.6	106
	TSEB-PT	304	180	36	70	88	0.491	0.770	39.1	75
	disTSEB	296	180	42	72	92	0.510	0.766	35.1	80
	ESVEP	219	193	-13	103	127	0.656	0.431	36.4	89
$R_n$	METRIC	250	456	-13	35	46	0.101	0.943	-	-
	TSEB-PT	331	441	12	38	51	0.116	0.919	-	-
	disTSEB	321	439	21	42	55	0.126	0.916	-	-
	ESVEP	322	440	5	38	50	0.114	0.918	-	-
G	METRIC	244	86	-16	47	56	0.655	0.597	-	-
	TSEB-PT	325	83	20	43	55	0.661	0.571	-	-
	disTSEB	315	80	11	44	54	0.669	0.529	-	-
	ESVEP	317	84	26	48	62	0.734	0.485	-	-

lower correlation ( $< 0.5$ ). Nevertheless, METRIC showed lower mean bias than the other models (9 and 15  $\text{W m}^{-2}$  respectively for  $\lambda\text{E}$  and H).

Regardless on the model use, both sharpening methods yield similar results, with  $T_{rad,DT}$  showing slightly better metrics than  $T_{rad,NN}$ . Therefore, in following analysis only results for  $T_{rad,DT}$  will be listed, while results regarding  $T_{rad,NN}$  are included as Supplementary Material.

Table 1.4: Error metrics for METRIC, TSEB-PT, disTSEB (TSEB-PT with flux disaggregation), and ESVEP modelled fluxes using Artificial Neural Network sharpened temperatures. N, number of valid cases;  $\overline{\text{Obs.}}$ ; mean of observed values ( $\text{W m}^{-2}$ ); bias, mean difference between predicted and observed ( $\text{W m}^{-2}$ ); MAE, Mean Absolute Error ( $\text{W m}^{-2}$ ), RMSE, Root Mean Square Error ( $\text{W m}^{-2}$ ); rRMSE, Relative RMSE (-); r, Pearson correlation coefficient (-); % points in plausible range from closure error; RDE, Range Difference Error ( $\text{W m}^{-2}$ )

variable	model	N	$\overline{\text{Obs.}}$	bias	MAE	RMSE	rRMSE	r	% points	RDE
H	METRIC	240	170	25	95	138	0.809	0.285	35.3	105
	TSEB-PT	309	176	-45	61	76	0.430	0.728	19.7	73
	disTSEB	300	177	-37	59	74	0.417	0.724	22.5	70
	ESVEP	262	167	69	141	212	1.273	0.426	21.3	175
$\lambda\text{E}$	METRIC	224	200	-1	101	131	0.656	0.348	33.3	101
	TSEB-PT	304	180	37	69	88	0.491	0.769	37.7	75
	disTSEB	296	180	42	72	92	0.510	0.765	33.7	81
	ESVEP	222	194	-13	104	129	0.667	0.419	34.4	91
$R_n$	METRIC	250	456	-13	35	46	0.101	0.943	-	-
	TSEB-PT	331	441	12	39	51	0.116	0.919	-	-
	disTSEB	321	439	21	42	55	0.126	0.915	-	-
	ESVEP	324	441	6	39	50	0.114	0.919	-	-
G	METRIC	244	86	-16	47	56	0.654	0.598	-	-
	TSEB-PT	325	83	20	44	55	0.664	0.566	-	-
	disTSEB	315	80	11	44	54	0.675	0.523	-	-
	ESVEP	319	84	26	48	61	0.729	0.495	-	-

In order to evaluate the model sensitivity and uncertainty to different vegetation types, we have split the results of Table 1.3 into four main vegetation types, depending on differences in aerodynamic roughness, horizontal homogeneity and/or seasonal dynamics/senescence (i.e. croplands, grasslands, savannas and forests, Table 1.5). At croplands and savannas, small differences are found between TSEB and METRIC, with moderate errors between 80 and 100  $\text{W m}^{-2}$  RMSE in  $\lambda\text{E}$  and between 65 and 90  $\text{W m}^{-2}$  in H, but ESVEP returned larger errors ( $> 120 \text{ W m}^{-2}$  RMSE in  $\lambda\text{E}$  and errors up to 276  $\text{W m}^{-2}$  in the case of RMSE in H in savannas). Despite of the relative similar performance between METRIC and TSEB, METRIC yielded fewer valid retrievals and still showed a lower agreement between the observed and the predicted when estimating H ( $r=0.27$  in METRIC vs. 0.62 in TSEB). As for the grasslands, TSEB outperformed the other two models, both in terms of RMSE (64  $\text{W m}^{-2}$  in H and 86  $\text{W m}^{-2}$  in  $\lambda\text{E}$ ) and correlation (0.81 in H and 0.76 in  $\lambda\text{E}$ ). Finally very few valid cases are available to evaluate the forest sites and hence the results are not very conclusive, with METRIC outperforming the other models but producing fewer valid retrievals.

Table 1.5: Error dependence on land cover for METRIC, TSEB-PT, disTSEB (TSEB-PT with flux disaggregation) and ESVEP modelled fluxes using Decision Trees sharpened temperatures. N, number of valid cases;  $\overline{\text{Obs.}}$ ; mean of observed values ( $\text{W m}^{-2}$ ); bias, mean difference between predicted and observed ( $\text{W m}^{-2}$ ); MAE, Mean Absolute Error ( $\text{W m}^{-2}$ ), RMSE, Root Mean Square Error ( $\text{W m}^{-2}$ ); rRMSE, Relative RMSE (-); r, Pearson correlation coefficient (-); % points in plausible range from closure error; RDE, Range Difference Error ( $\text{W m}^{-2}$ )

variable	land cover	model	N	$\overline{\text{Obs.}}$	bias	MAE	RMSE	rRMSE	r	% points	RDE
H	cropland	METRIC	79	135	48	77	91	0.672	0.271	57.7	39
		TSEB-PT	105	154	-55	66	83	0.542	0.617	20.2	81
		disTSEB	104	154	-48	64	81	0.524	0.593	18.4	77
		ESVEP	86	141	81	123	191	1.348	0.418	28.2	143
	grassland	METRIC	100	195	-16	117	170	0.870	0.221	24.5	160
		TSEB-PT	101	194	-26	51	64	0.328	0.814	20.0	59
		disTSEB	93	198	-25	51	65	0.328	0.809	20.7	60
		ESVEP	91	185	8	113	161	0.869	0.428	16.3	134
	savanna	METRIC	52	185	-18	52	67	0.362	0.777	39.2	56
		TSEB-PT	90	186	-47	57	69	0.371	0.779	18.2	67
		disTSEB	90	186	-30	54	65	0.352	0.762	26.0	59
		ESVEP	71	175	139	187	276	1.579	0.476	18.8	243
	forest	METRIC	11	140	221	222	354	2.535	0.528	45.5	237
		TSEB-PT	13	151	-72	94	115	0.763	0.502	30.8	113
		disTSEB	13	151	-62	88	107	0.706	0.566	30.8	104
		ESVEP	11	147	-12	116	138	0.941	0.480	27.3	106
$\lambda E$	cropland	METRIC	75	262	-46	85	104	0.397	0.614	41.9	45
		TSEB-PT	101	239	31	63	80	0.333	0.795	36.0	71
		disTSEB	100	239	39	66	85	0.355	0.774	35.4	78
		ESVEP	75	256	-56	96	122	0.477	0.462	37.8	72
	grassland	METRIC	94	130	99	132	163	1.254	0.306	22.7	161
		TSEB-PT	100	129	47	71	86	0.671	0.765	28.7	81
		disTSEB	93	126	57	77	93	0.738	0.756	19.5	89
		ESVEP	84	133	41	102	120	0.902	0.417	23.8	103
	savanna	METRIC	52	221	-65	82	98	0.442	0.352	54.9	35
		TSEB-PT	90	158	25	73	94	0.594	0.493	57.1	59
		disTSEB	90	158	26	70	92	0.583	0.533	53.2	61
		ESVEP	49	189	-45	107	136	0.719	0.144	56.8	72
	forest	METRIC	8	269	-50	68	83	0.310	0.934	50.0	29
		TSEB-PT	13	277	71	95	125	0.450	0.969	30.8	123
		disTSEB	13	277	76	100	130	0.467	0.972	30.8	128
		ESVEP	11	254	19	133	166	0.654	0.853	36.4	131

*Continued on next page*

Table 1.5 – Continued from previous page

variable	land cover	model	N	$\overline{\text{Obs.}}$	bias	MAE	RMSE	rRMSE	r	% points	RDE
$R_n$	cropland	METRIC	81	434	6	26	36	0.082	0.966	–	–
		TSEB-PT	124	429	23	32	41	0.096	0.961	–	–
		disTSEB	122	428	33	42	51	0.119	0.952	–	–
		ESVEP	120	430	9	33	43	0.099	0.941	–	–
	grassland	METRIC	104	470	-9	34	49	0.105	0.897	–	–
		TSEB-PT	104	470	0	35	49	0.104	0.897	–	–
		disTSEB	96	467	5	37	52	0.110	0.880	–	–
		ESVEP	102	469	-2	33	47	0.101	0.903	–	–
	savanna	METRIC	52	462	-53	53	57	0.123	0.990	–	–
		TSEB-PT	90	421	12	54	66	0.158	0.822	–	–
		disTSEB	90	421	22	51	67	0.159	0.849	–	–
		ESVEP	88	421	7	54	64	0.152	0.834	–	–
	forest	METRIC	13	459	4	20	27	0.060	0.994	–	–
		TSEB-PT	13	459	7	21	27	0.058	0.995	–	–
		disTSEB	13	459	10	22	28	0.061	0.995	–	–
		ESVEP	12	448	13	20	26	0.058	0.996	–	–
G	cropland	METRIC	81	40	22	41	50	1.264	0.271	–	–
		TSEB-PT	124	41	60	67	75	1.843	0.281	–	–
		disTSEB	122	39	51	59	66	1.690	0.295	–	–
		ESVEP	120	41	69	74	87	2.124	0.265	–	–
	grassland	METRIC	104	143	-64	66	72	0.503	0.811	–	–
		TSEB-PT	104	143	-25	34	43	0.300	0.781	–	–
		disTSEB	96	141	-31	38	51	0.361	0.715	–	–
		ESVEP	102	145	-25	36	43	0.298	0.763	–	–
	savanna	METRIC	52	57	16	20	23	0.415	0.971	–	–
		TSEB-PT	90	77	18	24	31	0.395	0.708	–	–
		disTSEB	90	77	-1	30	37	0.475	0.433	–	–
		ESVEP	88	78	26	29	34	0.435	0.778	–	–
	forest	METRIC	7	-2	18	18	27	-11.544	0.866	–	–
		TSEB-PT	7	-2	19	19	23	-10.063	0.789	–	–
		disTSEB	7	-2	13	13	16	-6.932	0.794	–	–
		ESVEP	7	-2	15	15	18	-7.762	0.784	–	–

The agriculture class was further split into herbaceous and woody types, with results shown in Table 1.6. The former sub-class represents crops such as corn, soybean or wheat while the latter represents olive groves and vineyards. TSEB model produces the most consistent results for both types of crops, although the RMSE of  $\lambda E$  in herbaceous crops ( $70 \text{ W m}^{-2}$ ) is significantly lower than in woody crops ( $84 \text{ W m}^{-2}$ ). ESVEP is very clearly performing better in herbaceous crops, while METRIC obtains better results in woody crops.

Similarly, Table 1.7 lists the model performance depending on whether sites are under Mediterranean or semi-arid climate (i.e. water limited sites), and sites under temperate climate (i.e. energy limited sites). First of all it is worth noting that due to cloud coverage conditions, more valid cases are obtained over semi-arid conditions than in temperate areas. TSEB model showed similar range of errors in both climatic conditions, with RMSE in  $\lambda E$  at  $90 \text{ W m}^{-2}$  and  $84 \text{ W m}^{-2}$  for semi-arid and temperate conditions, and correspondingly  $73$  and  $81 \text{ W m}^{-2}$  for H. ESVEP and METRIC yielded more varying results between

climates, with METRIC producing more accurate estimates of  $\lambda E$  but less accurate in H, over temperate climates than in semi-arid conditions. On the other hand, ESVEP showed overall better performance over temperate than in semi-arid climates.

Table 1.6: Crop type dependent errors for METRIC, TSEB and ESVEP modelled fluxes using Decision Tree sharpened temperatures. N, number of valid cases;  $\overline{\text{Obs.}}$ ; mean of observed values ( $\text{W m}^{-2}$ ); bias, mean difference between predicted and observed ( $\text{W m}^{-2}$ ); MAE, Mean Absolute Error ( $\text{W m}^{-2}$ ), RMSE, Root Mean Square Error ( $\text{W m}^{-2}$ ); rRMSE, Relative RMSE (-); r, Pearson correlation coefficient (-); % points in plausible range from closure error; RDE, Range Difference Error ( $\text{W m}^{-2}$ )

variable	crop type	model	N	$\overline{\text{Obs.}}$	bias	MAE	RMSE	rRMSE	r	% points	RDE
H	herbaceous	METRIC	39	123	71	80	93	0.759	0.571	31	20
		TSEB-PT	39	123	-54	63	82	0.667	0.479	8	81
		disTSEB	38	123	-52	60	79	0.646	0.550	8	78
		ESVEP	37	122	19	57	76	0.625	0.497	19	45
	woody	METRIC	40	146	25	73	88	0.601	0.014	14	51
		TSEB-PT	66	172	-56	68	84	0.490	0.620	13	81
		disTSEB	66	172	-46	66	81	0.473	0.569	11	76
		ESVEP	49	156	127	173	244	1.563	0.384	5	186
$\lambda E$	herbaceous	METRIC	35	246	-83	98	119	0.482	0.520	22	27
		TSEB-PT	35	246	-2	56	70	0.283	0.711	21	51
		disTSEB	34	246	1	55	70	0.284	0.729	20	55
		ESVEP	34	245	-54	69	92	0.375	0.619	22	26
	woody	METRIC	40	275	-14	75	89	0.322	0.772	9	57
		TSEB-PT	66	235	49	66	84	0.359	0.845	15	80
		disTSEB	66	235	58	72	92	0.390	0.823	15	88
		ESVEP	41	264	-58	119	142	0.538	0.404	6	95
$R_n$	herbaceous	METRIC	41	413	-7	26	36	0.087	0.967	-	-
		TSEB-PT	41	413	0	25	34	0.083	0.969	-	-
		disTSEB	39	408	4	26	36	0.088	0.968	-	-
		ESVEP	40	411	6	25	33	0.081	0.972	-	-
	woody	METRIC	40	455	20	25	35	0.077	0.974	-	-
		TSEB-PT	83	437	34	36	44	0.101	0.968	-	-
		disTSEB	83	437	46	49	57	0.129	0.959	-	-
		ESVEP	80	439	10	37	47	0.106	0.915	-	-
G	herbaceous	METRIC	41	45	2	39	47	1.026	0.487	-	-
		TSEB-PT	41	45	50	62	73	1.608	0.316	-	-
		disTSEB	39	42	45	58	68	1.629	0.311	-	-
		ESVEP	40	46	35	49	57	1.236	0.548	-	-
	woody	METRIC	40	34	42	44	53	1.586	0.335	-	-
		TSEB-PT	83	38	66	69	75	1.980	0.447	-	-
		disTSEB	83	38	54	60	66	1.720	0.447	-	-
		ESVEP	80	38	86	87	99	2.569	0.451	-	-

Table 1.7: Climate dependence of errors for METRIC, TSEB and ESVEP modelled fluxes using Decision Trees sharpened temperatures. N, number of valid cases;  $\overline{\text{Obs.}}$ ; mean of observed values ( $\text{W m}^{-2}$ ); bias, mean difference between predicted and observed ( $\text{W m}^{-2}$ ); MAE, Mean Absolute Error ( $\text{W m}^{-2}$ ), RMSE, Root Mean Square Error ( $\text{W m}^{-2}$ ); rRMSE, Relative RMSE (-); r, Pearson correlation coefficient (-); % points in plausible range from closure error; RDE, Range Difference Error ( $\text{W m}^{-2}$ )

variable	climate	model	N	$\overline{\text{Obs.}}$	bias	MAE	RMSE	rRMSE	r	% points	RDE
H	semi-arid	METRIC	177	192	-16	90	136	0.710	0.326	27.2	126
		TSEB-PT	242	193	-44	60	73	0.380	0.744	18.0	70
		disTSEB	234	194	-35	59	72	0.370	0.728	20.1	67
		ESVEP	198	182	80	157	231	1.268	0.369	11.9	194
	temperate	METRIC	65	113	101	107	169	1.496	0.486	72.3	100
		TSEB-PT	67	116	-45	59	81	0.700	0.472	26.9	79
		disTSEB	66	116	-42	56	77	0.668	0.539	27.3	76
		ESVEP	61	114	25	71	93	0.817	0.458	50.8	58
$\lambda E$	semi-arid	METRIC	171	180	37	108	137	0.760	0.298	28.8	122
		TSEB-PT	241	160	41	72	90	0.560	0.745	34.8	75
		disTSEB	234	160	47	75	93	0.583	0.736	32.2	81
		ESVEP	161	174	0	109	132	0.757	0.352	27.2	97
	temperate	METRIC	58	254	-72	88	107	0.422	0.747	62.1	29
		TSEB-PT	63	257	16	61	84	0.325	0.866	54.0	74
		disTSEB	62	257	24	63	87	0.338	0.875	45.2	80
		ESVEP	58	246	-47	85	111	0.452	0.712	60.3	62
$R_n$	semi-arid	METRIC	180	473	-19	38	49	0.103	0.921	-	-
		TSEB-PT	261	449	13	42	54	0.121	0.881	-	-
		disTSEB	253	447	22	46	59	0.131	0.876	-	-
		ESVEP	256	450	4	42	54	0.119	0.877	-	-
	temperate	METRIC	70	411	3	27	38	0.093	0.973	-	-
		TSEB-PT	70	411	9	26	38	0.092	0.975	-	-
		disTSEB	68	409	14	29	40	0.098	0.975	-	-
		ESVEP	66	405	11	24	34	0.084	0.981	-	-
G	semi-arid	METRIC	180	104	-22	52	61	0.591	0.485	-	-
		TSEB-PT	261	94	16	42	53	0.565	0.531	-	-
		disTSEB	253	92	5	44	53	0.581	0.493	-	-
		ESVEP	256	95	25	50	64	0.676	0.343	-	-
	temperate	METRIC	64	37	3	31	39	1.073	0.533	-	-
		TSEB-PT	64	37	38	48	60	1.645	0.426	-	-
		disTSEB	62	34	32	43	56	1.633	0.406	-	-
		ESVEP	61	37	30	40	48	1.318	0.600	-	-

## 1.4 Discussions

### 1.4.1 ET model intercomparison

Overall results listed in Table 1.3 or Table 1.4 show that TSEB produced more robust estimates of both sensible and latent heat fluxes, with lower error around  $90 \text{ W m}^{-2}$  and larger correlation coefficient, while at the same time it returns more valid cases than the other two models, METRIC and ESVEP. Those errors are within the expected and reported errors in literature, e.g. Kalma et al. [3] showed errors in  $\lambda E$  ranging between 24 and  $105 \text{ W m}^{-2}$  for a wide range of models, Chirouze et al. [58] reported errors for TSEB  $> 100 \text{ W m}^{-2}$  in a semi-arid area of Mexico,  $50 \text{ W m}^{-2}$  in Tang et al. [35], or Choi et al. [79] found TSEB and METRIC produced similar errors of  $54 \text{ W m}^{-2}$  in a watershed in Iowa, US. However it is worth noting that most of the reported errors in these studies [79, 58, 80, 35] used actual surface temperature at high spatial resolution (e.g. Landsat or ASTER), whereas in this study we used low resolution temperature sharpened at high spatial resolution, which provides an additional input uncertainty to the models. For that reason, section 1.4.2 is dedicated to this issue in depth.

TSEB was developed trying to solve some of the issues in sparse vegetation and semi-arid conditions previously raised by less complex models [36], and therefore it adapts better to a wider range of climatic and vegetation conditions [2] as it was shown in Tables 1.5 and 1.7. METRIC, on the other hand, was primarily designed for standard crops and requires concomitant presence of stressed and well watered-full vegetation conditions within the scene itself. Those cases in which, either due to the increased presence of clouds (i.e. fewer available pixels in the scene) or in regions where these hot and cold pixels cannot be simultaneously found, METRIC would produce more uncertain retrieval, as already pointed by Choi et al. [79] and Tang et al. [35] in humid or sub-humid areas, or even could not produce any valid data. That is the case for instance in the semi-arid sites, in particular Dahra and Taus (see Figs 1.8, 1.9, 1.22, or 1.23 in section 1.5), where no valid retrieval was obtained whatsoever.

Finally, it is worth pointing out that even *in situ* EC measurements are prone to uncertainty as it is confirmed for instance by the usual energy imbalance in those systems. Particularly is found a larger disagreement between observed and predicted net radiation in Dahra (Figs. 1.8 and 1.9), which affects specially the error metrics in  $R_n$  for TSEB and ESVEP compared to METRIC (as the latter model as it was commented before did not produce any valid case in this site to validate with). We hypothesise in two possible reasons, one is that our modelled irradiance, with depends on TCWV and aerosol optical thickness, are more noisy than the other sites, due to unaccounted dust aerosols in that site place over the Sahel. The second issue might be the actual  $R_n$  measurements as in this site only a NR-lite (Kipp & Zonen, Netherlands) is available to measure global  $R_n$  that might be less accurate than the radiometers in the other sites, which are measuring the four components of radiation. On the other hand, Harvard Forest site lacks *in situ* G measurements, which effects on the energy balance closure correction. This issue together with the fact that very few cases are available in forests (Table 1.5), leads to avoiding strong conclusions regarding the performance of the models in forests areas.

### 1.4.2 Sharpening and disaggregation

Comparison of the two sharpening methods evaluated (Decision Trees and Artificial Neural Network) leads to very similar results found between methods for all models tested, with very slightly better performance in  $T_{rad,DT}$  (Table 1.3) compared to  $T_{rad,NN}$  (Table 1.4). Furthermore, spatial patterns of sharpened temperature are very similar (see example comparison image in Fig 1.2). Therefore in the future, an assessment between sharpening methods needs to be carried out comparing the agreement in spatial patterns with actual surface temperature, for instance using concomitant Landsat  $T_{rad}$  imagery.

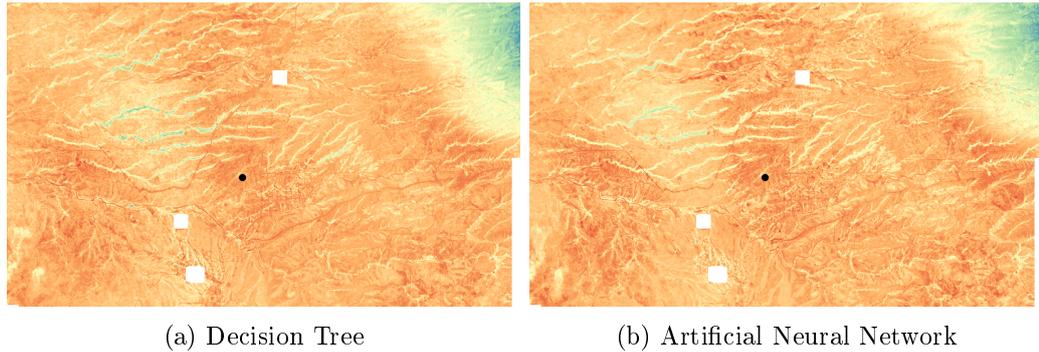


Figure 1.2: Spatial patterns comparison between both sharpened temperatures. Images correspond to the Walnut Gulch site on September 27th, 2017. The black dot shows the location of the flux tower

As it was previously mentioned thermal sharpening relates empirically or semi-empirically coarse resolution surface temperature with fine resolution multispectral and other ancillary data. This technique could be a sound alternative to the lack high resolution thermal imagery for operational activities. However, previous studies in thermal sharpening have already reported some uncertainties when compared to actual  $T_{rad}$  temperatures, with errors ranging up to 3.5 K [12, 5, 6, 7, 15, 13]. Therefore, for some applications requiring ET estimates at higher accuracy (i.e. precision agriculture), sharpening would not be considered as a substitute of  $T_{rad}$  but complementary to it, such as in the fusion approach by Knipper et al. [81].

In order to reduce flux retrieval errors with sharpened  $T_{rad}$  retrievals, we also tested a flux disaggregation method [8, 23]. Our results listed in Tables 1.3-1.7 show that disTSEB model, i.e. coarse S3 TSEB fluxes disaggregated with fine resolution sharpened  $T_{rad}$ , did not yield any improvement to the TSEB-PT model, i.e. running TSEB directly on the sharpened  $T_{rad}$  imagery. However, previous studies have shown the robustness of this approach to overcome limitations on the likely less reliable fine resolution  $T_{rad}$  images [82, 81, 4]. Furthermore, coarse input data must be produced beforehand for thermal sharpening and hence it is readily available for running the models at coarse resolution, which indeed is computationally inexpensive given the much lower number of pixels within a scene. Therefore, flux disaggregation would still be recommended when running TSEB with sharpened temperatures.

On the hand, the fact that coarse resolution  $T_{rad}$  image could be combined with fine resolution images on different days, with 4 days offset set to maximum corresponding to the nominal Sentinel 2 A & B combined overpass, might lead to additional uncertainties

when sharpening the temperature. This is caused by the fact that some changes in either in land cover properties, (e.g. vegetation growth, harvests, fires) or moisture conditions (e.g. rainfall or irrigation) might happen between the Sentinel 2 and 3 acquisitions. Fig. 1.3 shows how RMSE tends to increase for sensible heat flux with increasing days offset between S2 and S3 overpass. In this analysis H is plotted only since it is the energy component that is directly related to  $T_{rad}$ , and hence more prone to errors in sharpening.

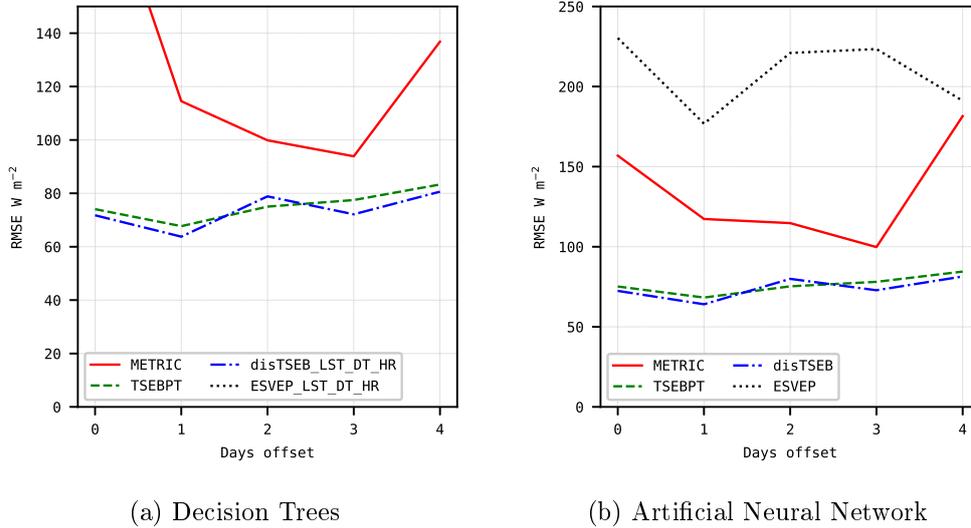


Figure 1.3: Root Mean Squared Error for sensible heat flux depending on offset days between a Sentinel-3  $T_{rad}$  image and the fine-scale Sentinel-2 multispectral image. RMSE computed for all sites together

Finally, some studies have reported larger errors than in this study, but they were using coarser resolution imagery [43]. This is probably due to the scale mismatch between the coarse pixel estimate and the measured by EC towers. However those differences are reduced when comparing flux estimates with measurements made at larger scale, for instance with Large Aperture Scintillometry Tang et al. [35] found errors using MODIS of  $50 \text{ W m}^{-2}$ . Therefore, even sharpened  $T_{rad}$  being more prone to errors than actual high-resolution  $T_{rad}$ , it has proven to be a good alternative to downscaling fluxes for model validation [4].

### 1.4.3 Effects of ancillary inputs

Ancillary data is required to characterise the canopy structure, since it effects both the radiation transmission through the canopy [53], and hence albedo and radiation partitioning, as well as the surface aerodynamic properties [72]. In this study we have used a static land cover map at global scale to assign some standard values to each land cover type (Table 1.2). However, the large difference in spatial resolution between the S2 data and CCI map can lead to visible artefacts in the output fluxes when modelled at 20 m resolution, especially on the edges of two classes with different vegetation properties (e.g. croplands and forests). However, those spatial artefacts seems not to have any influence in the validation results. Nevertheless, some discrepancies were found between the land cover type flagged

by the map and the actual coverage at the study site and around. In Majadas de Tiétar, CCI-LC flagged the site as cropland (CCI-LC=11), thus  $h_{c,MAX} = 0.5$  m,  $f_c = 1$  and  $l_w = 0.02$  m, but actually this site is a savanna with 8 m tress at 20 % coverage (CCI-LC=30). On the other hand the prescribed values that were assigned in Table 1.2 are very general, as they- were trying to fit a global-based land cover legend. Therefore they can significantly deviate from the site's actual values. Indeed, all croplands were assumed that they are not clumped ( $f_c = 1$ ) although row crops, like the vineyard in Borden, or orchards like the olive grove in Taous have very different canopy structure compared to a standard crop. Therefore, it could be expected a significant improvement if a more area-specific surface characteristics parametrizing was used, either using some ancillary remote sensing like SAR imagery or LiDAR or a regional/local oriented land cover classification.

To conclude, atmospheric forcing from numerical weather prediction models might add some uncertainty to the ET model compared to using local meteorological data, specially for precision agriculture where access to local agrometeorological stations is possible. This issue could be more evident in near-real-time applications as in that case the requirement for using forecast or analysis data, instead of the ensemble mean reanalysis data that was used in this study, could lead to increased uncertainties.

## 1.5 Conclusions

This study evaluated three thermal-based remote sensing methods (METRIC, ESVEP and TSEB) and two thermal sharpening methods (Artificial Neural Networks and Decision Trees) in order to produce operational estimates of water and energy fluxes using Sentinel data for the whole Globe. Regardless of the sharpening method, TSEB produced overall the most accurate estimates in terms of sensible heat and latent heat (i.e. evapotranspiration) fluxes, being robust in different land covers and climates. Despite of TSEB being the model with the largest demand amount of input data, this study proposed several approaches to retrieve some on the inputs, with special focus on exploiting the spectral capabilities of Sentinel-2, in particular with bands in the red edge region that is sensitive to leaf pigments. A simple empirical approach relating leaf bihemispherical reflectance and transmittance with the leaf biochemical properties, which can be derived with the ESA's off-the-shelf open-source SNAP software, resulted in accurate estimates of net radiation. More importantly, due to the larger uncertainty of TSEB models over senescent vegetation, we derived a method to obtain both total LAI and its green fraction based on [67] FA-PAR/FIPAR relationship. Nevertheless, more research is needed to systematically derive other vegetation properties such as canopy height/aerodynamic roughness or vegetation clumping.

## Supplementary material

Table 1.8: Error dependence on land cover for METRIC, TSEB and ESVEP modelled fluxes using Artificial Neural Network sharpened temperatures. N, number of valid cases;  $\overline{\text{Obs.}}$ ; mean of observed values ( $\text{W m}^{-2}$ ); bias, mean difference between predicted and observed ( $\text{W m}^{-2}$ ); MAE, Mean Absolute Error ( $\text{W m}^{-2}$ ), RMSE, Root Mean Square Error ( $\text{W m}^{-2}$ ); rRMSE, Relative RMSE (-); r, Pearson correlation coefficient (-); % points in plausible range from closure error; RDE, Range Difference Error ( $\text{W m}^{-2}$ )

variable	land cover	model	N	$\overline{\text{Obs.}}$	bias	MAE	RMSE	rRMSE	r	% points	RDE
H	cropland	METRIC	79	135	47	81	101	0.752	0.133	51.3	52
		TSEB-PT	105	154	-56	68	85	0.552	0.597	20.2	82
		disTSEB	104	154	-50	65	82	0.532	0.581	19.4	78
		ESVEP	86	143	86	130	199	1.390	0.453	28.2	151
	grassland	METRIC	100	194	20	111	161	0.828	0.240	25.5	143
		TSEB-PT	101	194	-27	51	64	0.330	0.813	20.0	59
		disTSEB	93	198	-25	52	66	0.332	0.807	21.8	61
		ESVEP	92	184	4	111	158	0.860	0.426	17.2	133
	savanna	METRIC	52	185	-30	59	77	0.415	0.737	31.4	68
		TSEB-PT	90	186	-48	57	69	0.374	0.775	16.9	67
		disTSEB	90	186	-30	54	65	0.352	0.761	26.0	59
		ESVEP	73	175	144	194	283	1.617	0.497	18.2	248
	forest	METRIC	9	121	193	248	304	2.512	0.272	22.2	142
		TSEB-PT	13	151	-71	95	116	0.765	0.488	30.8	113
		disTSEB	13	151	-62	88	107	0.706	0.555	30.8	103
		ESVEP	11	147	-10	119	143	0.975	0.478	18.2	108
$\lambda E$	cropland	METRIC	75	262	-44	88	112	0.429	0.533	37.8	57
		TSEB-PT	101	239	32	62	79	0.332	0.790	37.0	71
		disTSEB	100	239	39	65	85	0.354	0.770	31.3	78
		ESVEP	76	253	-62	104	132	0.520	0.419	34.7	80
	grassland	METRIC	91	133	71	118	152	1.145	0.326	17.6	144
		TSEB-PT	100	129	48	71	87	0.674	0.765	25.5	82
		disTSEB	93	126	57	77	94	0.744	0.754	20.7	90
		ESVEP	86	135	43	101	118	0.870	0.479	22.0	103
	savanna	METRIC	52	221	-53	77	94	0.428	0.335	54.9	38
		TSEB-PT	90	158	25	72	93	0.593	0.497	54.5	59
		disTSEB	90	158	26	69	92	0.581	0.536	51.9	60
		ESVEP	49	190	-40	103	133	0.700	0.119	56.8	71
	forest	METRIC	6	274	-98	205	243	0.884	0.577	16.7	133
		TSEB-PT	13	277	70	96	125	0.451	0.969	30.8	123
		disTSEB	13	277	75	100	130	0.467	0.970	30.8	127
		ESVEP	11	254	17	137	171	0.675	0.841	36.4	134

*Continued on next page*

Table 1.8 – *Continued from previous page*

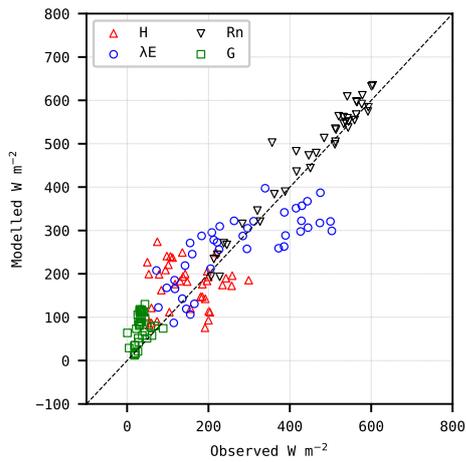
variable	land cover	model	N	$\overline{\text{Obs.}}$	bias	MAE	RMSE	rRMSE	r	% points	RDE
$R_n$	cropland	METRIC	81	434	7	26	35	0.082	0.967	–	–
		TSEB-PT	124	429	24	33	42	0.097	0.961	–	–
		disTSEB	122	428	34	42	51	0.120	0.953	–	–
		ESVEP	121	431	12	33	42	0.097	0.947	–	–
	grassland	METRIC	104	470	-9	34	49	0.104	0.898	–	–
		TSEB-PT	104	470	0	34	49	0.104	0.898	–	–
		disTSEB	96	467	5	37	51	0.110	0.881	–	–
		ESVEP	102	469	-1	33	47	0.100	0.904	–	–
	savanna	METRIC	52	462	-54	54	58	0.125	0.990	–	–
		TSEB-PT	90	421	11	54	66	0.157	0.823	–	–
		disTSEB	90	421	21	52	67	0.158	0.847	–	–
		ESVEP	89	420	4	55	64	0.153	0.829	–	–
	forest	METRIC	13	459	4	20	27	0.059	0.994	–	–
		TSEB-PT	13	459	7	20	26	0.057	0.995	–	–
		disTSEB	13	459	10	22	28	0.061	0.995	–	–
		ESVEP	12	448	13	20	26	0.058	0.996	–	–
G	cropland	METRIC	81	40	22	41	50	1.255	0.275	–	–
		TSEB-PT	124	41	61	67	75	1.857	0.278	–	–
		disTSEB	122	39	52	60	67	1.707	0.299	–	–
		ESVEP	121	41	68	73	86	2.097	0.273	–	–
	grassland	METRIC	104	143	-64	66	72	0.504	0.810	–	–
		TSEB-PT	104	143	-25	34	43	0.300	0.782	–	–
		disTSEB	96	141	-31	39	51	0.363	0.715	–	–
		ESVEP	102	145	-25	36	43	0.297	0.763	–	–
	savanna	METRIC	52	57	16	20	24	0.417	0.971	–	–
		TSEB-PT	90	77	18	23	30	0.392	0.706	–	–
		disTSEB	90	77	-1	30	37	0.476	0.436	–	–
		ESVEP	89	78	26	29	34	0.443	0.775	–	–
	forest	METRIC	7	-2	18	18	27	-11.557	0.865	–	–
		TSEB-PT	7	-2	19	19	23	-10.016	0.799	–	–
		disTSEB	7	-2	12	12	16	-6.761	0.801	–	–
		ESVEP	7	-2	15	15	18	-7.710	0.795	–	–

Table 1.9: Crop type dependent errors for METRIC, TSEB and ESVEP modelled fluxes using Artificial Neural Network sharpened temperatures. N, number of valid cases;  $\overline{\text{Obs.}}$ ; mean of observed values ( $\text{W m}^{-2}$ ); bias, mean difference between predicted and observed ( $\text{W m}^{-2}$ ); MAE, Mean Absolute Error ( $\text{W m}^{-2}$ ), RMSE, Root Mean Square Error ( $\text{W m}^{-2}$ ); rRMSE, Relative RMSE (-); r, Pearson correlation coefficient (-); % points in plausible range from closure error; RDE, Range Difference Error ( $\text{W m}^{-2}$ )

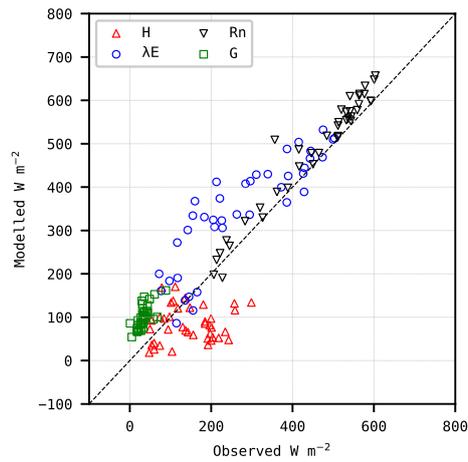
variable	crop type	model	N	$\overline{\text{Obs.}}$	bias	MAE	RMSE	rRMSE	r	% points	RDE
H	herbaceous	METRIC	39	123	65	82	97	0.789	0.439	29	30
		TSEB-PT	39	123	-56	65	84	0.683	0.436	7	83
		disTSEB	38	123	-54	61	81	0.660	0.524	8	80
		ESVEP	36	122	16	61	80	0.657	0.454	19	52
	woody	METRIC	40	146	28	80	105	0.720	-0.121	11	67
		TSEB-PT	66	172	-56	70	85	0.496	0.596	14	82
		disTSEB	66	172	-47	68	82	0.479	0.554	12	77
		ESVEP	50	159	136	180	252	1.591	0.440	5	193
$\lambda E$	herbaceous	METRIC	35	246	-74	94	117	0.475	0.514	19	31
		TSEB-PT	35	246	0	56	70	0.283	0.712	20	52
		disTSEB	34	246	4	55	71	0.287	0.727	19	57
		ESVEP	33	243	-52	70	94	0.386	0.581	20	32
	woody	METRIC	40	275	-18	82	108	0.392	0.601	9	73
		TSEB-PT	66	235	49	65	84	0.358	0.840	17	80
		disTSEB	66	235	58	71	91	0.387	0.819	12	87
		ESVEP	43	261	-71	130	154	0.593	0.368	6	103
$R_n$	herbaceous	METRIC	41	413	-7	26	36	0.087	0.968	-	-
		TSEB-PT	41	413	0	25	34	0.083	0.970	-	-
		disTSEB	39	408	5	26	36	0.087	0.968	-	-
		ESVEP	40	411	6	25	33	0.081	0.973	-	-
	woody	METRIC	40	455	21	26	35	0.077	0.977	-	-
		TSEB-PT	83	437	35	38	45	0.103	0.969	-	-
		disTSEB	83	437	47	50	57	0.131	0.961	-	-
		ESVEP	81	440	15	38	45	0.103	0.927	-	-
G	herbaceous	METRIC	41	45	2	39	46	1.024	0.490	-	-
		TSEB-PT	41	45	50	62	73	1.614	0.312	-	-
		disTSEB	39	42	46	58	69	1.634	0.316	-	-
		ESVEP	40	46	35	49	57	1.240	0.551	-	-
	woody	METRIC	40	34	42	44	53	1.570	0.339	-	-
		TSEB-PT	83	38	66	70	76	2.000	0.448	-	-
		disTSEB	83	38	55	61	66	1.744	0.453	-	-
		ESVEP	81	38	85	85	97	2.527	0.455	-	-

Table 1.10: Climate dependence of errors for METRIC, TSEB and ESVEP modelled fluxes using Artificial Neural Network sharpened temperatures. N, number of valid cases;  $\overline{\text{Obs.}}$ ; mean of observed values ( $\text{W m}^{-2}$ ); bias, mean difference between predicted and observed ( $\text{W m}^{-2}$ ); MAE, Mean Absolute Error ( $\text{W m}^{-2}$ ), RMSE, Root Mean Square Error ( $\text{W m}^{-2}$ ); rRMSE, Relative RMSE (-); r, Pearson correlation coefficient (-); % points in plausible range from closure error; RDE, Range Difference Error ( $\text{W m}^{-2}$ )

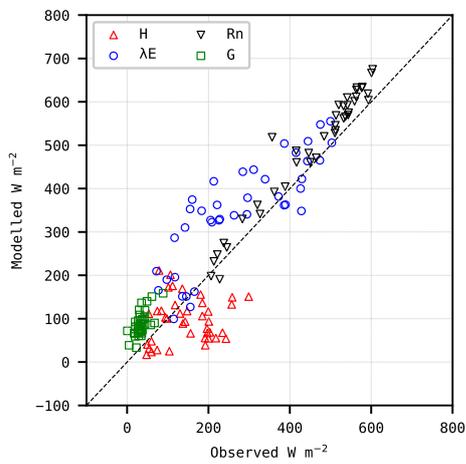
variable	climate	model	N	$\overline{\text{Obs.}}$	bias	MAE	RMSE	rRMSE	r	% points	RDE
H	semi-arid	METRIC	177	191	2	91	135	0.704	0.315	23.7	118
		TSEB-PT	242	193	-44	61	74	0.383	0.737	18.0	71
		disTSEB	234	194	-35	59	72	0.373	0.723	21.0	68
		ESVEP	201	183	84	161	236	1.294	0.386	12.8	199
	temperate	METRIC	63	109	88	106	145	1.324	0.340	66.7	60
		TSEB-PT	67	116	-46	60	82	0.711	0.449	25.4	81
		disTSEB	66	116	-43	57	79	0.677	0.523	27.3	77
		ESVEP	61	114	23	74	96	0.843	0.440	47.5	62
$\lambda E$	semi-arid	METRIC	168	182	22	101	132	0.725	0.343	28.1	113
		TSEB-PT	241	160	42	72	90	0.560	0.743	34.4	75
		disTSEB	234	160	47	74	93	0.583	0.733	30.8	81
		ESVEP	164	174	-2	110	134	0.770	0.321	27.3	99
	temperate	METRIC	56	254	-70	100	129	0.508	0.597	48.2	52
		TSEB-PT	63	257	18	61	83	0.324	0.867	49.2	74
		disTSEB	62	257	25	63	87	0.337	0.875	43.5	80
		ESVEP	58	248	-42	86	113	0.458	0.713	53.4	66
$R_n$	semi-arid	METRIC	180	473	-19	38	49	0.103	0.920	-	-
		TSEB-PT	261	449	13	42	54	0.121	0.881	-	-
		disTSEB	253	447	22	46	59	0.132	0.875	-	-
		ESVEP	258	450	4	42	54	0.119	0.879	-	-
	temperate	METRIC	70	411	4	27	38	0.091	0.974	-	-
		TSEB-PT	70	411	9	26	37	0.091	0.976	-	-
		disTSEB	68	409	14	29	40	0.098	0.975	-	-
		ESVEP	66	405	11	24	34	0.084	0.981	-	-
G	semi-arid	METRIC	180	104	-23	52	61	0.590	0.487	-	-
		TSEB-PT	261	94	16	42	53	0.568	0.525	-	-
		disTSEB	253	92	6	44	54	0.586	0.484	-	-
		ESVEP	258	95	24	49	63	0.671	0.361	-	-
	temperate	METRIC	64	37	3	31	39	1.072	0.534	-	-
		TSEB-PT	64	37	38	48	61	1.651	0.423	-	-
		disTSEB	62	34	32	44	56	1.641	0.407	-	-
		ESVEP	61	37	30	40	49	1.324	0.602	-	-



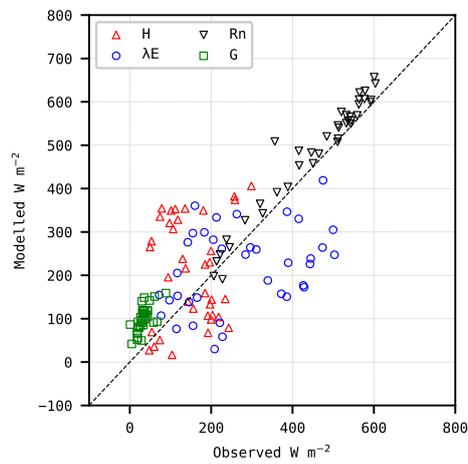
(a) METRIC



(b) TSEB-PT

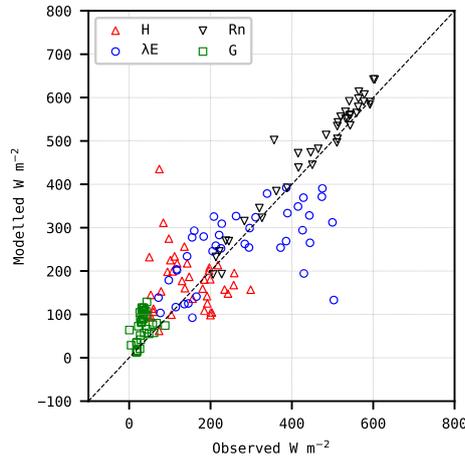


(c) disTSEB

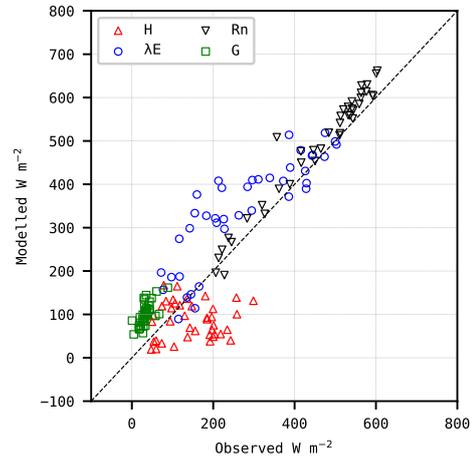


(d) ESVEP

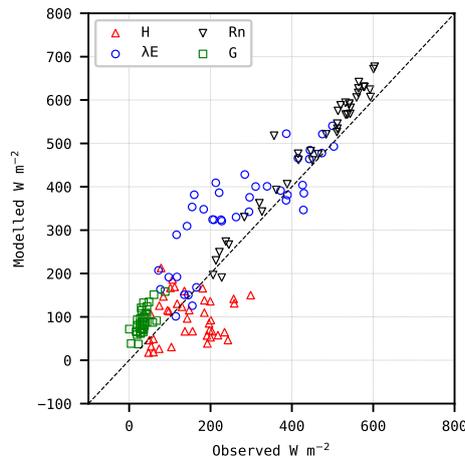
Figure 1.4: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Decision Trees for Borden



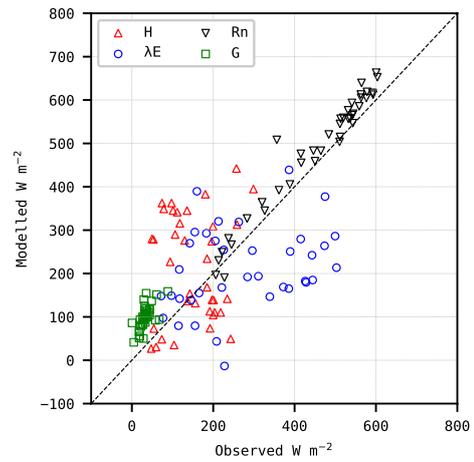
(a) METRIC



(b) TSEB-PT

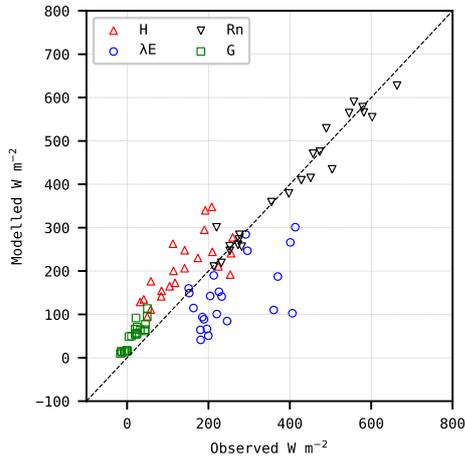


(c) disTSEB

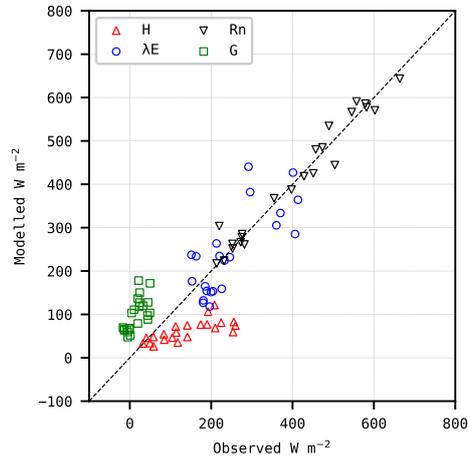


(d) ESVEP

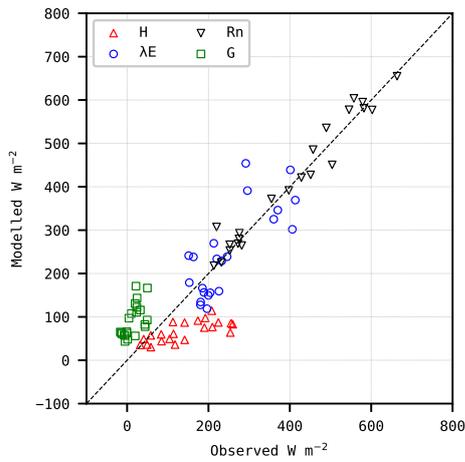
Figure 1.5: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Artificial Neural Network for Borden



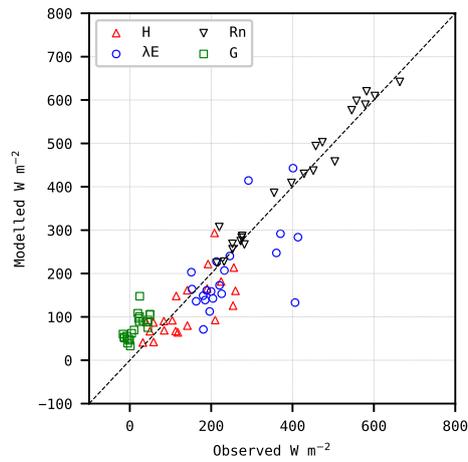
(a) METRIC



(b) TSEB-PT

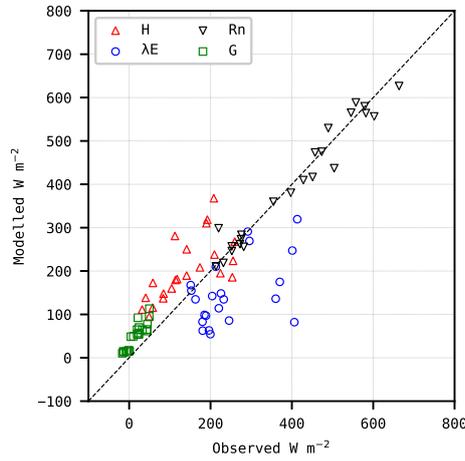


(c) distTSEB

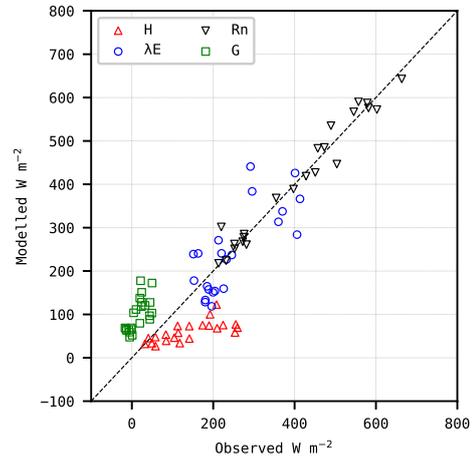


(d) ESVEP

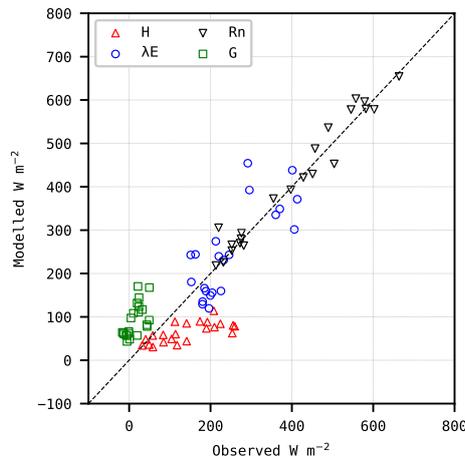
Figure 1.6: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Decision Trees for Choptank



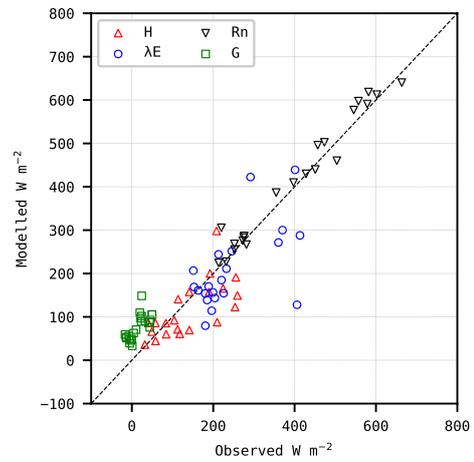
(a) METRIC



(b) TSEB-PT

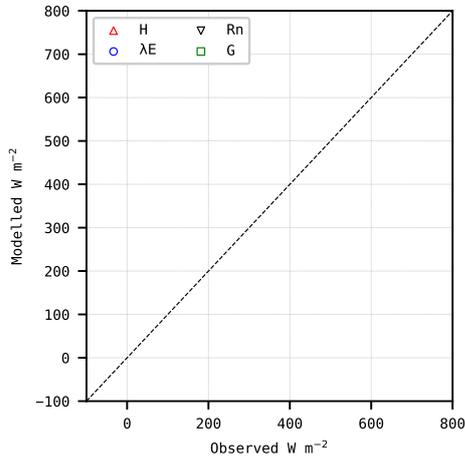


(c) disTSEB

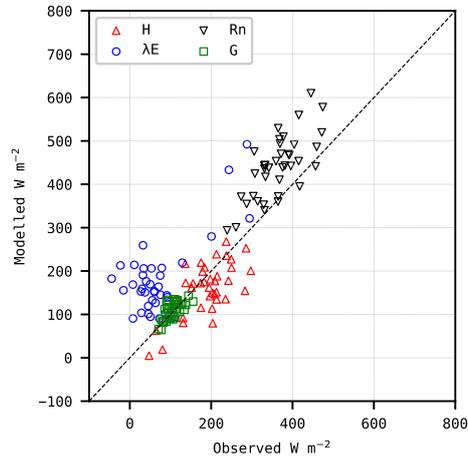


(d) ESVEP

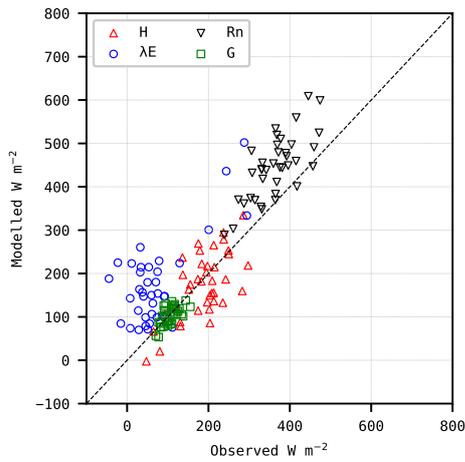
Figure 1.7: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Artificial Neural Network for Choptank



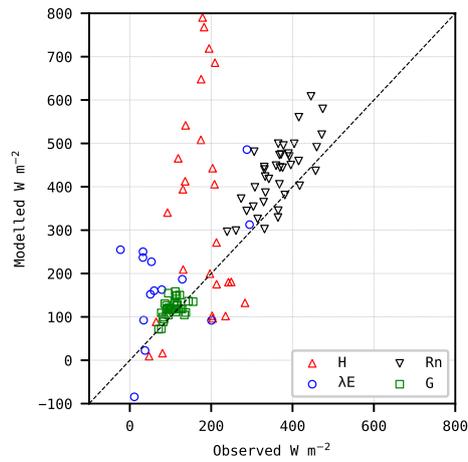
(a) METRIC



(b) TSEB-PT

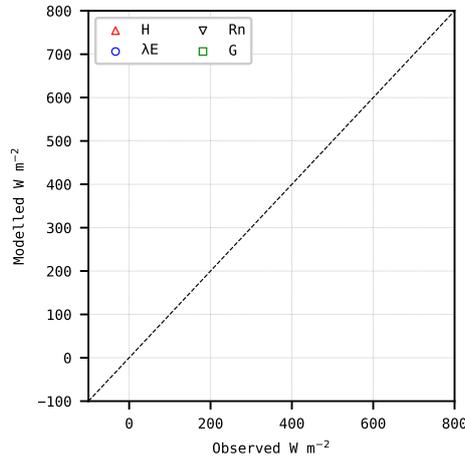


(c) distTSEB

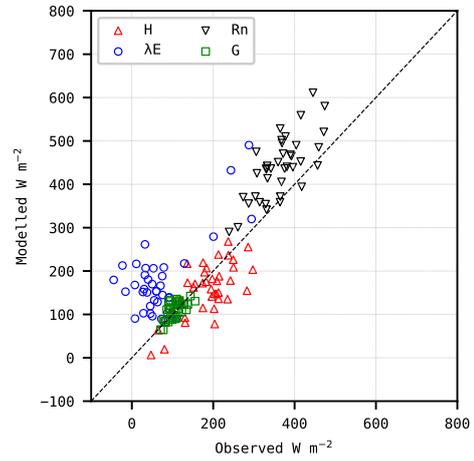


(d) ESVEP

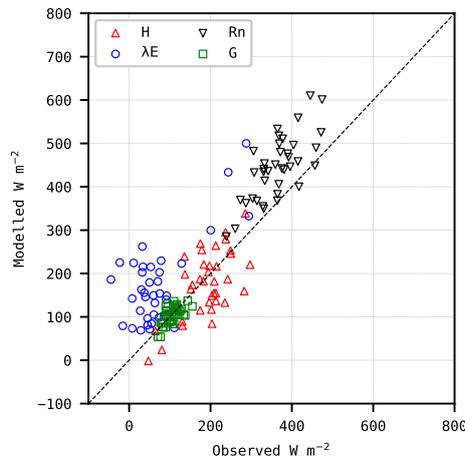
Figure 1.8: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Decision Trees for Dahra



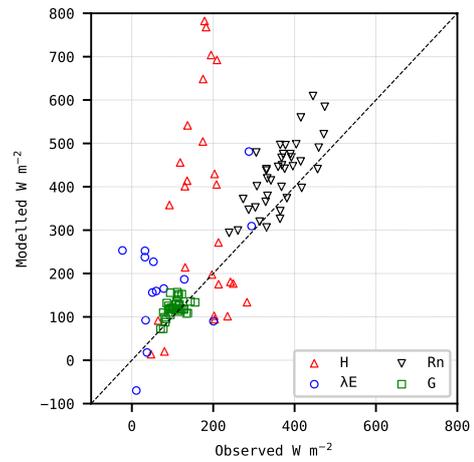
(a) METRIC



(b) TSEB-PT

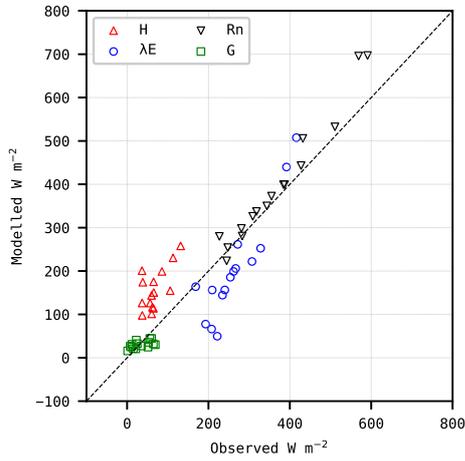


(c) disTSEB

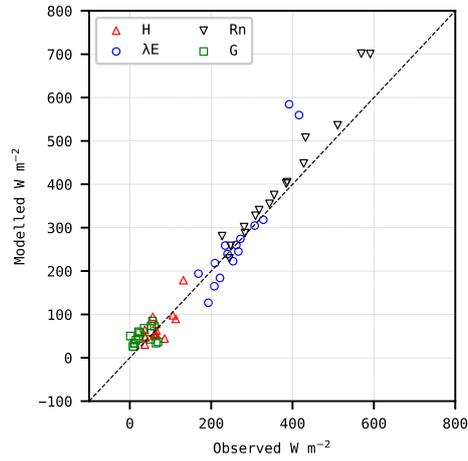


(d) ESVEP

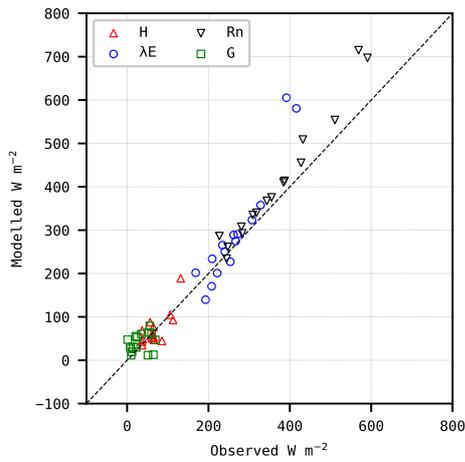
Figure 1.9: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Artificial Neural Network for Dahra



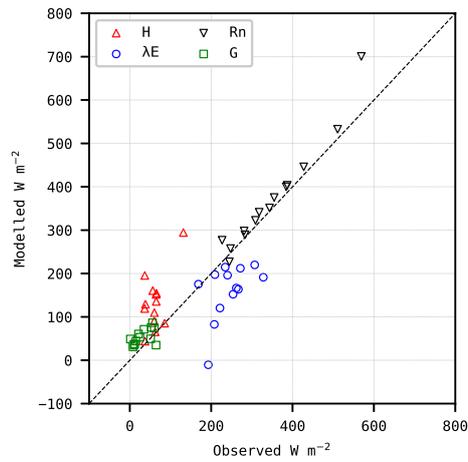
(a) METRIC



(b) TSEB-PT

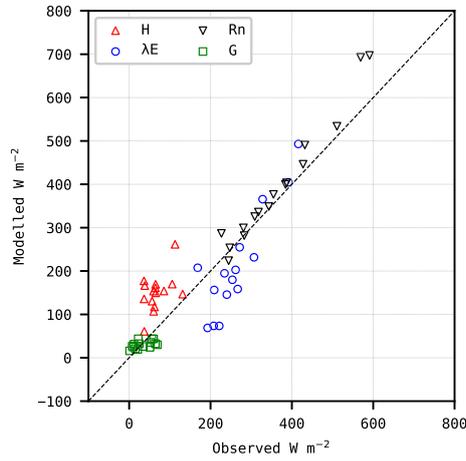


(c) distTSEB

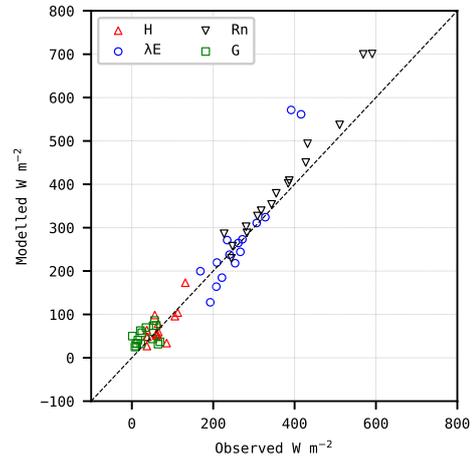


(d) ESVEP

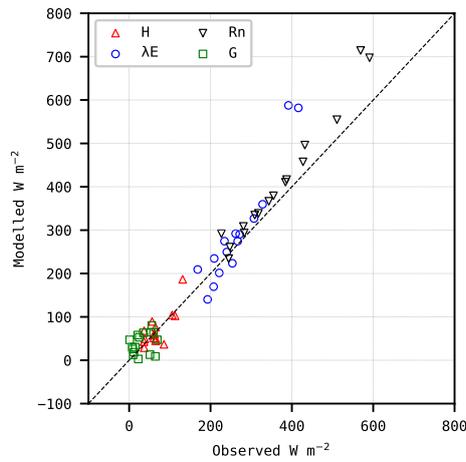
Figure 1.10: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Decision Trees for Grillenburg



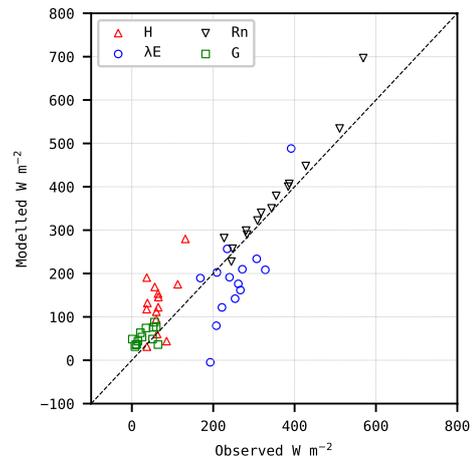
(a) METRIC



(b) TSEB-PT

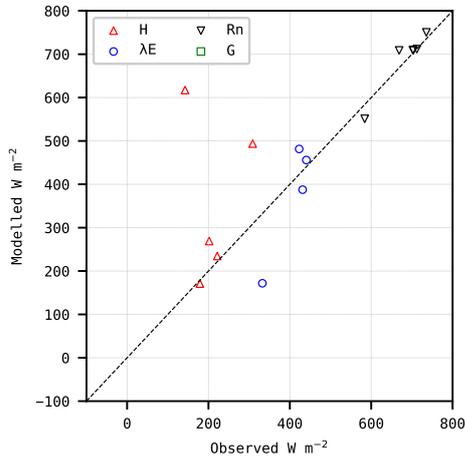


(c) disTSEB

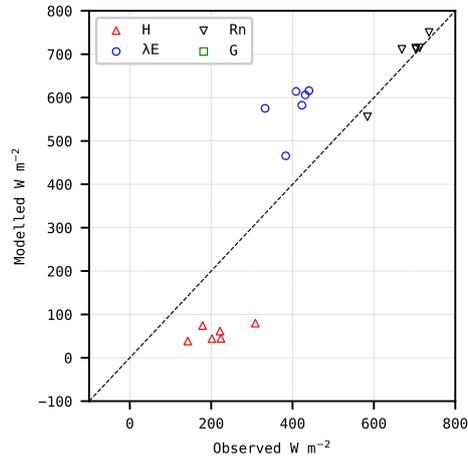


(d) ESVEP

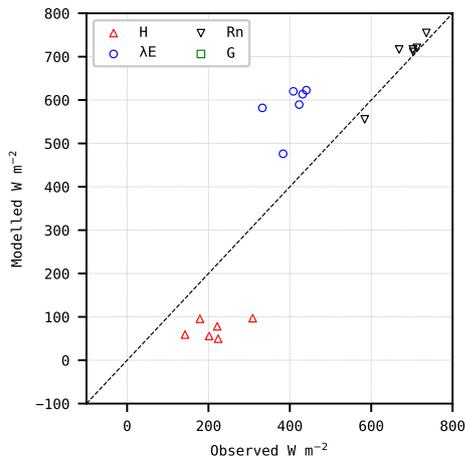
Figure 1.11: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Artificial Neural Network for Grillenburg



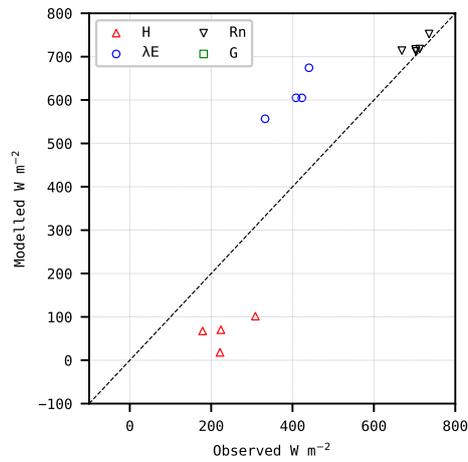
(a) METRIC



(b) TSEB-PT

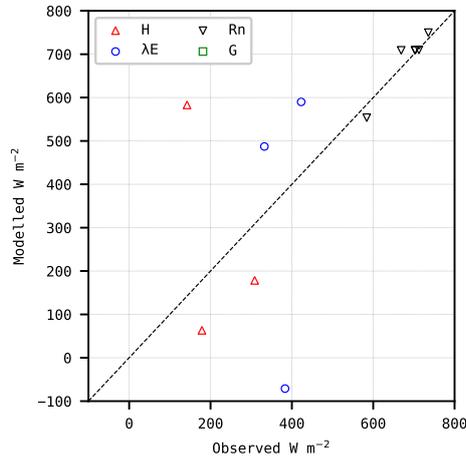


(c) distTSEB

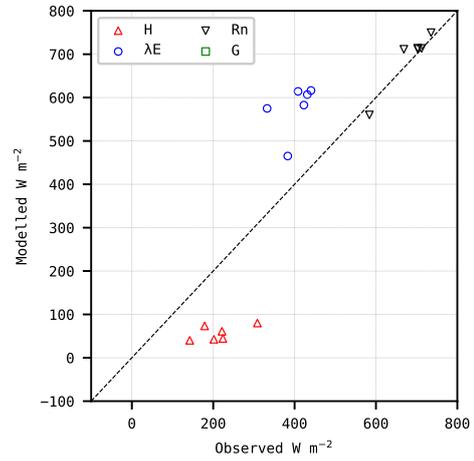


(d) ESVEP

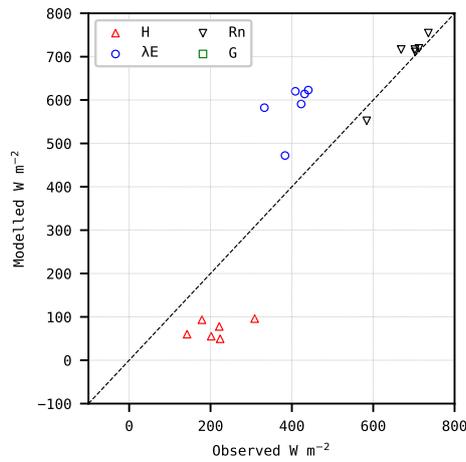
Figure 1.12: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Decision Trees for Harvard



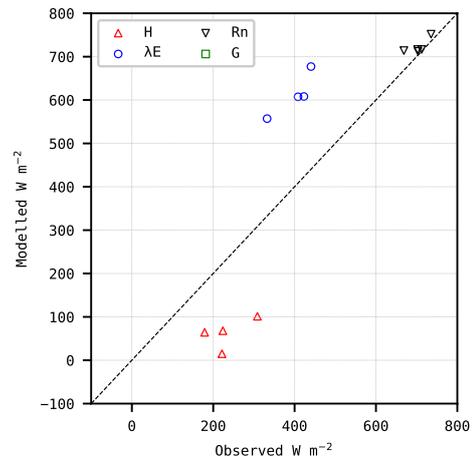
(a) METRIC



(b) TSEB-PT

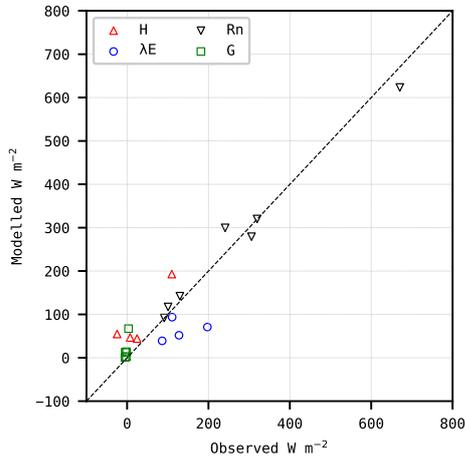


(c) disTSEB

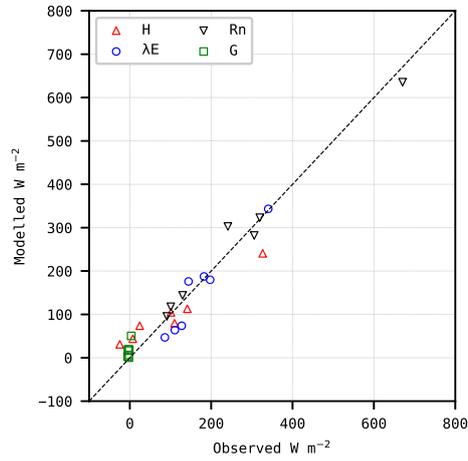


(d) ESVEP

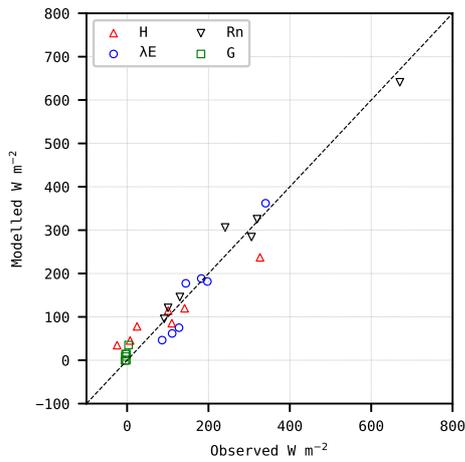
Figure 1.13: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Artificial Neural Network for Harvard



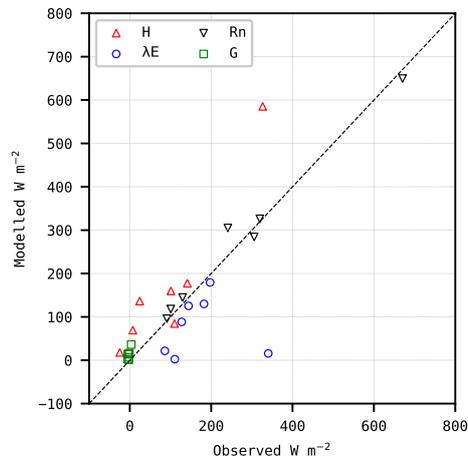
(a) METRIC



(b) TSEB-PT

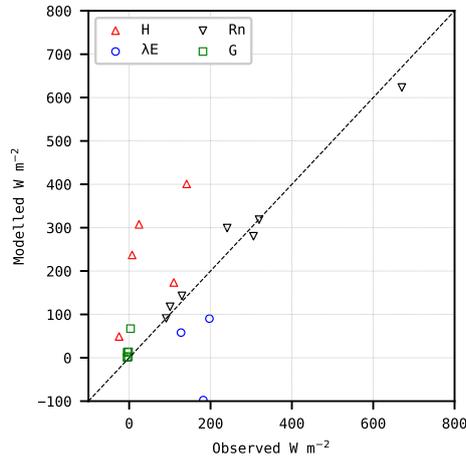


(c) distTSEB

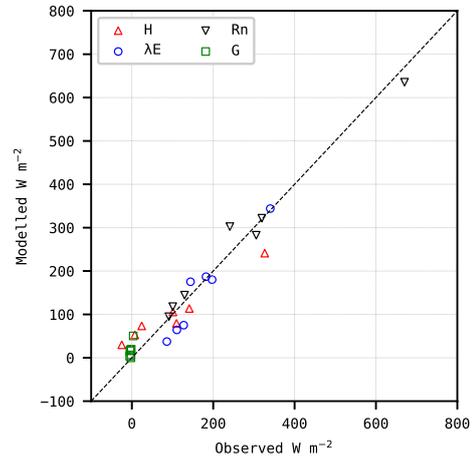


(d) ESVEP

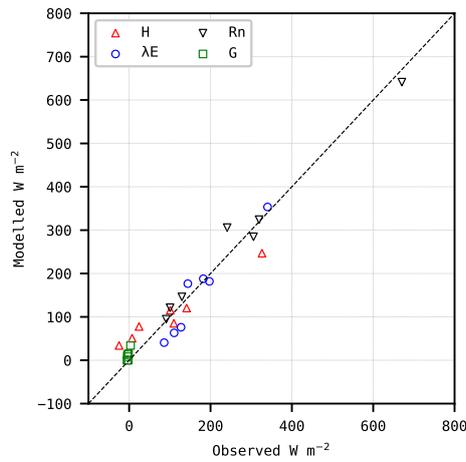
Figure 1.14: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Decision Trees for Hyltemossa



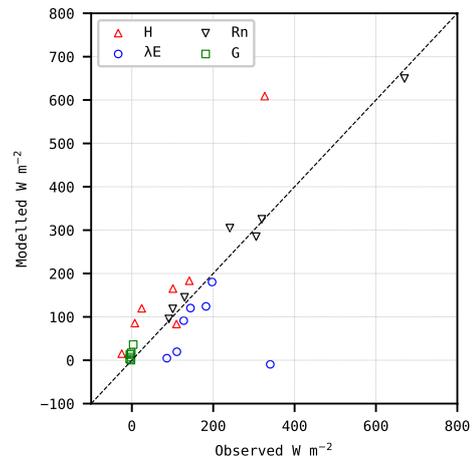
(a) METRIC



(b) TSEB-PT

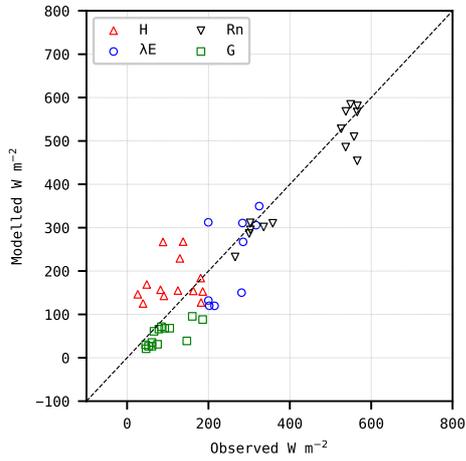


(c) disTSEB

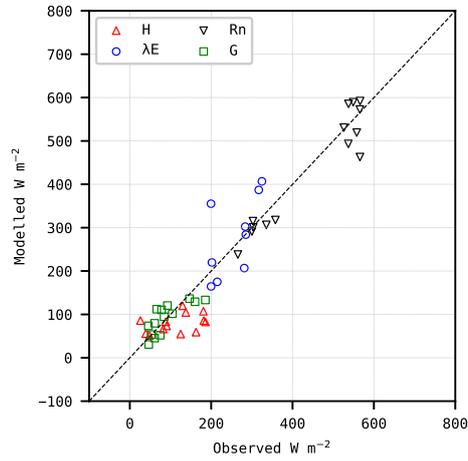


(d) ESVEP

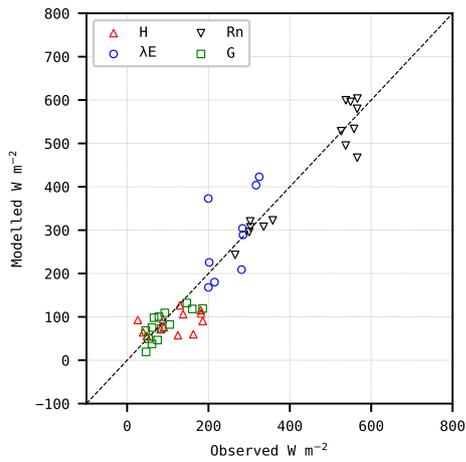
Figure 1.15: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Artificial Neural Network for Hyltemossa



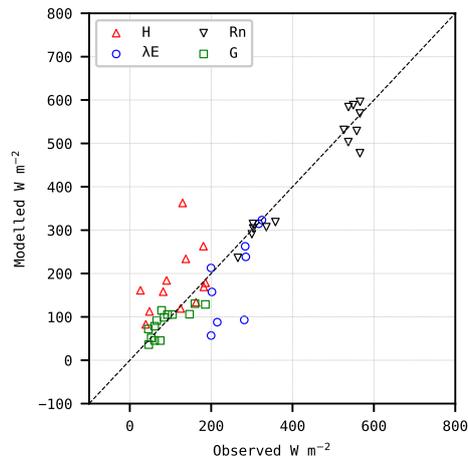
(a) METRIC



(b) TSEB-PT

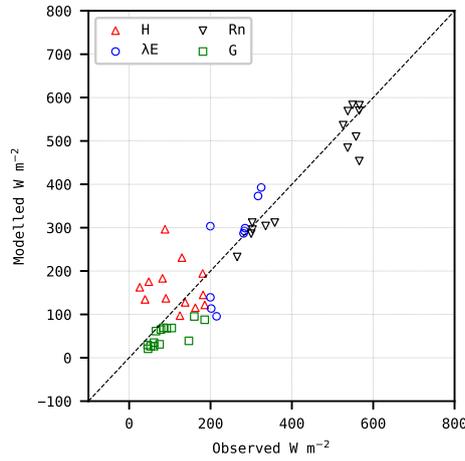


(c) distTSEB

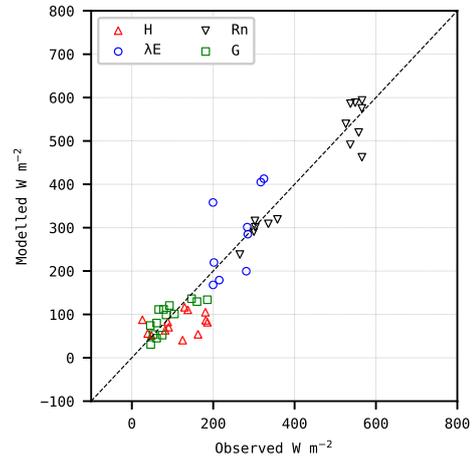


(d) ESVEP

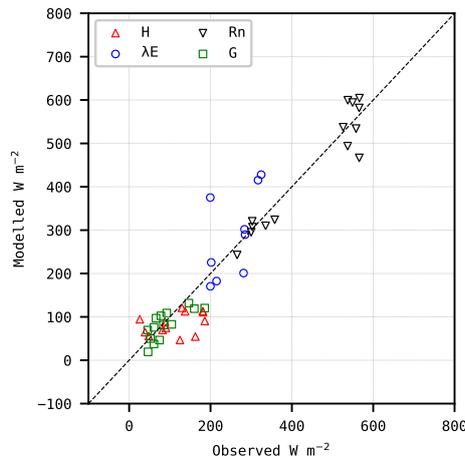
Figure 1.16: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Decision Trees for Klingenberg



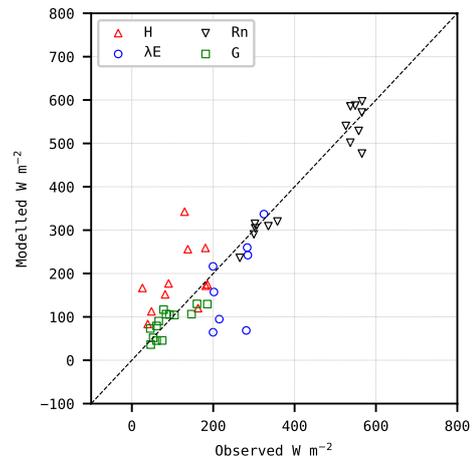
(a) METRIC



(b) TSEB-PT

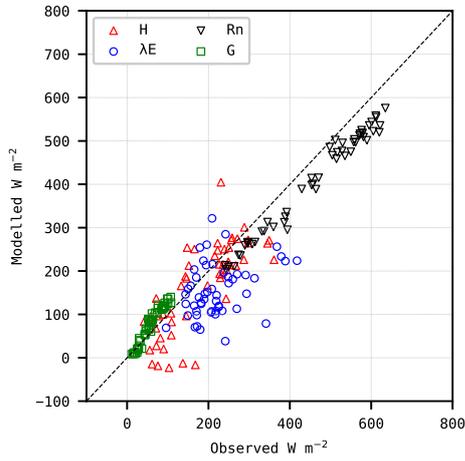


(c) disTSEB

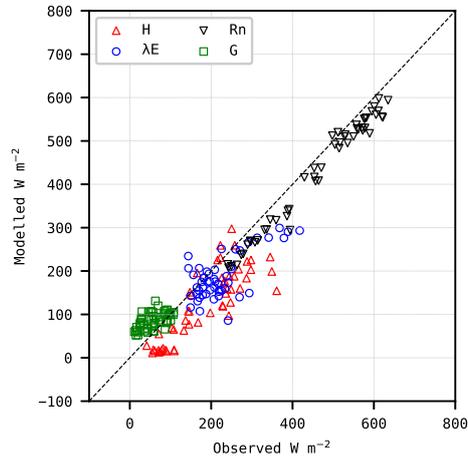


(d) ESVEP

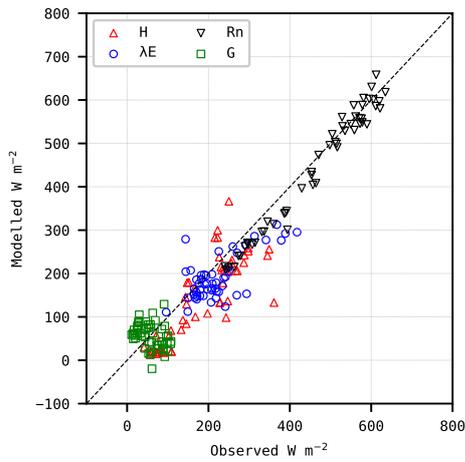
Figure 1.17: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Artificial Neural Network for Klingenberg



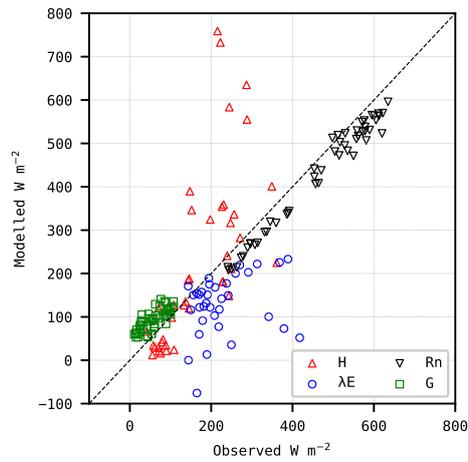
(a) METRIC



(b) TSEB-PT

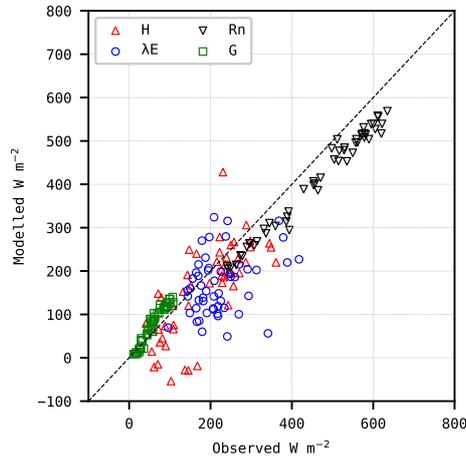


(c) distTSEB

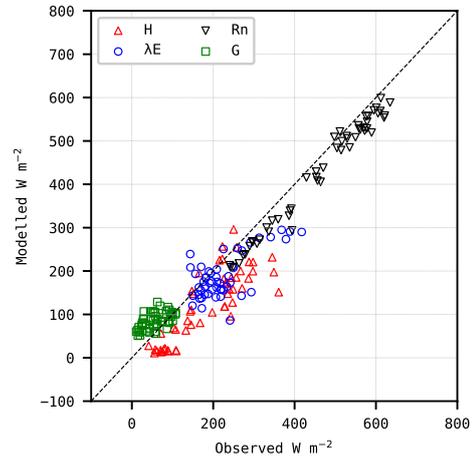


(d) ESVEP

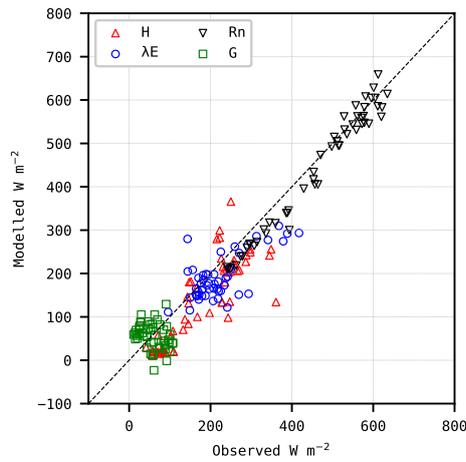
Figure 1.18: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Decision Trees for Majadas



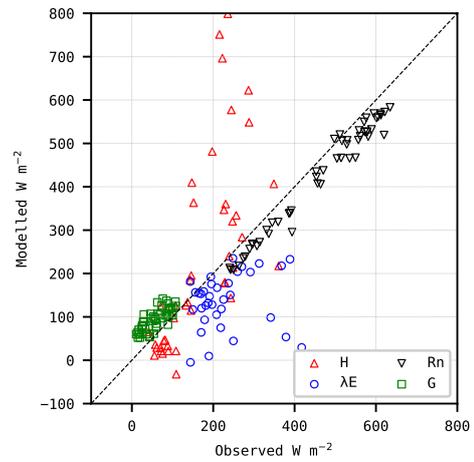
(a) METRIC



(b) TSEB-PT

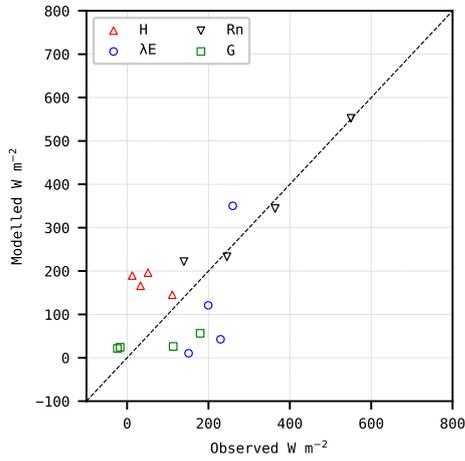


(c) disTSEB

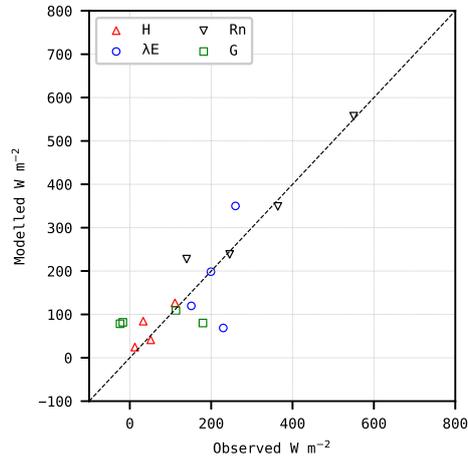


(d) ESVEP

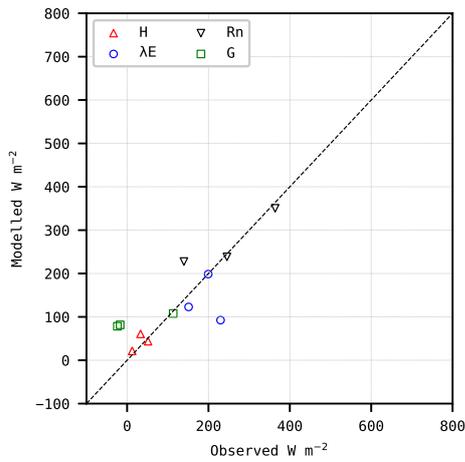
Figure 1.19: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Artificial Neural Network for Majadas de Tiétar



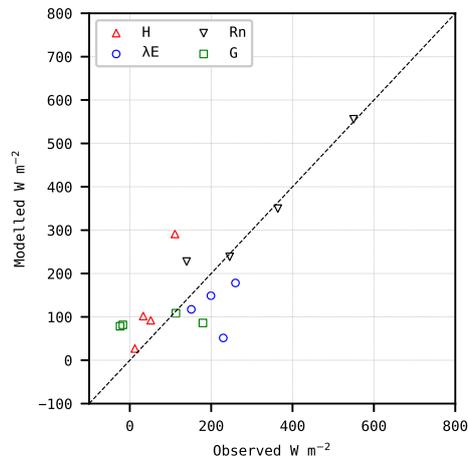
(a) METRIC



(b) TSEB-PT

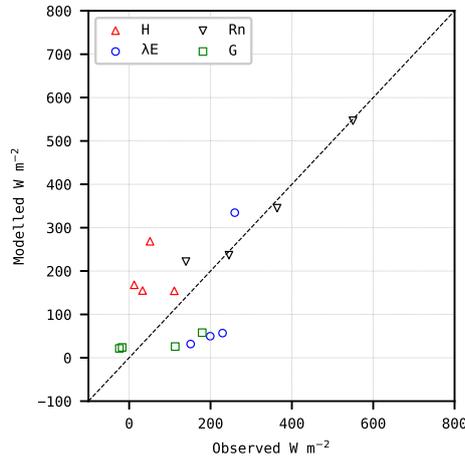


(c) distTSEB

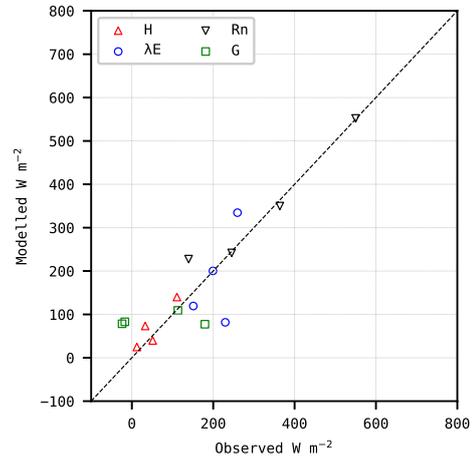


(d) ESVEP

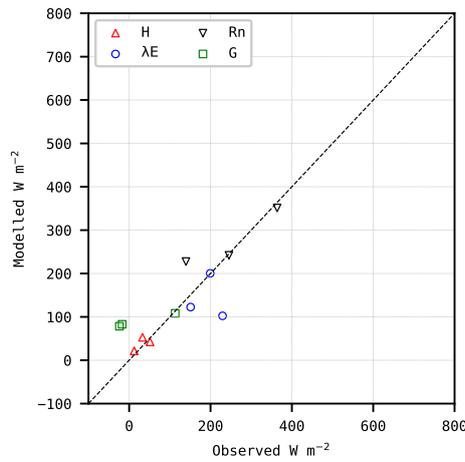
Figure 1.20: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Decision Trees for Selhausen



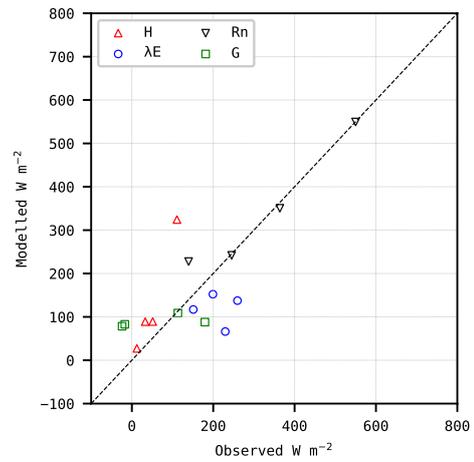
(a) METRIC



(b) TSEB-PT

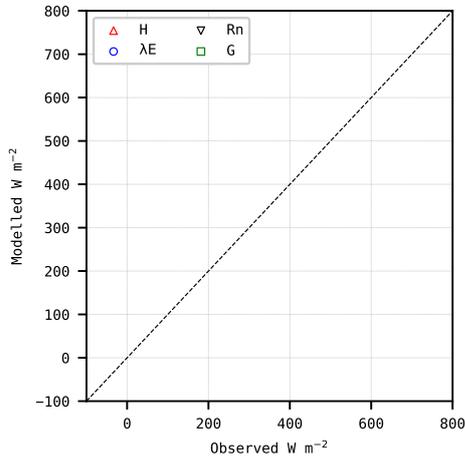


(c) disTSEB

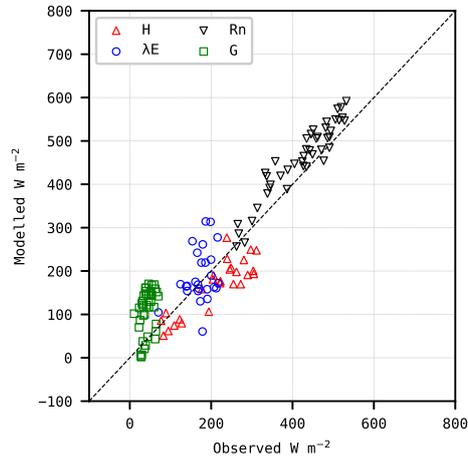


(d) ESVEP

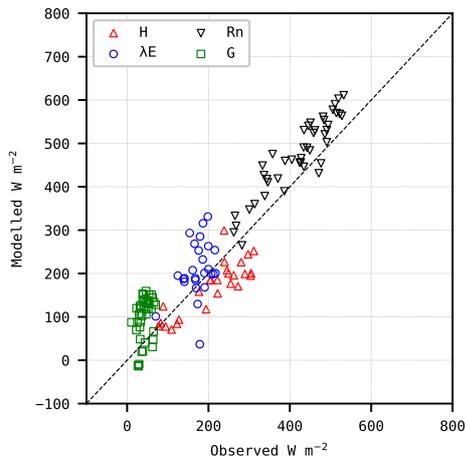
Figure 1.21: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Artificial Neural Network for Selhausen



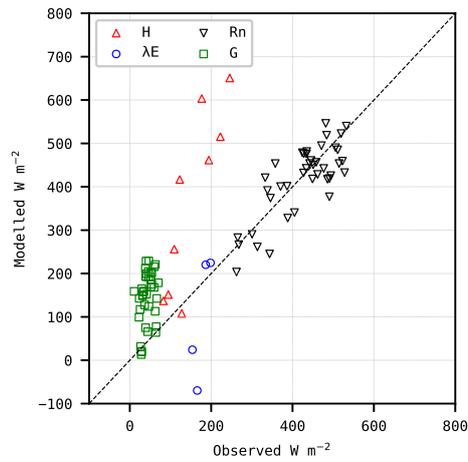
(a) METRIC



(b) TSEB-PT

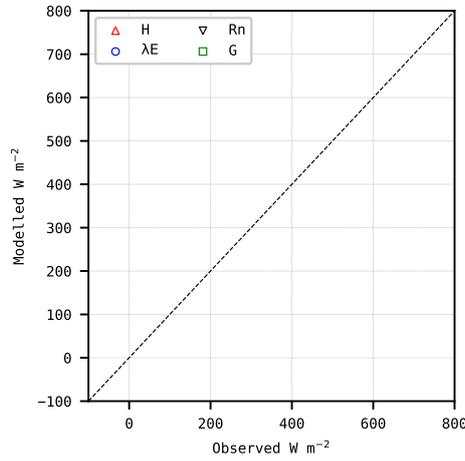


(c) distTSEB

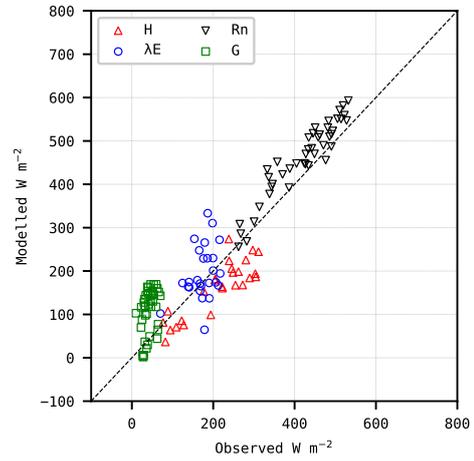


(d) ESVEP

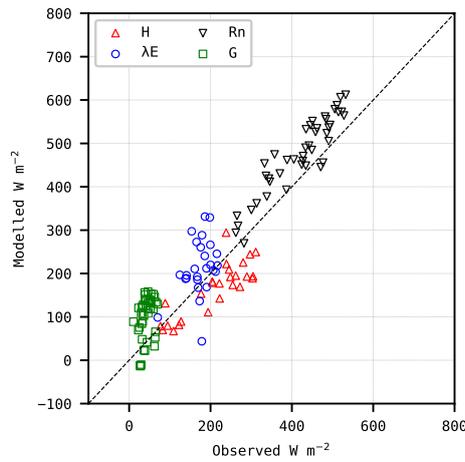
Figure 1.22: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Decision Trees for Taous



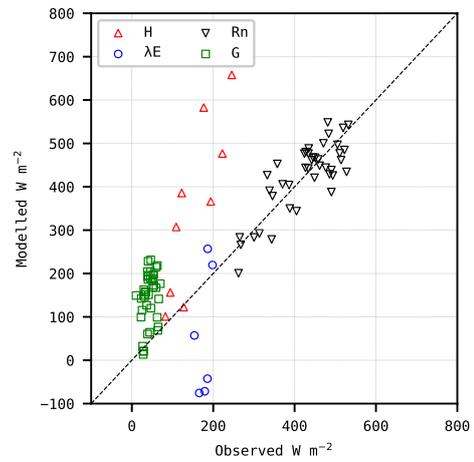
(a) METRIC



(b) TSEB-PT

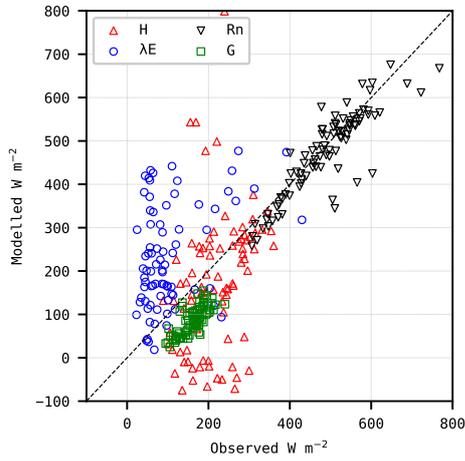


(c) disTSEB

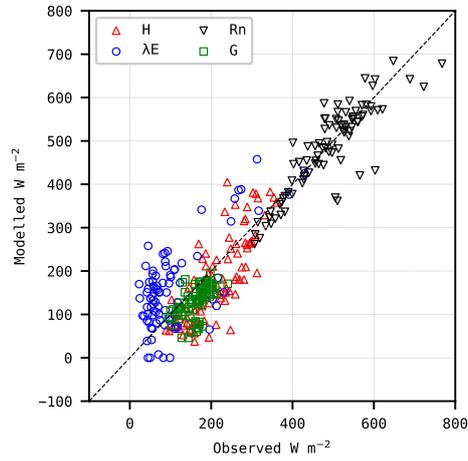


(d) ESVEP

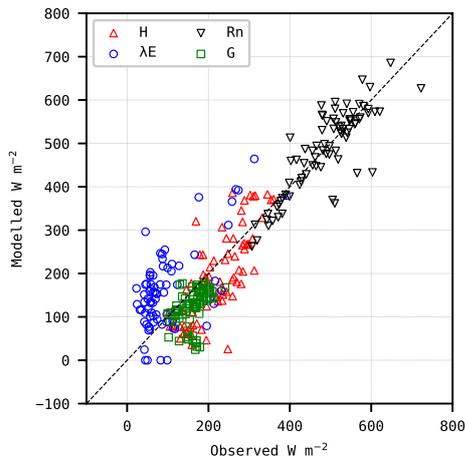
Figure 1.23: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Artificial Neural Network for Taous



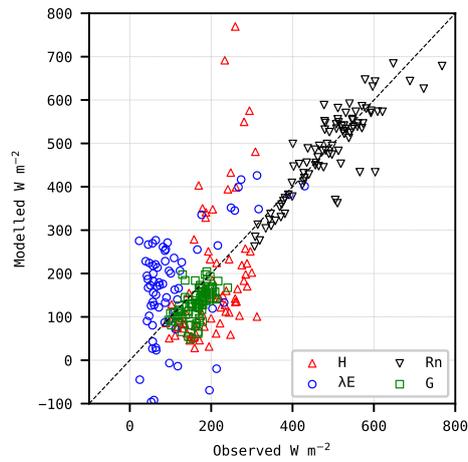
(a) METRIC



(b) TSEB-PT

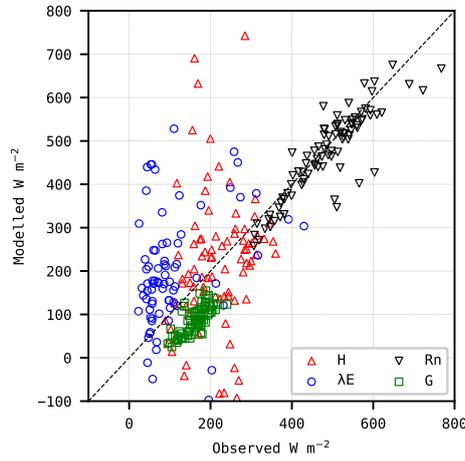


(c) distTSEB

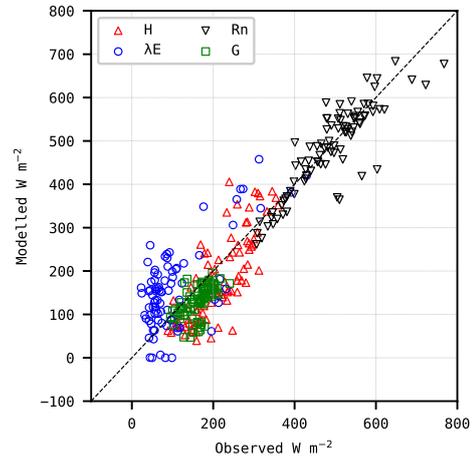


(d) ESVEP

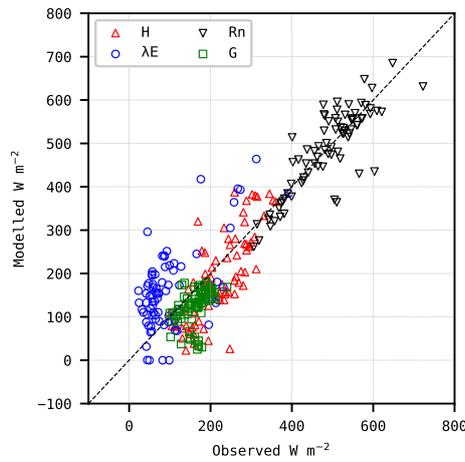
Figure 1.24: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Decision Trees for Walnut Gulch



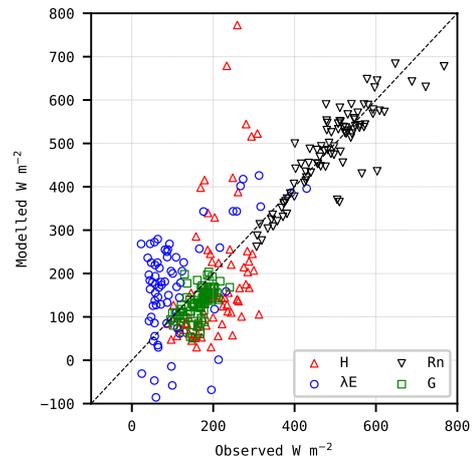
(a) METRIC



(b) TSEB-PT



(c) disTSEB



(d) ESVEP

Figure 1.25: Model scatterplot of predicted vs. EC observed using sharpened  $T_{rad}$  with Artificial Neural Network for Walnut Gulch



## Chapter 2

# Model benchmarking and selection

At the beginning of the project, three evapotranspiration (ET) models and two land surface temperature (LST) sharpening methods were selected for prototyping and benchmarking, after a literature review [1]. The purpose was to find an optimal combination for estimating high-resolution (10 - 20 m) ET using Sentinel-2 (S2) and Sentinel-3 (S3) observations. This combination will be implemented as an open-source software and made freely available to the Earth Observation (EO) community in the second phase of SEN-ET project. The three selected ET models were: METRIC [32, 57], TSEB-PT [36, 50] and ESVEP [44]. Both of the sharpening approaches used the Data Mining Sharpener (DMS) [6, 4] framework with Bagging ensemble of either Neural Networks (NN) or Decision Trees (DT). The prototypes of all the selected ET models and sharpening methods were implemented in Python and are available in GitHub: <https://github.com/hectornieto/pyMETRIC>, <https://github.com/hectornieto/pyTSEB>, <https://github.com/dhi-gras/py-esvep> and <https://github.com/radosuav/pyDMS>.

The high-resolution fluxes used for the benchmarking were produced using only publicly available and global data sources. The low resolution LST was estimated using S3 SLSTR L1 data and the approach of [70], since SLSTR L2 data was not available for the whole of the study period. The LST was then sharpened to S2 resolution using atmospherically corrected 10m and 20 m S2 bands, SRTM digital elevation model and solar illumination conditions, before being used as an input to the ET models. S2 data were also used to derive plant biophysical parameters, such as leaf area index (LAI) or albedo, predominantly using SNAP software functionality. Meteorological inputs were obtained from European Centre for Medium Range Weather Forecast (ECMWF) ERA-5 reanalysis dataset, and ancillary inputs (e.g. vegetation height) were based on ESA Climate Change Initiative land-cover map for 2015 and a look-up table.

### 2.1 Benchmarking criteria

The selection of the optimal model was based on 6 evaluation criteria.

1. Accuracy - evaluate models on their performance when run in "operational mode" (standard setting) and in "research mode".

This involves looking at overall model accuracy across all validation sites when run with globally available datasets (i.e. operational mode). Special focus should be placed on accuracy in the SEN-ET focus application (i.e. agriculture). Finally, it

should be evaluated how the accuracy would improve if localised input data were used in the models (i.e. research mode).

2. Applicability - evaluate application of models in certain areas and conditions.

This criteria looks at issues such as, whether the models can be applied globally and how they perform in various vegetation types and climatological conditions.

3. Data Accessibility - evaluate access to input data.

In this criteria we consider whether the models require data which is not free and open or which is not operationally acquired and produced. In addition, the number of separate data sources is taken into account.

4. Functionality - evaluate how user friendly it is to run the models.

This criteria evaluates if accurate results can be obtained when running the models with default values or whether user interaction is required and how much EO or ET knowledge is required for such customisation. In addition, it considers the outputs which the models produce and the ease with which the models can be integrated with SNAP.

5. Performance - evaluate how computationally performant are the models.

We look at the computing execution time (e.g. over a whole Sentinel-2 scene), robustness in returning valid data, stability when running over long time series, any special software or hardware requirements and suitability for execution in a cloud environment.

6. Other considerations - additional items to consider regarding the models.

This includes issues such as popularity of the model in the research and applied communities, existing implementations of the models and potential for uptake and community development.

Each of the models was quantitatively (when relevant) and qualitatively evaluated against those six criteria by the members of the consortium and the project Technical Officer. Afterwards, the models were ranked on how well they satisfy the criteria and final decision on model selection was taken.

## 2.2 Benchmarking results

The evaluation of the ET models and LST sharpening approaches based on the six criteria is presented in the sections below. This is followed by a summary and selection of methods for implementation in the second phase of SEN-ET project.

### 2.2.1 Accuracy

The accuracy of the models was validated at 11 flux tower sites located in Europe, Africa and North America (see Figure 2.1). Five of those sites were located in agricultural fields (three in herbaceous crops and two in woody crops), two in savannah, two in grassland and one each in conifer and broadleaved forests. Validation was performed using all the

### Validation Sites Location



Figure 2.1: Map showing locations of validation sites. Only Primary sites were used for accuracy assesment.

Sentinel-2 observations in 2017 which were cloud free in the vicinity of the towers and Sentinel-3 observations which were acquired on the day of the Sentinel-2 observation or within four days afterwards.

The two sharpening approaches, produced very similar high-resolution LST maps, although when visual spatial comparison was conducted certain differences were apparent (see Figure 2.2). However, no significant differences between the NN and DT approaches were observed when using their output high-resolution LST to estimate land surface energy fluxes and validating them again flux tower measurements. For brevity, the results obtained using DT sharpener are used throughout the rest of the benchmarking evaluation.

Regarding the ET models, the aggregated results from all the sites are shown in Table 2.1. The results from the agricultural sites are shown in Table 2.2. In both cases, TSEB-PT achieves consistently lower Root Mean Square Error (RMSE) and higher correlation in the turbulent flux (sensible heat -  $H$  and latent heat -  $\lambda E$ ) estimates than either METRIC or ESVEP. The situation is more complex when considering ESVEP and METRIC with the accuracy ranking varying between the two models. For example ESVEP appears to be more accurate in herbaceous crops while METRIC performs better in woody crops. Looking at overall statistics, METRIC produced  $H$  with significantly lower RMSE but also lower correlation, while the RMSE of  $\lambda E$  produced by both models is similar with ESVEP obtaining higher correlation.

Finally, the accuracy of TSEB-PT is expected to improve more than that of other

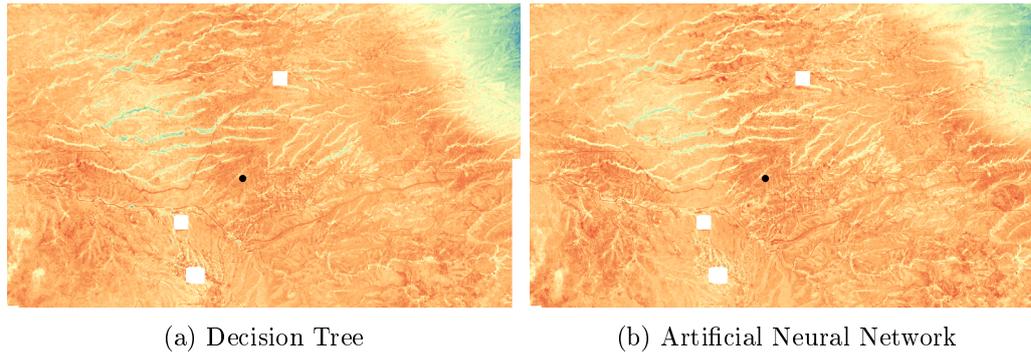


Figure 2.2: Spatial patterns comparison between decision-tree and neural-network sharpened temperatures. Images correspond to one of the validation sites on September 27th, 2017. The black dot shows the location of the flux tower.

models if localised data on parameters such as vegetation height or fraction of vegetation which is green is available.

**Ranking ET models:** 1 - TSEB-PT, 2 -, 3 - METRIC & ESVEP

**Ranking sharpening approaches:** 1 - NN & DT, 2 -

### 2.2.2 Applicability

Applicability was assessed based on the same 11 validation sites as accuracy, but this time grouped by land cover and climatology. Once again the two sharpening approaches performed very similarly in all cases.

The results for the ET models are shown in Table 2.3 (land cover) and Table 2.4 (climatology). For brevity, only the accuracy statistics for H and  $\lambda E$  are shown in those tables. TSEB-PT once again outperforms the other models in accuracy metrics across all land cover classes, with the exception of latent heat flux in forests where METRIC has significantly lower RMSE. However, METRIC was able to produce  $\lambda E$  in 8 cases in the forest class while TSEB-PT produced 13 valid values. When comparing METRIC and ESVEP, the first model produced more accurate results in croplands and savannah, while the second was more accurate in grasslands. Considering climatology, TSEB-PT achieved the most accurate outputs and, especially in semi-arid climate, was able to produce more valid outputs than the other two models. METRIC performed better than ESVEP when estimating H in semi-arid locations, while the situation was reversed in temperate sites. The difference between those two models was not very significant in both climatologies when it came to estimating latent heat flux.

The reason for larger number of valid cases produced by TSEB-PT as compared to both METRIC and ESVEP, particularly in semi-arid climate, is due to intrinsic assumptions by the latter models. Both of them scale the ET value in a given pixel between the values of theoretical hot (no ET) and cold (potential ET) pixels. In METRIC those hot and cold pixels are retrieved from a specified area of interest (AOI) in the satellite image, while in ESVEP they are derived from the pixel itself. In certain locations, such as in rainfed olive groves surrounded by dry natural vegetation, it might not be possible to reliably identify

Table 2.1: Error metrics for METRIC, TSEB-PT and ESVEP modelled fluxes using sharpened temperatures. N, number of valid cases;  $\overline{\text{Obs.}}$ ; mean of observed values ( $\text{W m}^{-2}$ ); bias, mean difference between predicted and observed ( $\text{W m}^{-2}$ ); MAE, Mean Absolute Error ( $\text{W m}^{-2}$ ), RMSE, Root Mean Square Error ( $\text{W m}^{-2}$ ); rRMSE, Relative RMSE (-); r, Pearson correlation coefficient (-).

variable	model	N	$\overline{\text{Obs.}}$	bias	MAE	RMSE	rRMSE	r
H	METRIC	242	171	15	95	146	0.854	0.286
	TSEB-PT	309	176	-44	60	75	0.426	0.734
	ESVEP	259	166	67	137	207	1.246	0.413
$\lambda\text{E}$	METRIC	229	199	9	103	130	0.654	0.334
	TSEB-PT	304	180	36	70	88	0.491	0.770
	ESVEP	219	193	-13	103	127	0.656	0.431
$R_n$	METRIC	250	456	-13	35	46	0.101	0.943
	TSEB-PT	331	441	12	38	51	0.116	0.919
	ESVEP	322	440	5	38	50	0.114	0.918
G	METRIC	244	86	-16	47	56	0.655	0.597
	TSEB-PT	325	83	20	43	55	0.661	0.571
	ESVEP	317	84	26	48	62	0.734	0.485

the hot and cold pixels within the AOI or to derive their values. In those cases, METRIC and ESVEP do not produce valid results. It should be noted that in an irrigated area surrounded by arid landscape, arguably a case of most need of ET estimates, METRIC and ESVEP should perform well.

**Ranking ET models:** 1 - TSEB-PT, 2 -, 3 - METRIC & ESVEP

**Ranking sharpening approaches:** 1 - NN & DT, 2 -

### 2.2.3 Data Accessibility

The two LST sharpening approaches used exactly the same inputs in this study: high-resolution reflectances, DEM and solar illumination conditions and low-resolution LST.

The major inputs of the three ET models are listed in Table 2.5. All the models require inputs from three major data sources: optical satellite observations, thermal satellite observations and meteorological models. In this project, Sentinel-2 is used as the source of optical data, Sentinel-3 as the source of thermal data and ERA-5 dataset from ECMWF as source of meteorological data. Those three data sources are operationally produced and disseminated in the free and open manner through the European Union's Copernicus programme.

TSEB-PT and ESVEP additionally require some parameters which cannot be easily determined through satellite observations or modelling and therefore are set using a land-cover based look-up table. The most important of those is canopy-height since both ESVEP and TSEB-PT are quite sensitive to this parameter. In METRIC, vegetation height is used to separate the land cover classes into tall and short vegetation during end-member pixel search and to compute surface roughness. However, it is expected that METRIC is less sensitive to vegetation height estimation than ESVEP and TSEB-PT.

Table 2.2: Crop type dependence errors for METRIC, TSEB and ESVEP modelled fluxes using sharpened temperatures. N, number of valid cases;  $\overline{\text{Obs.}}$ ; mean of observed values ( $\text{W m}^{-2}$ ); bias, mean difference between predicted and observed ( $\text{W m}^{-2}$ ); MAE, Mean Absolute Error ( $\text{W m}^{-2}$ ), RMSE, Root Mean Square Error ( $\text{W m}^{-2}$ ); rRMSE, Relative RMSE (-); r, Pearson correlation coefficient (-).

variable	crop type	model	N	$\overline{\text{Obs.}}$	bias	MAE	RMSE	rRMSE	r
H	herbaceous	METRIC	39	123	71	80	93	0.759	0.571
		TSEB-PT	39	123	-54	63	82	0.667	0.479
		ESVEP	37	122	19	57	76	0.625	0.497
	woody	METRIC	40	146	25	73	88	0.601	0.014
		TSEB-PT	66	172	-56	68	84	0.490	0.620
		ESVEP	49	156	127	173	244	1.563	0.384
$\lambda E$	herbaceous	METRIC	35	246	-83	98	119	0.482	0.520
		TSEB-PT	35	246	-2	56	70	0.283	0.711
		ESVEP	34	245	-54	69	92	0.375	0.619
	woody	METRIC	40	275	-14	75	89	0.322	0.772
		TSEB-PT	66	235	49	66	84	0.359	0.845
		ESVEP	41	264	-58	119	142	0.538	0.404
$R_n$	herbaceous	METRIC	41	413	-7	26	36	0.087	0.967
		TSEB-PT	41	413	0	25	34	0.083	0.969
		ESVEP	40	411	6	25	33	0.081	0.972
	woody	METRIC	40	455	20	25	35	0.077	0.974
		TSEB-PT	83	437	34	36	44	0.101	0.968
		ESVEP	80	439	10	37	47	0.106	0.915
G	herbaceous	METRIC	41	45	2	39	47	1.026	0.487
		TSEB-PT	41	45	50	62	73	1.608	0.316
		ESVEP	40	46	35	49	57	1.236	0.548
	woody	METRIC	40	34	42	44	53	1.586	0.335
		TSEB-PT	83	38	66	69	75	1.980	0.447
		ESVEP	80	38	86	87	99	2.569	0.451

Table 2.3: Error dependence on land cover for METRIC, TSEB-PT and ESVEP modelled fluxes using sharpened temperatures. N, number of valid cases;  $\overline{\text{Obs.}}$ ; mean of observed values ( $\text{W m}^{-2}$ ); bias, mean difference between predicted and observed ( $\text{W m}^{-2}$ ); MAE, Mean Absolute Error ( $\text{W m}^{-2}$ ), RMSE, Root Mean Square Error ( $\text{W m}^{-2}$ ); rRMSE, Relative RMSE (-); r, Pearson correlation coefficient (-).

variable	land cover	model	N	$\overline{\text{Obs.}}$	bias	MAE	RMSE	rRMSE	r
H	cropland	METRIC	79	135	48	77	91	0.672	0.271
		TSEB-PT	105	154	-55	66	83	0.542	0.617
		ESVEP	86	141	81	123	191	1.348	0.418
	grassland	METRIC	100	195	-16	117	170	0.870	0.221
		TSEB-PT	101	194	-26	51	64	0.328	0.814
		ESVEP	91	185	8	113	161	0.869	0.428
	savanna	METRIC	52	185	-18	52	67	0.362	0.777
		TSEB-PT	90	186	-47	57	69	0.371	0.779
		ESVEP	71	175	139	187	276	1.579	0.476
	forest	METRIC	11	140	221	222	354	2.535	0.528
		TSEB-PT	13	151	-72	94	115	0.763	0.502
		ESVEP	11	147	-12	116	138	0.941	0.480
$\lambda E$	cropland	METRIC	75	262	-46	85	104	0.397	0.614
		TSEB-PT	101	239	31	63	80	0.333	0.795
		ESVEP	75	256	-56	96	122	0.477	0.462
	grassland	METRIC	94	130	99	132	163	1.254	0.306
		TSEB-PT	100	129	47	71	86	0.671	0.765
		ESVEP	84	133	41	102	120	0.902	0.417
	savanna	METRIC	52	221	-65	82	98	0.442	0.352
		TSEB-PT	90	158	25	73	94	0.594	0.493
		ESVEP	49	189	-45	107	136	0.719	0.144
	forest	METRIC	8	269	-50	68	83	0.310	0.934
		TSEB-PT	13	277	71	95	125	0.450	0.969
		ESVEP	11	254	19	133	166	0.654	0.853

Table 2.4: Climate dependence of errors for METRIC, TSEB and ESVEP modelled fluxes using sharpened temperatures. N, number of valid cases;  $\overline{\text{Obs.}}$ ; mean of observed values ( $\text{W m}^{-2}$ ); bias, mean difference between predicted and observed ( $\text{W m}^{-2}$ ); MAE, Mean Absolute Error ( $\text{W m}^{-2}$ ), RMSE, Root Mean Square Error ( $\text{W m}^{-2}$ ); rRMSE, Relative RMSE (-); r, Pearson correlation coefficient (-).

variable	climate	model	N	$\overline{\text{Obs.}}$	bias	MAE	RMSE	rRMSE	r
H	semi-arid	METRIC	177	192	-16	90	136	0.710	0.326
		TSEB-PT	242	193	-44	60	73	0.380	0.744
		ESVEP	198	182	80	157	231	1.268	0.369
	temperate	METRIC	65	113	101	107	169	1.496	0.486
		TSEB-PT	67	116	-45	59	81	0.700	0.472
		ESVEP	61	114	25	71	93	0.817	0.458
$\lambda\text{E}$	semi-arid	METRIC	171	180	37	108	137	0.760	0.298
		TSEB-PT	241	160	41	72	90	0.560	0.745
		ESVEP	161	174	0	109	132	0.757	0.352
	temperate	METRIC	58	254	-72	88	107	0.422	0.747
		TSEB-PT	63	257	16	61	84	0.325	0.866
		ESVEP	58	246	-47	85	111	0.452	0.712

Table 2.5: Main input parameters for the ET models.

Parameter	Source	TSEB-PT	ESVEP	METRIC
Land surface temperature	S3	X	X	X
Air temperature	ECMWF	X	X	X
Wind speed	ECMWF	X	X	X
Incoming solar radiation	ECMWF	X	X	X
Leaf Area Index	S2	X	X	X
Albedo	S2	X	X	X
Fraction of vegetation which is green	S2	X		
Canopy height	Look-up	X	X	X
Leaf parameters (orientation, width, etc.)	Look-up	X	X	

TSEB-PT requires one additional input: fraction of vegetation which is green. It is particularly important in woody vegetation or in herbaceous vegetation during senescence [43]. Although this parameter can be determined from S2 observations, the method for deriving it is still not fully established and therefore it is not yet of operational quality.

**Ranking ET models:** 1 - METRIC, 2 - ESVEP, 3 - TSEB-PT

**Ranking sharpening approaches:** 1 - NN & DT, 2 -

#### 2.2.4 Functionality

As discussed in the section above, the models vary in the number and types of inputs they need. This is due to their varying complexity, with TSEB-PT being the most complex and METRIC the simplest. However, this does not directly translate to their levels of user friendliness.

When discussing accuracy and applicability, we used results obtained with default model settings and inputs. In this comparison TSEB-PT came out on top and therefore it is viable for the users to run it without fine-tuning any of the settings. To improve the results the users could change the values in the land-cover based look-up table using their knowledge of their area of interest. The parameters in the look-up table, such as vegetation height or leaf angle orientation, are easy to understand for non-EO experts, such as agronomists. Since ESVEP relies on a similar look-up table, potentially it could also produce more accurate results with location specific modifications.

The simplest way for non-EO expert to modify METRIC outputs is to change the AOI over which the search for hot and cold pixels is conducted. This is a more subjective choice compared to look-up table modification, and therefore might require some trial and error before the user gains enough experience to improve the model results.

Integrating any of the models into SNAP should present similar amount of effort. The GUI for TSEB-PT and ESVEP might look a bit more complex than that of METRIC due to larger number of inputs but this should not be a critical issue.

Since TSEB-PT and ESVEP are two-source models, they produce plant transpiration and soil evaporation as two separate outputs. During our discussion with the end-users, they indicated that such separation of ET is of high interest to them. It should be noted however, that the two components of ET have been rarely validated separately (due to difficulty in measuring them separately in the field) and therefore their accuracy is still unknown.

The DT and NN sharpening approaches have very similar user interface and could be integrated in SNAP in a very similar fashion. However, it might be simpler for an average user to understand and fine tune DT model parameters (e.g. number of leaf nodes) than to do the same for NN (e.g. number of hidden layers). In addition, the inputs to NN models have to be normalised before being used. This is done automatically in the script but still might place some constraints on the types of input data (e.g. land-cover map) the user might want to use during the sharpening. DT, on the other hand, can handle heterogeneous inputs more easily.

**Ranking ET models:** 1 - TSEB-PT & ESVEP, 2 - METRIC, 3 -

**Ranking sharpening approaches:** 1 - DT, 2 - NN

### 2.2.5 Performance

The computational performance of the three ET models was benchmark by timing their execution over five, mostly cloud-free Sentinel-2 images (100 km by 100 km each). This resulted in running the models for over 125 000 000 pixels. The test was run on a cloud-based Ubuntu machine with 8 virtual cores and 64 GB of RAM and data located on a virtual disk. The results are presented in Table 2.6. It should be noted that the tests were performed on prototype code and in the final implementation the performance is expected to significantly improve.

Table 2.6: Computational performance of the ET models as tested over five Sentinel-2 images on a cloud-based Ubuntu machine with 8 virtual cores and 64 GB of RAM and data located on a virtual disk.

	TSEB-PT	ESVEP	METRIC
Average execution time per S2 image (minutes)	17	12	6
Maximum RAM usage per S2 image (GB)	34	29	17

TSEB-PT is the most complex of the evaluated models, requiring the largest number of model equations and parameters. In addition it has two nested iterative convergence loops. This complexity is reflected in the longest execution time and highest usage of RAM out of all three models.

ESVEP uses many of the same equations as TSEB-PT but through certain approximations and assumptions it avoids some of parameters and one of the convergence loops. This leads to ESVEP being less computationally demanding compared to TSEB-PT.

Both TSEB-PT and ESVEP are based on the two-source approach, in which the fluxes from the canopy and soil are calculated and treated separately. METRIC, on the other hand, is a one-source model in which bulk canopy and soil fluxes are estimated. Therefore, it has to perform the least number of calculations and after the hot and cold pixels are selected it is very quick in deriving the flux values. This makes it the most performant of the three models.

When running the models over a large (country scale) area it is often required to use cloud computing resources and to employ coding techniques such as data chunking or parallelisation to achieve efficient execution. TSEB-PT and ESVEP are executed purely on per-pixel basis, which makes it easier to implement them using those techniques. METRIC cold and hot pixel selection needs to be performed on a specified AOI and this might somewhat complicate efficient code implementation. The AOI selection requires user input, which again makes METRIC less suitable for fully automatic execution, compared to TSEB-PT and ESVEP. When processing a time-series of images, all three models consider images one time-step at a time so if one image cannot be processed it does not have any impact on subsequent execution.

The performance of the sharpening methods was benchmarked on the same dataset and in the same environment as the ET models and the results are shown in Table 2.7. DT method is much more memory efficient and therefore it was possible to execute the workload on six processors in parallel, leading to short execution time. For NN approach, only two processors could be used in parallel before the machine reached its limit of 64 GB RAM. This contributed to the longer execution time of this approach.

Table 2.7: Computational performance of the LST sharpening methods (NN - neural networks, DT - decision trees) as tested over five Sentinel-2 images on a cloud-based Ubuntu machine with 8 virtual cores and 64 GB of RAM and data located on a virtual disk.

	DT	NN
Average execution time per S2 image (minutes)	5.5	21
Maximum RAM usage per S2 image (GB)	26	43

**Ranking ET models:** 1 - METRIC, 2 - ESVEP, 3 - TSEB-PT

**Ranking sharpening approaches:** 1 - DT, 2 - NN

### 2.2.6 Other considerations

Both METRIC and TSEB-PT are established and well researched ET models, while ESVEP is rather new and has not yet been evaluated in many scientific studies. Therefore, it would be harder to build a community around ESVEP compared to the other two models.

An online search has shown that while a number of open-source implementations of METRIC are available for download, no such implementation exists for TSEB-PT apart from the one used in this project. Therefore it might be simpler to focus community contributions to code, documentation and testing on our Python implementation of TSEB-PT than it would be for METRIC implementation.

Regarding the sharpening methods, both should have similar uptake and contribution from the community.

**Ranking ET models:** 1 - TSEB-PT, 2 - METRIC, 3 - ESVEP

**Ranking sharpening approaches:** 1 - NN & DT, 2 -

### 2.2.7 Overall assessment and model selection

Table 2.8 summarizes the output of the ET model benchmarking. TSEB-PT model obtained the most accurate flux estimates when compared to flux tower measurements and consistently outperformed the other models across different vegetation types and climatologies. This is partly due to this model's increased complexity which is reflected in input data requirements and computational performance. However, it is expected that once the prototype model code is refactored and optimised the computational performance will increase significantly. Therefore, the TSEB-PT model is chosen to be implemented as as SNAP plugin in the second phase of the SEN-ET project.

Regarding the LST sharpening approaches, they performed very similarly with respect to accuracy and applicability. They also require exactly the same inputs. The NN network approach might need more care if new input data is to be added, since it requires normalisation, and it is less computationally efficient compared to the DT. Therefore, the DT version of DMS will be the focus of further development in phase 2 of SEN-ET.

Table 2.8: Summary of the evapotranspiration model benchmarking.

Criteria	METRIC	TSEB-PT	ESVEP
Accuracy	3	1	3
Applicability	3	1	3
Data Accessibility	1	3	2
Functionality	2	1	1
Performance	1	3	2
Others	2	1	3

## Chapter 3

# Detailed model implementations

### 3.1 Net radiation

The code used is available at [https://github.com/hectornieto/pyTSEB/net\\_radiation.py](https://github.com/hectornieto/pyTSEB/net_radiation.py) and is based on Campbell and Norman [51] radiative transfer model in which net short-wave radiation is calculated as

$$S_{n,C,\lambda} = (1 - \rho_{C,\lambda})(1 - \tau_{C,\lambda})S_{\lambda}^{\downarrow} \quad (3.1)$$

$$S_{n,S,\lambda} = (1 - \rho_{S,\lambda})\tau_{C,\lambda}S_{\lambda}^{\downarrow} \quad (3.2)$$

where  $S_{\lambda}^{\downarrow}$  ( $\text{W m}^{-2}$ ) is the incoming spectral shortwave radiation,  $\tau_{C,\lambda}$  is spectral transmittance through the canopy,  $\rho_{S,\lambda}$  and  $\rho_{C,\lambda}$  are respectively soil and canopy spectral hemispherical reflectances (i.e. spectral albedo). The subscript  $\lambda$  is due to that shortwave transmission through the canopy depends on the wavelength due to vegetation absorbing a greater portion of the photosynthetically active radiation (PAR,  $\lambda = [400 - 700]$  nm) than near-infrared radiation (NIR,  $\lambda = [700 - 2500]$  nm) wavelengths.  $\tau_{C,\lambda}$  and  $\alpha_{C,\lambda}$  are partitioned into four components each having a view factor for either direct-beam or diffuse irradiance ( $W_{d,\lambda}$ ) and the ratio of PAR to total shortwave irradiance. The canopy transmittance and reflectance for a horizontally homogeneous canopy are defined as

$$\tau_{C,\lambda} = (1 - W_{d,\lambda})\tau_{C,b,\lambda} + W_d\tau_{C,d,\lambda} \quad (3.3)$$

$$\rho_{C,\lambda} = (1 - W_{d,\lambda})\rho_{C,b,\lambda} + W_d\rho_{C,d,\lambda} \quad (3.4)$$

where  $\tau_{C,b,\lambda}$  and  $\tau_{C,d,\lambda}$  the spectral transmittance for beam direct and diffuse, irradiance respectively, with

$$\tau_{C,b,\lambda} = \frac{(\rho_{C,\lambda}^* - 1) \exp(-\sqrt{\alpha_{leaf,\lambda}\kappa_{b,\theta}}\Omega_{\theta}PAI)}{(\rho_{C,\lambda}^*\rho_{S,\lambda} - 1) + \rho_{C,\lambda}^*(\rho_{C,\lambda}^* - \rho_{S,\lambda}) \exp(-2\sqrt{\alpha_{leaf,\lambda}\kappa_{b,\theta}}\Omega_{\theta}PAI)} \quad (3.5)$$

$$\rho_{C,b,\lambda} = \frac{\rho_{C,\lambda}^* + \frac{\rho_{C,\lambda}^* - \rho_{S,\lambda}}{\rho_{C,\lambda}^* \rho_{S,\lambda} - 1} \exp(-\sqrt{\alpha_{leaf,\lambda}\kappa_{b,\theta}}\Omega_{\theta}PAI)}{1 + \rho_{C,\lambda}^* + \frac{\rho_{C,\lambda}^* - \rho_{S,\lambda}}{\rho_{C,\lambda}^* \rho_{S,\lambda} - 1} \exp(-\sqrt{\alpha_{leaf,\lambda}\kappa_{b,\theta}}\Omega_{\theta}PAI)} \quad (3.6)$$

where  $\rho_{C,\lambda}^*$  is the beam spectral reflectance factor for a deep canopy with non-horizontal leaves,  $\alpha_{leaf,\lambda}$  is the leaf spectral absorption,  $\Omega_{\theta}PAI$  is the effective plant area index ( $\text{m}^2$ )

m<sup>2</sup>), and  $\rho_{S,\lambda}$  is the soil bihemispherical spectral reflectance. Diffuse canopy spectral transmittance ( $\tau_{C,d,\lambda}$ ) and reflectance ( $\rho_{C,d,\lambda}$ ) are calculated by numerically integrating Eq 3.5 over the hemisphere. The extinction coefficient for direct-beam ( $\kappa_{b,\theta}$ ) is calculated based on Campbell [52] ellipsoidal leaf angle distribution function:

$$\kappa_{b,\theta} = \frac{\sqrt{\chi^2 + \tan^2 \theta}}{\chi + 1.774(\chi + 1.182)^{-0.733}} \quad (3.7)$$

where  $\chi$  is the ratio of horizontal to vertical projected unit of area of leaves ( $\chi = 1$  for a spherical leaf angle distribution), and  $\theta$  is the incidence zenith angle

The increased downwelling radiation that is reflected by the soil and then scattered by the canopy back down to the ground surface is accounted for in the  $\rho_{C,\lambda}^*$  and  $\rho_{S,\lambda}$  terms.

$$\rho_{C,\theta,\lambda}^* = \frac{2\kappa_{b,\theta}\rho_{h,\lambda}}{1 + \kappa_{b,\theta}} \quad (3.8)$$

$$\rho_{h,\lambda} = \frac{1 - \sqrt{\alpha_{leaf,\lambda}}}{1 + \sqrt{\alpha_{leaf,\lambda}}} \quad (3.9)$$

leaf absorption is computed as  $\zeta_\lambda = 1 - \rho_{leaf,\lambda} - \tau_{leaf,\lambda}$ , with  $\rho_{leaf,\lambda}$  and  $\tau_{leaf,\lambda}$  with being the leaf bihemispherical spectral reflectance and transmittance, respectively.

$\Omega$  is the vegetation clumping factor proposed by Kustas and Norman [50] for sparsely vegetated areas, which can be set to 1 for homogeneous canopies. Otherwise the nadir viewing clumping index ( $\Omega_0$ ) is defined as the factor that modifies the leaf area index of a clumped canopy ( $F = PAI/f_{cover}$ ) in a homogeneous canopy with  $LAI_{eff} = \Omega_0 F$  such as its at-nadir gap fraction  $G_0$  is the same as the gap fraction of the canopy.

$$\Omega_0 = \frac{-\log(G_0)}{\kappa_{b,0} F} \quad (3.10)$$

Then, the off-nadir clumping index is calculated using the empirical formula for randomly placed canopies proposed by Kustas and Norman [50]

$$\Omega_\theta = \frac{\Omega_0}{\Omega_0 + (1 - \Omega_0) \exp(-2.2\theta^p)} \quad (3.11)$$

$$p = 3.8 - 0.46 \frac{h_c}{w_c} \quad ; 1 \leq \frac{h_c}{w_c} \leq 3.34 \quad (3.12)$$

where  $h_c$  and  $w_c$  are canopy height and width respectively.

On the other hand, net longwave radiation is computed as

$$L_{n,C} = [1 - \exp(-\kappa_l PAI)] \left[ \epsilon_C (L^\downarrow + L_S) - 2L_C \right] \quad (3.13)$$

$$L_{n,S} = \epsilon_S \exp(-\kappa_l PAI) L^\downarrow + \epsilon_S [1 - \exp(-\kappa_l PAI)] L_C - L_S \quad (3.14)$$

with  $\kappa_l \approx 0.95$  is the long-wave radiation extinction coefficient, and  $L_S$ ,  $L_C$  and  $L^\downarrow$  (W m<sup>-2</sup>) are the long-wave emissions from soil, canopy and sky. The Stefan-Boltzman equation based on soil, canopy and air temperatures can be used to compute  $L_S$ ,  $L_C$  and  $L^\downarrow$ .

$$L_S = \epsilon_S \sigma T_S^4 \quad (3.15)$$

$$L_C = \epsilon_C \sigma T_C^4 \quad (3.16)$$

$$L^\downarrow = \epsilon_{atm} \sigma T_{air}^4 \quad (3.17)$$

where  $\epsilon_S$ ,  $\epsilon_C$ , and  $\epsilon_{atm}$  are emissivity for soil, canopy and air respectively, and  $\sigma$  is the Stefan-Boltzman constant.

However, Eq. 3.13 requires the estimation of both canopy and soil temperatures and hence it is only valid for TSEB models. Therefore, for METRIC, like in other OSEB models, net longwave radiation calculation is simplified to

$$L_n = \epsilon L^\downarrow - \epsilon \sigma T_{rad}^4 \quad (3.18)$$

with  $\epsilon$  being the surface emissivity ( $\epsilon = f_c \epsilon_C + (1 - f_c) \epsilon_S$ ).

## 3.2 METRIC

The code used is available at <https://github.com/hectornieto/pyMETRIC>

Sensible heat flux is computed in METRIC as

$$H = \rho_a C_p \frac{\delta T}{R_{AH}} \quad (3.19)$$

with  $\rho_a C_p$  is the volumetric heat capacity of air ( $\text{J m}^{-3} \text{K}^{-1}$ ),  $R_{AH}$  is the aerodynamic resistance to heat transport ( $\text{s m}^{-1}$ ), and  $\Delta T$  is computed as

$$\delta T = c + m T_{rad} \quad (3.20)$$

Linear coefficients  $c$  and  $m$  are computed from a set of two linear equations function of endmember of hot and cold pixels

$$m = \frac{\delta T_{hot} - \delta T_{cold}}{T_{hot} - T_{cold}} \quad (3.21)$$

$$c = \delta T_{hot} - m T_{hot} \quad (3.22)$$

with  $\Delta T_{hot}$  and  $\Delta T_{cold}$  computed as

$$\delta T_{cold} = (R_{n,cold} - G_{cold} - \lambda E_{cold}) \frac{R_{AH,cold}}{\rho_{a,cold} C_p} \quad (3.23)$$

$$\delta T_{hot} = (R_{n,hot} - G_{hot} - \lambda E_{hot}) \frac{R_{AH,hot}}{\rho_{a,hot} C_p} \quad (3.24)$$

where  $\lambda E_{hot} \approx 0$  and  $\lambda E_{cold}$  is an empirical function of the ASCE's [56] 0.5 m tall alfalfa reference evapotranspiration ( $\lambda E_{cold} = 1.05 f_c \lambda E_{ASCE}$ )

Net radiation calculation follows the procedure described in 3.1, whereas soil heat flux is computed as in [31]

$$G = R_n [(T_{rad} - 273.15) * (0.0038 + 0.0074 * \alpha) * (1 - 0.98 * f_c^4)] \quad (3.25)$$

where  $\alpha$  is the surface albedo ( $\alpha = 1 - S_n/S^\downarrow$ ), and  $f_c$  is the green fractional cover ( $f_c = 1 - \exp(-0.5 LAI)$ )

Finally, the aerodynamic resistance to heat transport is calculated based on Allen et al. [32]:

$$R_{AH} = \frac{\log(z_2/z_1) - \Psi_H(z_2) + \Psi_H(z_1)}{\kappa u_*} \quad (3.26)$$

where  $z_2$  is at 2m,  $z_1 = 0.1$  m,  $\Psi_H(\zeta)$  are adiabatic correction functions for heat transport,  $u_*$  is friction velocity ( $\text{m s}^{-1}$ ) and  $\kappa = 0.41$  is the von Kàrman constant.

### 3.2.1 Automatic search of cold and hot pixels

We adopted the Exhaustive Search Algorithm described in Bhattarai et al. [57], which is based on dynamic thresholds of a vegetation index ( $f_{cover}$  in our case), surface temperature ( $T_{rad}$ ), and surface albedo ( $\alpha$ ).

In a first step we filter all candidate pixels -excluding those pixels classified as water bodies, urban, snow/ice and shadow/cloud- based on their homogeneity compared to neighbouring pixels. Homogeneity calculations are based on a 11x11 pixels (i.e. 220x220 m) moving window, where candidate pixels are those which the coefficients of variation for  $f_{cover}$  and  $\alpha$  are lower than 0.25, and  $T_{rad}$  standard deviation is lower than 1.5 K.

Then histograms are built from  $T_{rad}$  and  $f_{cover}$  values in order to remove possible outliers. Histogram bins are created from 0.25 K size in  $T_{rad}$  0.01 in  $f_{cover}$ , while ensuring that the minimum and maximum bins contain at least 50 pixels. All pixels within the first and last bin for each variable ( $T_{rad}$  and  $f_{cover}$ ) are excluded as outliers.

Now the method looks for the coldest pixels with the highest fractional cover, but instead of searching for a single pixel, the method aims at searching as set of pixels in an iterative approach. It starts selecting those pixels that are within the 1% highest  $f_{cover}$  (i.e. percentile 99%) and 1% lowest  $T_{rad}$ , decreasing the  $f_{cover}$  percentile threshold to 1 point up to a maximum threshold decrease of 10% (i.e. percentile 90%), until at least 10 pixels are found. If not enough pixels are found, then the  $T_{rad}$  percentile threshold is increased 1 point (i.e. 2%) and the search is started again, until a maximum  $T_{rad}$  percentile threshold of 10%. In case that after all these iterations (i.e. when the number of pixels on the 10% highest  $f_{cover}$  and 10% lowest  $T_{rad}$  is less than 10) the maximum allowed percentile thresholds are relaxed and increased to 5 points (i.e. minimum  $f_{cover}$  percentile threshold during the iterative search is now 85%). This step is repeated until at least 10 pixels are found.

A similar approach is done for searching the hot pixels, which iterative search starts from those pixels with lowest vegetation cover (1% percentile) and highest temperature (99% percentile). Finally, the hot and cold pixels used in METRIC are the ones with the highest  $T_{rad} - f_{cover}$  and ( $f_{cover} - T_{rad}$ ) value, respectively.

## 3.3 TSEB

The code used is available at <https://github.com/hectornieto/pyTSEB> The basic equation of the energy balance at the surface can be expressed following Eq. 3.27.

$$R_n \approx H + \lambda E + G \quad (3.27a)$$

$$R_{n,S} \approx H_S + \lambda E_S + G \quad (3.27b)$$

$$R_{n,C} \approx H_C + \lambda E_C \quad (3.27c)$$

with  $R_n$  being the net radiation,  $H$  the sensible heat flux,  $\lambda E$  the latent heat flux or ET, and  $G$  the soil heat flux. “C” and “S” subscripts refer to canopy and soil layers respectively. The symbol “ $\approx$ ” appears since there are additional components of the energy balance that are usually neglected, such as heat advection, storage of energy in the canopy layer or energy for the fixation of CO<sub>2</sub> [83]

The key in TSEB models is the partition of sensible heat flux into the canopy and soil layers, which depends on the soil and canopy temperatures ( $T_S$  and  $T_C$  respectively). If we assume that there is an interaction between the fluxes of canopy and soil, due to an expected heating of the in-canopy air by heat transport coming from the soil, the resistances network in TSEB can be considered to be in series. In that case  $H$  can be estimated as in Eq. 3.28 [36, Eqs. A1-A3]

$$\begin{aligned} H = H_C + H_S &= \rho_{air} C_p \frac{T_{AC} - T_A}{R_a} \\ &= \rho_{air} C_p \left[ \frac{T_C - T_{AC}}{R_x} + \frac{T_S - T_{AC}}{R_s} \right] \end{aligned} \quad (3.28)$$

where  $\rho_{air}$  is the density of air ( $\text{kg m}^{-3}$ ),  $C_p$  is the heat capacity of air at constant pressure ( $\text{J kg}^{-1} \text{K}^{-1}$ ),  $T_{AC}$  is the air temperature at the canopy interface, equivalent to the aerodynamic temperature  $T_0$ , computed with Eq. 3.29 [36, Eq. 4].

$$T_{AC} = \frac{\frac{T_A}{R_a} + \frac{T_C}{R_x} + \frac{T_S}{R_s}}{\frac{1}{R_a} + \frac{1}{R_x} + \frac{1}{R_s}} \quad (3.29)$$

Here  $R_a$  is the aerodynamic resistance to heat transport ( $\text{s m}^{-1}$ ),  $R_s$  is the resistance to heat flow in the boundary layer immediately above the soil surface ( $\text{s m}^{-1}$ ), and  $R_x$  is the boundary layer resistance of the canopy of leaves ( $\text{s m}^{-1}$ ). The mathematical expressions of these resistances are detailed in Eq. 3.30 and in Norman et al. [36] and Kustas and Norman [53] and discussed in Kustas et al. [28].

$$R_a = \frac{\ln\left(\frac{z_T - d_0}{z_{0H}}\right) - \Psi_h\left(\frac{z_T - d_0}{L}\right) + \Psi_h\left(\frac{z_{0H}}{L}\right)}{\kappa' u_*} \quad (3.30a)$$

$$R_s = \frac{1}{c(T_S - T_A)^{1/3} + b u_s} \quad (3.30b)$$

$$R_x = \frac{C'}{\text{LAI}} \left( \frac{l_w}{U_{d_0 + z_{0M}}} \right)^{1/2} \quad (3.30c)$$

where  $u_*$  is the friction velocity ( $\text{m s}^{-1}$ ) computed as:

$$u_* = \frac{\kappa' u}{\left[ \ln\left(\frac{z_u - d_0}{z_{0M}}\right) - \Psi_m\left(\frac{z_u - d_0}{L}\right) + \Psi_m\left(\frac{z_{0M}}{L}\right) \right]} \quad (3.31)$$

In Eq. 3.31  $z_u$  and  $z_T$  are the measurement heights for wind speed  $u$  ( $\text{m s}^{-1}$ ) and air temperature  $T_A$  (K), respectively.  $d_0$  is the zero-plane displacement height,  $z_{0M}$  and  $z_{0H}$  are the roughness length for momentum and heat transport respectively (all those magnitudes expressed in m), with  $z_{0H} = z_{0M} \exp(-kB^{-1})$ . In the series version of TSEB  $z_{0H}$  is assumed equal to  $z_{0M}$  since the term  $R_x$  already accounts for the different efficiency between heat and momentum transport [36], and therefore  $kB^{-1} = 0$ . The value of  $\kappa' = 0.4$  is the von Karman's constant. The  $\Psi_m(\zeta)$  terms in Eqs. 3.30a and 3.31 are the adiabatic correction factors for momentum. The formulations of these two factors are described in Brutsaert [26] and Brutsaert [25]. These corrections depend on the atmospheric stability, which is expressed using the Monin-Obukhov length  $L$  (m):

$$L = \frac{-u_*^3 \rho_{air}}{k g [H / (T_A C_p) + 0.61 E]} \quad (3.32)$$

where  $H$  is the bulk sensible heat flux ( $\text{W m}^{-2}$ ),  $E$  is the rate of surface evaporation ( $\text{kg s}^{-1}$ ), and  $g$  the acceleration of gravity ( $\text{m s}^{-2}$ )

The coefficients  $b$  and  $c$  in Eq 3.30b depend on turbulent length scale in the canopy, soil-surface roughness and turbulence intensity in the canopy, which are discussed in Sauer et al. [84], Kondo and Ishida [85] and Kustas et al. [28].  $C'$  is assumed to be  $90 \text{ s}^{1/2} \text{ m}^{-1}$  and  $l_w$  is the average leaf width (m)

Wind speed at the heat source-sink ( $z_{0M} + d_0$ ) and near the soil surface was originally estimated using Goudriaan [86] wind attenuation model (Eq 3.33)

$$U(z) = U_C \exp[-a_G(1 - z/h_c)] \quad (3.33)$$

$$a_G = 0.28 \text{LAI}^{2/3} h_c^{1/3} l_w^{-1/3} \quad (3.34)$$

Since Eqs. 3.28-3.32 are interrelated, an iterative scheme is performed until the convergence of  $L$  and  $u_*$  is reached. The iterative process is as follows: neutral conditions are firstly assumed ( $L \rightarrow \infty$ ,  $\Psi_M(\zeta) = 0$  and  $\Psi_H(\zeta) = 0$ ) and an initial estimate of  $H$  is calculated using Eqs. 3.31 to 3.28, and  $E$  with Eq. 3.27. An initial value of  $L$  is then obtained from Eq. 3.32 and the stability functions are then calculated, which gives a new friction velocity (Eq. 3.31) and resistance set (Eq. 3.30) and new estimates of  $H$  and  $E$  (Eqs. 3.29, 3.28 and 3.27).  $L$  is recalculated again and the process continues (Eqs. 3.32-3.28) until the change in  $L$  and  $u_*$  between two successive iterations is lower than a certain threshold.

When only a single observation of  $T_{rad}$  is available (i.e. measurement at a single angle), partitioning of  $T_{rad}$  requires some assumptions to help to define  $T_C$  or  $T_S$ . One approach developed for TSEB [36] starts with an initial estimate that assumes plants are transpiring at a potential rate, as defined by the Priestley and Taylor [87] relationship, applied to the canopy divergence of net radiation ( $R_{n,C}$ )

$$\lambda E_C = \alpha_{PT} f_g \frac{\Delta}{\Delta + \gamma} R_{n,C} \quad (3.35)$$

where  $\alpha_{PT}$  is the Priestley-Taylor coefficient, initially set to 1.26,  $f_g$  is the fraction of vegetation that is green and hence capable of transpiring,  $\Delta$  is the slope of the saturation vapour pressure versus temperature curve, and  $\gamma$  is the psychrometric constant. This allows the canopy sensible heat flux to be calculated using the energy-balance at the canopy layer ( $H_c = R_{n,C} - \lambda E_C$ ) and hence an estimate of  $T_C$  to be obtained by inverting Eq. 3.28 [36, Eqs. A7, A11 and A12]. Then  $T_S$  is derived from Eq. 3.36 having both  $T_{rad}$  and  $T_C$  and an estimate of  $f_c(\theta)$  the fraction of vegetation observed by the sensor view zenith angle  $\theta$ .

$$T_{rad}^4(\theta) = f_c(\theta) T_C^4 + [1 - f_c(\theta)] T_S^4 \quad (3.36)$$

The value of  $f_c(\theta)$  is typically estimated as an exponential function of the leaf area index, which includes a clumping factor or index  $\Omega$  where the LAI is concentrated for plants sparsely [50, 82] and has the following form.

$$f_c(\theta) = 1 - \exp\left(\frac{-0.5\Omega LAI}{\cos\theta}\right) \quad (3.37)$$

If the initial  $T_C$  implied by this approximation is unusually low in comparison with the observed  $T_{rad}$ ,  $T_S$  will likely be overestimated and therefore produce unrealistic estimates of soil latent heat flux (negative values during daytime). In this case, the  $\alpha_{PT}$  coefficient is iteratively reduced assuming the canopy is stressed and transpiring at sub-potential levels until soil latent heat flux becomes zero or positive.

### 3.4 ESVEP

The code used is available at <https://github.com/DHI-GRAS/py-esvep>.

All the equations in Section 3.3 are also applied to ESVEP.

### 3.5 sharpening models

The code used is available at <https://github.com/radosuav/pyDMS>



## Chapter 4

# Data preprocessing details

The code used is available at <https://github.com/DHI-GRAS/sen-et-input-scripts/> upon access granted by GRAS-DHI. The atmospheric correction of the MSI observations was performed using the Sen2Cor processor [65] v2.5.5 in order to obtain bottom-of-atmosphere (BOA) reflectance values. The Sen2Cor processor was configured to use rural aerosol type, with both atmospheric profile (summer or winter) and ozone concentration chosen automatically from the scene metadata. Aerosol optical thickness and total column water vapour content are determined from the MSI observations themselves using all possible spectral bands. The output of the processor was set at 20 m spatial resolution, and included 9 spectral bands (2-7, 8A, 11 and 12) as well as land cover classification and a cloud mask. Due to the inaccuracies present in the cloud mask, it was first cleaned up using a 5x5 pixel majority filter, before being dilated using 20x20 pixel filter. The BOA reflectance values were then used as input to the Biophysical Processor [66] available in the SNAP software v6.0.1 (step.esa.int - last accessed 28.11.2018) in order to obtain green Leaf Area Index (LAI), Fractional Vegetation Cover (FVC), Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), Canopy Chlorophyll Content (CCC) and Canopy Water Content (CWC).

The thermal data needed to drive the evapotranspiration model was obtained from the Sea and Land Surface Temperature Radiometer (SLSTR) on board of the Sentinel-3A satellite [69]. At this moment there was not total availability of SLSTR L2 LST product, and therefore we used the SLSTR L1 radiance product to convert two SLSTR brightness temperature (BT) bands (S8 - centred on 10.85  $\mu\text{m}$  and S9 - centred on 12  $\mu\text{m}$ ) were used to retrieve  $T_{rad}$  using a split-window algorithm described by [70]. Apart from the BTs the algorithm requires Total Column Water Vapour (TCWV) and land surface emissivity estimates. TCWV was obtained from European Centre for Medium-range Weather Forecast (ECMWF) reanalysis model. The surface emissivity was estimated by mixing the emissivity of the soil and vegetation using FVC [70]. Soil emissivity was assumed to take constant values in the two SLSTR thermal bands, as described in [70], although this assumption does not take soil moisture into account. Vegetation emissivity was set to 0.99 for green vegetation and to 0.91 for dry vegetation [88] and combined using the  $f_g$  parameter. The surface emissivity was estimated at the S2 spatial resolution before being averaged to the SLSTR spatial resolution. Cloud masking was based on the internal mask provided in the SLSTR data file but similarly to the case of S2 images it was cleaned up with a 5x5 pixel majority filter followed by 5x5 pixel dilation.



## Chapter 5

# Scientific questions to be addressed in Phase 2, and beyond

There are still remaining some scientific questions that could be answered during phase 2 and/or carried out based on future projects and calls.

- What is the effect/sensitivity of considering an explicit model for heat and momentum transfer in complex landscapes such as forest, savannahs and row crops/orchards.

Recent studies have suggested modifications in radiation partitioning and wind attenuation models that explicitly account for horizontal and vertical canopy heterogeneity [89, 90]. Several of the sites in Table 1.1 could be used to evaluate how those sub-models perform and extent the validation of these approaches such as Borden (vineyard), the savannas of Dahra and Majadas and the olive grove in Taous

- What is the sensitivity in two source models of using different resistance formulations of soil and canopy heat transport.

Besides of the formulations for soil and canopy resistance expressed in Eq. 3.30, other alternative calculations could be tested, such as the ones in Choudhury and Monteith [91], McNaughton and Van Den Hurk [92], or Li et al. [93], and see how they perform at different landscape and/or globally.

- Can other forms of TSEB be more robust

The Priestley-Taylor TSEB is a simple approach in terms of computing the initial canopy transpiration. Other methods such as the use of Penman-Monteith potential ET [37] or SPARSE model [38] could be evaluated, especially in sites with some heat advection as these models usually produce more reliable results in very high vapour pressure deficit conditions. On the other hand, multiangular thermal observations by SLSTR could be used to directly retrieve soil and canopy temperatures [94, 95], without the need to estimate any potential transpiration nor green fraction.

- How robust is the total LAI and its green fraction  $f_g$  derived with Sentinel multi-spectral data.

Majadas de Tiétar in one the most ideal sites to evaluate this variable as it is a site with a very intensive *in situ* data collection and the grassland presents a strong

seasonality, with a senescent phase starting around May and a later regreening in autumn.

- How can we upscale instantaneous ET to daily ET.

Many applications, such as in precision agriculture, require estimates of daily ET (usually mm/day) but satellite ET models produce instantaneous ET. Methods to integrate to daily scales are needed to be explored and tested. Usually these methods rely on the assumption of the daily self-preservation of a flux metric such as evaporative fraction ( $EF = \frac{\lambda E}{R_n - G}$ ), in which EF (or other possible metric like  $\frac{\lambda E}{S^{\downarrow}}$  or  $\frac{\lambda E}{\lambda E_0}$ , is usually assumed constant during daytime and hence  $\lambda E_{daily} = EF_{instantaneous} (R_n - G)_{daily}$

- What is the most suitable ET index that could be used to derive crop water stress.

For irrigation practices not only actual ET is needed but also we need to know the maximum crop water needs (i.e. the crop potential ET). Furthermore, for regulated deficit irrigation practices, it is crucial to assess the crop water stress and keep it under certain stress thresholds.

- Can the disaggregated ET product detect spatial differences of annual water use/water consumption due to different irrigation/fertilization treatments

Considering the uncertainties in sharpened temperatures, can accumulated metrics, in which some random errors might be cancelled out, detect spatial or inter-annual variability in water use, irrigation or fertilization practices? For instance several sites described in section 1.2.2 have more than one eddy covariance tower, such as in the case of Borden, with two towers in two adjacent vineyards, or Majadas de Tiétar with three EC systems on which each one received different NPK fertilization rates.

- What is the uncertainty associated of cloud cover conditions in thermal sharpening. Or in other words, what is the minimum number of valid pixels to be used in the machine learning calibration for thermal sharpening?

- What is the best approach to produce gap-filled daily continuous ET estimates

Optical and thermal satellite remote sensing has the disadvantage that is affected by clouds. Therefore spatiotemporal gaps will be present that needs to be filled somehow. Several alternatives could be evaluated, from simple empirical interpolation and extrapolation gap filling methods, to linking the satellite information into land surface model via data assimilation. Being the latest one more sound as it does not only fill the gaps but it is also able to forecast future crop status/water availability.

## Chapter 6

# Suggestions for future ESA thermal mission candidate

While the best ET models performed well based on the available input data (Sentinel 2 and Sentinel 3), neither of the models reached an accuracy of 20% which is desirable for ET assessment for agricultural applications. The current S3 input  $T_{rad}$  data were not able to resolve the resolution of an agricultural field or the foot-print of the flux tower used for validation, so sharpening of the data is necessary. This will have negative consequences on the accuracy of the modelled ET, in particular in heterogeneous landscapes, however effect is difficult to quantify, but the downscaling method seem to influence the result. Secondly, the time of the passage of Sentinel 3 during the morning, before the surface heating has reached its maximum, leads to less accurate ET measurements.

Based on the prototype implementation of the four ET models, the need for accessibility of high quality ancillary data for operational processing of ET is highlighted. Land cover maps form the basis for model dependent input parameters to evapotranspiration models, for instance vegetation height, thus a up to date land cover map is essential. Land surface characterization (e.g. albedo, leaf area index) are required at the same spatial resolution as  $T_{rad}$  data, e.g. obtained from Sentinel-2 observations as well as Vegetation height/profile or surface roughness at the similar spatial resolution as  $T_{rad}$  data (e.g. from LiDAR observations) are desirable ancillary data. A list of essential and optional variables is given in the table below.

The overall recommendations to future LST missions can thus be stated:

- High spatial resolution is required to resolve the spatial resolution of agricultural fields or ideally irrigation sectors
- Overpass time needs to be aligned with time of the peak of  $T_{rad}$  and sensible heat flux ( noon). There should be a compromise between maximum fluxes and minimum global cloud coverage.
- Ancillary input data essential for operational ET modeling

Input	Parameter	Source
Land Cover Map		CCI or Sen2Agri
Vegetation state	Leaf Area Index Fraction of green vegetation FAPAR Canopy Chlorophyll Content	Sentinel 2
Radiation	Instantaneous longwave irradiance Instantaneous shortwave irradiance, direct and diffuse	ECMWF
Meteorological forcing	Air temperature 100m Wind speed 100m Vapour pressure/dew temperature Surface pressure Atmospheric water vapour Aerosol optical thickness	ECMWF
DEM	Optional	

# Bibliography

- [1] Héctor Nieto, Radoslaw Guzinski, Inge Sandholt, Mads Olander Rasmussen, Joaquim Bellvert, Christian Jofré, Lars Boye Hansen, Jonas Blüthgen Sølvsteen, Torsten Bondo, and Benjamin Koetz. Sentinels for evapotranspiration. prototype evaluation. Deliverable Project Nr 4000121772/17/I-NB D1.1, European Space Agency, March 2018.
- [2] W. P. Kustas and M. C. Anderson. Advances in thermal infrared remote sensing for land surface modeling. *Agricultural and Forest Meteorology*, 149(12):2071–2081, 2009.
- [3] J.D. Kalma, T.R. McVicar, and M.F. McCabe. Estimating land surface evaporation: A review of methods using remotely sensed surface temperature data. *Surveys in Geophysics*, 29(4):421–469, 2008.
- [4] Radoslaw Guzinski and Héctor Nieto. Evaluating the feasibility of using Sentinel-2 and Sentinel-3 satellites for high-resolution evapotranspiration estimations. *Remote Sensing of Environment*, 221:157–172, February 2019. ISSN 0034-4257. doi: 10.1016/j.rse.2018.11.019.
- [5] Olivier Merlin, Benoit Duchemin, Olivier Hagolle, Frédéric Jacob, Benoit Coudert, Ghani Chehbouni, Gérard Dedieu, Jaime Garatuza, and Yann Kerr. Disaggregation of MODIS surface temperature over an agricultural area using a time series of Formosat-2 images. *Remote Sensing of Environment*, 114(11):2500–2512, 2010.
- [6] Feng Gao, William P. Kustas, and Martha C. Anderson. A Data Mining approach for Sharpening thermal satellite imagery over land. *Remote Sensing*, 4(11):3287–3319, 2012. doi: 10.3390/rs4113287.
- [7] V.M. Bindhu, B. Narasimhan, and K.P. Sudheer. Development and verification of a non-linear disaggregation method (NL-DisTrad) to downscale MODIS land surface temperature to the spatial scale of Landsat thermal data to estimate evapotranspiration. *Remote Sensing of Environment*, 135(Supplement C):118 – 129, 2013. doi: 10.1016/j.rse.2013.03.023.
- [8] JM Norman, MC Anderson, WP Kustas, AN French, John Mecikalski, Ray Torn, GR Diak, TJ Schmugge, and BCW Tanner. Remote sensing of surface energy fluxes at 10<sup>-1</sup>m pixel resolutions. *Water Resources Research*, 39(8):–, 2003.
- [9] S. N. Goward, R. H. Waring, D. G. Dye, and J. L. Yang. Ecological remote-sensing at OTTER: Satellite macroscale observations. *Ecological Applications*, 4(2):322–343, 1994.

- [10] S. D. Prince, S. J. Goetz, R. O. Dubayah, K. P. Czajkowski, and M. Thawley. Inference of surface and air temperature, atmospheric precipitable water and vapor pressure deficit using Advanced Very High-Resolution Radiometer satellite observations: Comparison with field observations. *Journal of Hydrology*, 213(1-4):230–249, 1998.
- [11] Pietro Ceccato, Stéphane Flasse, Stefano Tarantola, Stéphane Jacquemoud, and Jean-Marie Grégoire. Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sensing of Environment*, 77(1):22–33, 2001. doi: 10.1016/S0034-4257(01)00191-2.
- [12] Nurit Agam, William P. Kustas, Martha C. Anderson, Fuqin Li, and Christopher M.U. Neale. A vegetation index based technique for spatial sharpening of thermal imagery. *Remote Sensing of Environment*, 107(4):545 – 558, 2007. doi: 10.1016/j.rse.2006.10.006.
- [13] Ji Zhou, Shaomin Liu, Mingsong Li, Wenfeng Zhan, Ziwei Xu, and Tongren Xu. Quantification of the scale effect in downscaling remotely sensed land surface temperature. *Remote Sensing*, 8(12):975, 2016. doi: 10.3390/rs8120975.
- [14] M. Bisquert, J. M. Sánchez, and V. Caselles. 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, 2016. doi: 10.1109/JSTARS.2016.2519099.
- [15] Christopher Hutengs and Michael Vohland. Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sensing of Environment*, 178:127–141, 2016. doi: 10.1016/j.rse.2016.03.006.
- [16] G. Yang, R. Pu, W. Huang, J. Wang, and C. Zhao. A novel method to estimate subpixel temperature by fusing solar-reflective and thermal-infrared remote-sensing data with an artificial neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 48(4):2170–2178, 2010.
- [17] Xuehong Chen, Wentao Li, Jin Chen, Yuhan Rao, and Yasushi Yamaguchi. A combination of TsHARP and thin plate spline interpolation for spatial sharpening of thermal imagery. *Remote Sensing*, 6(4):2845–2863, 2014. doi: 10.3390/rs6042845.
- [18] Z. Q. Peng, X. Xin, J. J. Jiao, T. Zhou, and Q. Liu. Remote sensing algorithm for surface evapotranspiration considering landscape and statistical effects on mixed pixels. *Hydrology and Earth System Sciences*, 20(11):4409–4438, 2016. doi: 10.5194/hess-20-4409-2016.
- [19] Yinghai Ke, Jungho Im, Seonyoung Park, and Huili Gong. Downscaling of MODIS one kilometer evapotranspiration using Landsat-8 data and machine learning approaches. *Remote Sensing*, 8(3), 2016. doi: 10.3390/rs8030215.
- [20] M. Bisquert, J. M. Sánchez, R. López-Urrea, and V. Caselles. Estimating high resolution evapotranspiration from disaggregated thermal images. *Remote Sensing of Environment*, 187:423–433, 2016. doi: 10.1016/j.rse.2016.10.049.

- [21] Shen Tan, Bingfang Wu, Nana Yan, and Weiwei Zhu. An NDVI-based statistical ET downscaling method. *Water*, 9(12):995, 2017. doi: 10.3390/w9120995.
- [22] Wenfeng Zhan, Yunhao Chen, Ji Zhou, Jinfei Wang, Wenyu Liu, James Voogt, Xiaolin Zhu, Jinling Quan, and Jing Li. Disaggregation of remotely sensed land surface temperature: Literature survey, taxonomy, issues, and caveats. *Remote Sensing of Environment*, 131:119 – 139, 2013. doi: 10.1016/j.rse.2012.12.014.
- [23] R. Guzinski, H. Nieto, R. Jensen, and G. Mendiguren. Remotely sensed land-surface energy fluxes at sub-field scale in heterogeneous agricultural landscape and coniferous plantation. *Biogeosciences*, 11(18):5021–5046, 2014.
- [24] Kathryn A. Semmens, Martha C. Anderson, William P. Kustas, Feng Gao, Joseph G. Alfieri, Lynn McKee, John H. Prueger, Christopher R. Hain, Carmelo Cammalleri, Yun Yang, Ting Xia, Luis Sanchez, M. Mar Alsina, and Mónica Vélez. Monitoring daily evapotranspiration over two California vineyards using Landsat 8 in a multi-sensor data fusion approach. *Remote Sensing of Environment*, 185:155–170, 2016. doi: 10.1016/j.rse.2015.10.025.
- [25] Wilfried Brutsaert. *Hydrology. An Introduction*. Cambridge University Press, 2005. ISBN 13 978-0-521-82479-8.
- [26] Wilfried Brutsaert. Aspects of bulk atmospheric boundary layer similarity under free-convective conditions. *Reviews of Geophysics*, 37:439–451, 1999.
- [27] J-P Lhomme, Bruno Monteny, and Mustapha Amadou. Estimating sensible heat flux from radiometric temperature over sparse millet. *Agricultural and Forest Meteorology*, 68(1-2):77–91, 1994.
- [28] W. P. Kustas, H. Nieto, L. Morillas, M. C. Anderson, J. G. Alfieri, L. E. Hipps, L. Villagarcía, F. Domingo, and M. García. Revisiting the paper “Using radiometric surface temperature for surface energy flux estimation in Mediterranean drylands from a two-source perspective”. *Remote Sensing of Environment*, 184:645–653, 2016.
- [29] JR Garratt and BB Hicks. Momentum, heat and water vapour transfer to and from natural and artificial surfaces. *Quarterly Journal of the Royal Meteorological Society*, 99(422):680–687, 1973.
- [30] A Verhoef, HAR De Bruin, and BJJM Van Den Hurk. Some practical notes on the parameter  $kB^{-1}$  for sparse vegetation. *Journal of Applied Meteorology*, 36(5):560–572, 1997.
- [31] WGM Bastiaanssen, M Menenti, RA Feddes, and AAM Holtslag. A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. *Journal of Hydrology*, 212:198–212, 1998.
- [32] R. Allen, M. Tasumi, and R. Trezza. Satellite-Based Energy Balance for Mapping Evapotranspiration with Internalized Calibration (METRIC)\97model. *Journal of Irrigation and Drainage Engineering*, 133(4):380–394, 2007. doi: 10.1061/(ASCE)0733-9437(2007)133:4(380).

- [33] Z Su. The Surface Energy Balance system (SEBS) for estimation of turbulent heat fluxes. *Hydrology and Earth System Sciences Discussions*, 6(1):85–100, 2002.
- [34] William P. Kustas, Bhaskar J. Choudhury, Kenneth E. Kunkel, and Lloyd W. Gay. Estimate of the aerodynamic roughness parameters over an incomplete canopy cover of cotton. *Agricultural and Forest Meteorology*, 46(1-2):91 – 105, 1989.
- [35] Ronglin Tang, Zhao-Liang Li, Yuanyuan Jia, Chuanrong Li, Xiaomin Sun, William P. Kustas, and Martha C. Anderson. An intercomparison of three remote-sensing-based energy balance models using large aperture scintillometer measurements over a wheat-corn production region. *Remote Sensing of Environment*, 115(12):3187–3202, 2011. doi: 10.1016/j.rse.2011.07.004.
- [36] J. M. Norman, W. P. Kustas, and K. S. Humes. Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature. *Agricultural and Forest Meteorology*, 77(3-4):263–293, 1995.
- [37] Paul D. Colaizzi, William P. Kustas, Martha C. Anderson, Nurit Agam, Judy A. Tolck, Steven R. Evett, Terry A. Howell, Prasanna H. Gowda, and Susan A. O’Shaughnessy. Two-source energy balance model estimates of evapotranspiration using component and composite surface temperatures. *Advances in Water Resources*, 50:134 – 151, 2012. doi: 10.1016/j.advwatres.2012.06.004.
- [38] G. Boulet, B. Mougenot, J.-P. L’homme, P. Fanise, Z. Lili-Chabaane, A. Olioso, M. Bahir, V. Rivalland, L. Jarlan, O. Merlin, B. Coudert, S. Er-Raki, and J.-P. Lagouarde. The SPARSE model for the prediction of water stress and evapotranspiration components from thermal infra-red data and its evaluation over irrigated and rainfed wheat. *Hydrology and Earth System Sciences*, 19(11):4653–4672, 2015. doi: 10.5194/hess-19-4653-2015.
- [39] G.J Roerink, Z Su, and M Menenti. S-SEBI: A simple remote sensing algorithm to estimate the surface energy balance. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 25(2):147–157, 2000. doi: 10.1016/S1464-1909(99)00128-8.
- [40] Inge Sandholt, Kjeld Rasmussen, and Jens Andersen. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sensing of Environment*, 79(2):213–224, 2002.
- [41] Olivier Merlin, Jonas Chirouze, Albert Olioso, Lionel Jarlan, Ghani Chehbouni, and Gilles Boulet. An image-based four-source surface energy balance model to estimate crop evapotranspiration from solar reflectance/thermal emission data (SEB-4S). *Agricultural and Forest Meteorology*, 184:188–203, 2014. doi: 10.1016/j.agrformet.2013.10.002.
- [42] S. Stisen, I. Sandholt, A. Nørgaard, R. Fensholt, and K.H. Jensen. Combining the triangle method with thermal inertia to estimate regional evapotranspiration - applied to MSG-SEVIRI data in the Senegal River basin. *Remote Sensing of Environment*, 112(3):1242–1255, 2008.

- [43] R. Guzinski, M.C. Anderson, W.P. Kustas, H. Nieto, and I. Sandholt. Using a thermal-based two source energy balance model with time-differencing to estimate surface energy fluxes with day-night MODIS observations. *Hydrology and Earth System Sciences*, 17(7):2809–2825, 2013.
- [44] R. Tang and Z. L. Li. An end-member-based two-source approach for estimating land surface evapotranspiration from remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 55(10):5818–5832, 2017. doi: 10.1109/TGRS.2017.2715361.
- [45] Mick J. Russell, Manuel Nunez, Mark A. Chladil, Jose A. Valiente, and Ernesto Lopez-Baeza. Conversion of nadir, narrowband reflectance in red and near-infrared channels to hemispherical surface albedo. *Remote Sensing of Environment*, 61(1):16 – 23, 1997. doi: 10.1016/S0034-4257(96)00218-0.
- [46] B. Franch, E.F. Vermote, J.A. Sobrino, and Y. Julien. Retrieval of surface albedo on a daily basis: Application to MODIS data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(12):7549–7558, 2014. doi: 10.1109/TGRS.2014.2313842.
- [47] W. Lucht, C.B. Schaaf, and A.H. Strahler. An algorithm for the retrieval of albedo from space using semiempirical BRDF models. *IEEE Transactions on Geoscience and Remote Sensing*, 38(2):977–998, 2000. doi: 10.1109/36.841980.
- [48] F Jacob, A Oliosio, M Weiss, F Baret, and O Hautecoeur. Mapping short-wave albedo of agricultural surfaces using airborne POLDER data. *Remote Sensing of Environment*, 80(1):36 – 46, 2002. ISSN 0034-4257. doi: [http://dx.doi.org/10.1016/S0034-4257\(01\)00265-6](http://dx.doi.org/10.1016/S0034-4257(01)00265-6).
- [49] Frédéric Jacob and Albert Oliosio. Derivation of diurnal courses of albedo and reflected solar irradiance from airborne POLDER data acquired near solar noon. *Journal of Geophysical Research: Atmospheres*, 110(D10), 2005. ISSN 2156-2202. doi: 10.1029/2004JD004888.
- [50] William P Kustas and John M Norman. Evaluation of soil and vegetation heat flux predictions using a simple two-source model with radiometric temperatures for partial canopy cover. *Agricultural and Forest Meteorology*, 94(1):13–29, 1999. doi: 10.1016/S0168-1923(99)00005-2.
- [51] G.S. Campbell and J.M. Norman. *An Introduction to Environmental Biophysics*. Springer-Verlag New York, New York, 2nd edition edition, 1998.
- [52] G. S. Campbell. Extinction coefficients for radiation in plant canopies calculated using an ellipsoidal inclination angle distribution. *Agricultural and Forest Meteorology*, 36(4):317–321, 1986.
- [53] W.P. Kustas and J.M. Norman. A two-source energy balance approach using directional radiometric temperature observations for sparse canopy covered surfaces. *Agronomy Journal*, 92(5):847–854, 2000.

- [54] B.J. Choudhury, S.B. Idso, and R.J. Reginato. Analysis of an empirical model for soil heat flux under a growing wheat crop for estimating evaporation by an infrared-temperature based energy balance equation. *Agricultural and Forest Meteorology*, 39(4):283–297, 1987.
- [55] J.A. Santanello Jr and M.A. Friedl. Diurnal covariation in soil heat flux and net radiation. *Journal of Applied Meteorology*, 42(6):851–862, 2003.
- [56] I. A. Walter, R. G. Allen, R. Elliott, M. E. Jensen, D. Itenfisu, B. Mecham, T. A. Howell, R. Snyder, P. Brown, S. Echings, T. Spofford, M. Hattendorf, R. H. Cuenca, J. L. Wright, and D. Martin. ASCE’s standardized reference evapotranspiration equation. In *Watershed Management and Operations Management 2000*. 2000.
- [57] Nishan Bhattarai, Lindi J. Quackenbush, Jungho Im, and Stephen B. Shaw. A new optimized algorithm for automating endmember pixel selection in the SEBAL and METRIC models. *Remote Sensing of Environment*, 196:178 – 192, 2017. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2017.05.009>.
- [58] J. Chirouze, G. Boulet, L. Jarlan, R. Fieuzal, J. C. Rodriguez, J. Ezzahar, S. Er-Raki, G. Bigeard, O. Merlin, J. Garatuza-Payan, C. Watts, and G. Chehbouni. Inter-comparison of four remote-sensing-based energy balance methods to retrieve surface evapotranspiration and water stress of irrigated fields in semi-arid climate. *Hydrology and Earth System Sciences*, 18(3):1165–1188, 2014. doi: 10.5194/hess-18-1165-2014.
- [59] Yuting Yang and Songhao Shang. A hybrid dual-source scheme and trapezoid framework-based evapotranspiration model (HTEM) using satellite images: Algorithm and model test: HTEM. *Journal of Geophysical Research: Atmospheres*, 118(5):2284–2300, 2013. doi: 10.1002/jgrd.50259.
- [60] Di Long and Vijay P. Singh. A Two-source Trapezoid Model for Evapotranspiration (TTME) from satellite imagery. *Remote Sensing of Environment*, 121(Supplement C): 370 – 388, 2012. doi: 10.1016/j.rse.2012.02.015.
- [61] Matteo Detto, Nicola Montaldo, John D. Albertson, Marco Mancini, and Gaby Katul. Soil moisture and vegetation controls on evapotranspiration in a heterogeneous Mediterranean ecosystem on Sardinia, Italy. *Water Resources Research*, 42(8), 2006. doi: 10.1029/2005WR004693.
- [62] Cheng-I Hsieh, Gabriel Katul, and Tze wen Chi. An approximate analytical model for footprint estimation of scalar fluxes in thermally stratified atmospheric flows. *Advances in Water Resources*, 23(7):765 – 772, 2000. doi: [https://doi.org/10.1016/S0309-1708\(99\)00042-1](https://doi.org/10.1016/S0309-1708(99)00042-1).
- [63] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini. Sentinel-2: ESA’s optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120:25–36, 2012. ISSN 0034-4257. doi: 10.1016/j.rse.2011.11.026.

- [64] Markus Immitzer, Francesco Vuolo, and Clement Atzberger. First experience with Sentinel-2 data for crop and tree species classifications in Central Europe. *Remote Sensing*, 8(3):166, 2016. doi: 10.3390/rs8030166.
- [65] JÃ©rÃ©me Louis, Vincent Debaecker, Bringfried Pflug, Magdalena Main-Knorn, Jakub Bieniarz, Uwe Mueller-Wilm, Enrico Cadau, and Ferran Gascon. Sentinel-2 Sen2cor: L2A processor for users. In L. Ouwehand, editor, *Proceedings Living Planet Symposium 2016*, volume SP-740, pages 1–8, Prague, Czech Republic, August 2016. Spacebooks Online. ISBN 978-92-9221-305-3.
- [66] Marie Weiss and Fred Baret. S2ToolBox Level 2 products: LAI, FAPAR, FCOVER - Algorithm Theoretical Basis Document, February 2016. URL [http://step.esa.int/docs/extra/ATBD\\_S2ToolBox\\_L2B\\_V1.1.pdf](http://step.esa.int/docs/extra/ATBD_S2ToolBox_L2B_V1.1.pdf).
- [67] Joshua B. Fisher, Kevin P. Tu, and Dennis D. Baldocchi. Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. *Remote Sensing of Environment*, 112(3):901 – 919, 2008. doi: 10.1016/j.rse.2007.06.025.
- [68] J.-B. F  ret, A.A. Gitelson, S.D. Noble, and S. Jacquemoud. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. *Remote Sensing of Environment*, 193:204 – 215, 2017. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2017.03.004>.
- [69] C. Donlon, B. Berruti, A. Buongiorno, M. H. Ferreira, P. F  m  nias, J. Frerick, P. Goryl, U. Klein, H. Laur, C. Mavrocordatos, J. Nieke, H. Rebhan, B. Seitz, J. Stroede, and R. Sciarra. The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission. *Remote Sensing of Environment*, 120:37–57, 2012. ISSN 0034-4257. doi: 10.1016/j.rse.2011.07.024.
- [70] J. A. Sobrino, J. C. Jim  nez-Mu  oz, G. S  ria, A. B. Ruescas, O. Danne, C. Brockmann, D. Ghent, J. Remedios, P. North, C. Merchant, M. Berger, P. P. Mathieu, and F. M. G  tttsche. Synergistic use of MERIS and AATSR as a proxy for estimating Land Surface Temperature from Sentinel-3 data. *Remote Sensing of Environment*, 179:149–161, June 2016. doi: 10.1016/j.rse.2016.03.035.
- [71] S Bontemps, P Defourny, J Radoux, E Van Bogaert, C Lamarche, F Achard, P Mayaux, M Boettcher, C Brockmann, G Kirches, et al. Consistent global land cover maps for climate modelling communities: current achievements of the ESA’s land cover CCI. In *Proceedings of the ESA Living Planet Symposium*, pages 9–13, 2013.
- [72] M. R. Raupach. Simplified expressions for vegetation roughness length and zero-plane displacement as functions of canopy height and area index. *Boundary-Layer Meteorology*, 71(1):211–216, 1994. ISSN 1573-1472.
- [73] K.J Schaudt and R.E Dickinson. An approach to deriving roughness length and zero-plane displacement height from satellite data, prototyped with BOREAS data. *Agricultural and Forest Meteorology*, 104(2):143 – 155, 2000. ISSN 0168-1923. doi: [https://doi.org/10.1016/S0168-1923\(00\)00153-2](https://doi.org/10.1016/S0168-1923(00)00153-2).

- [74] H. Hersbach. The ERA5 atmospheric reanalysis. *AGU Fall Meeting Abstracts*, December 2016.
- [75] J.-J. Morcrette, O. Boucher, L. Jones, D. Salmond, P. Bechtold, A. Beljaars, A. Benedetti, A. Bonet, J. W. Kaiser, M. Razinger, M. Schulz, S. Serrar, A. J. Simmons, M. Sofiev, M. Suttie, A. M. Tompkins, and A. Untch. Aerosol analysis and forecast in the European Centre for Medium-range Weather Forecasts integrated forecast system: Forward modeling. *Journal of Geophysical Research: Atmospheres*, 114(D6), 2009. doi: 10.1029/2008JD011235.
- [76] Pierre Ineichen and Richard Perez. A new air mass independent formulation for the Linke turbidity coefficient. *Solar Energy*, 73(3):151 – 157, 2002. ISSN 0038-092X. doi: [https://doi.org/10.1016/S0038-092X\(02\)00045-2](https://doi.org/10.1016/S0038-092X(02)00045-2).
- [77] Pierre Ineichen. Conversion function between the Linke turbidity and the atmospheric water vapor and aerosol content. *Solar Energy*, 82(11):1095 – 1097, 2008. ISSN 0038-092X. doi: <https://doi.org/10.1016/j.solener.2008.04.010>.
- [78] Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996. doi: 10.1007/BF00058655.
- [79] Minha Choi, William P. Kustas, Martha C. Anderson, Richard G. Allen, Fuqin Li, and Jeppe H. Kjaersgaard. An intercomparison of three remote sensing-based surface energy balance algorithms over a corn and soybean production region (Iowa, U.S.) during SMACEX. *Agricultural and Forest Meteorology*, 149(12):2082–2097, 2009. doi: 10.1016/j.agrformet.2009.07.002.
- [80] M.P. González-Dugo, C.M.U. Neale, L. Mateos, W.P. Kustas, J.H. Prueger, M.C. Anderson, and F. Li. A comparison of operational remote sensing-based models for estimating crop evapotranspiration. *Agricultural and Forest Meteorology*, 149(11):1843–1853, 2009. doi: 10.1016/j.agrformet.2009.06.012.
- [81] Kyle R Knipper, William P Kustas, Martha C Anderson, Joseph G Alfieri, John H Prueger, Christopher R Hain, Feng Gao, Yun Yang, Lynn G McKee, Hector Nieto, Lawrence E. Hipps, Maria Mar Alsina, and Luis Sánchez. Evapotranspiration estimates derived using thermal-based satellite remote sensing and data fusion for irrigation management in California vineyards. *Irrigation Science*, pages 1–19, 2018. doi: 10.1007/s00271-018-0591-y.
- [82] Martha C. Anderson, J. M. Norman, William P. Kustas, Fuqin Li, John H. Prueger, and John R. Mecikalski. Effects of vegetation clumping on Two-Source Model estimates of surface energy fluxes from an agricultural landscape during SMACEX. *Journal of Hydrometeorology*, 6(6):892–909, 2005.
- [83] D. Hillel. *Environmental Soil Physics*. Academic Press, 1998. ISBN 9780123485250.
- [84] T. J. Sauer, J. M. Norman, C. B. Tanner, and T. B. Wilson. Measurement of heat and vapor transfer coefficients at the soil surface beneath a maize canopy using source plates. *Agricultural and Forest Meteorology*, 75(1-3):161 – 189, 1995. ISSN 0168-1923.

- [85] Junsei Kondo and Sachinobu Ishida. Sensible heat flux from the Earth's surface under natural convective conditions. *Journal of the Atmospheric Sciences*, 54(4):498–509, 1997.
- [86] J. Goudriaan. Crop micrometeorology: A simulation stud. Technical report, Center for Agricultural Publications and Documentation, Wageningen, 1977.
- [87] C. H. B. Priestley and R. J. Taylor. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, 100(2):81–92, 1972.
- [88] A.M. Baldridge, S.J. Hook, C.I. Grove, and G. Rivera. The aster spectral library version 2.0. *Remote Sensing of Environment*, 113(4):711 – 715, 2009. doi: <https://doi.org/10.1016/j.rse.2008.11.007>.
- [89] Joseph G. Alfieri, William P. Kustas, Hector Nieto, John H. Prueger, Lawrence E. Hipps, Lynn G. McKee, Feng Gao, and Sebastian Los. Influence of wind direction on the surface roughness of vineyards. *Irrigation Science*, 2018. doi: [10.1007/s00271-018-0610-z](https://doi.org/10.1007/s00271-018-0610-z). URL <https://doi.org/10.1007/s00271-018-0610-z>.
- [90] Héctor Nieto, William P. Kustas, Joseph G. Alfieri, Feng Gao, Lawrence E. Hipps, Sebastian Los, John H. Prueger, Lynn G. McKee, and Martha C. Anderson. Impact of different within-canopy wind attenuation formulations on modelling sensible heat flux using TSEB. *Irrigation Science*, 2018. ISSN 1432-1319. doi: [10.1007/s00271-018-0611-y](https://doi.org/10.1007/s00271-018-0611-y).
- [91] B. J. Choudhury and J. L. Monteith. A four-layer model for the heat budget of homogeneous land surfaces. *Quarterly Journal of the Royal Meteorological Society*, 114(480):373–398, 1988. URL <http://dx.doi.org/10.1002/qj.49711448006>.
- [92] K.G. McNaughton and B.J.J.M. Van Den Hurk. A “lagrangian” revision of the resistors in the two-layer model for calculating the energy budget of a plant canopy. *Boundary-Layer Meteorology*, 74(3):261–288, 1995. doi: [10.1007/BF00712121](https://doi.org/10.1007/BF00712121).
- [93] Yan Li, William P. Kustas, Chunlin Huang, Hector Nieto, Erfan Haghghi, Martha C. Anderson, Francisco Domingo, Monica Garcia, and Russell L. Scott. Evaluating soil resistance formulations in thermal-based two source energy balance (TSEB) model: Implications for heterogeneous semiarid and arid regions. *Water Resources Research*, 0(ja). doi: [10.1029/2018WR022981](https://doi.org/10.1029/2018WR022981).
- [94] William P. Kustas and John M. Norman. A two-source approach for estimating turbulent fluxes using multiple angle thermal infrared observations. *Water Resources Research*, 33(6):1495–1508, 1997. doi: <http://dx.doi.org/10.1029/97WR00704>.
- [95] A. Chehbouni, Y. Nouvellon, J.P. Lhomme, C. Watts, G. Boulet, YH Kerr, MS Moran, and DC Goodrich. Estimation of surface sensible heat flux using dual angle observations of radiative surface temperature. *Agricultural and Forest Meteorology*, 108(1): 55–65, 2001.